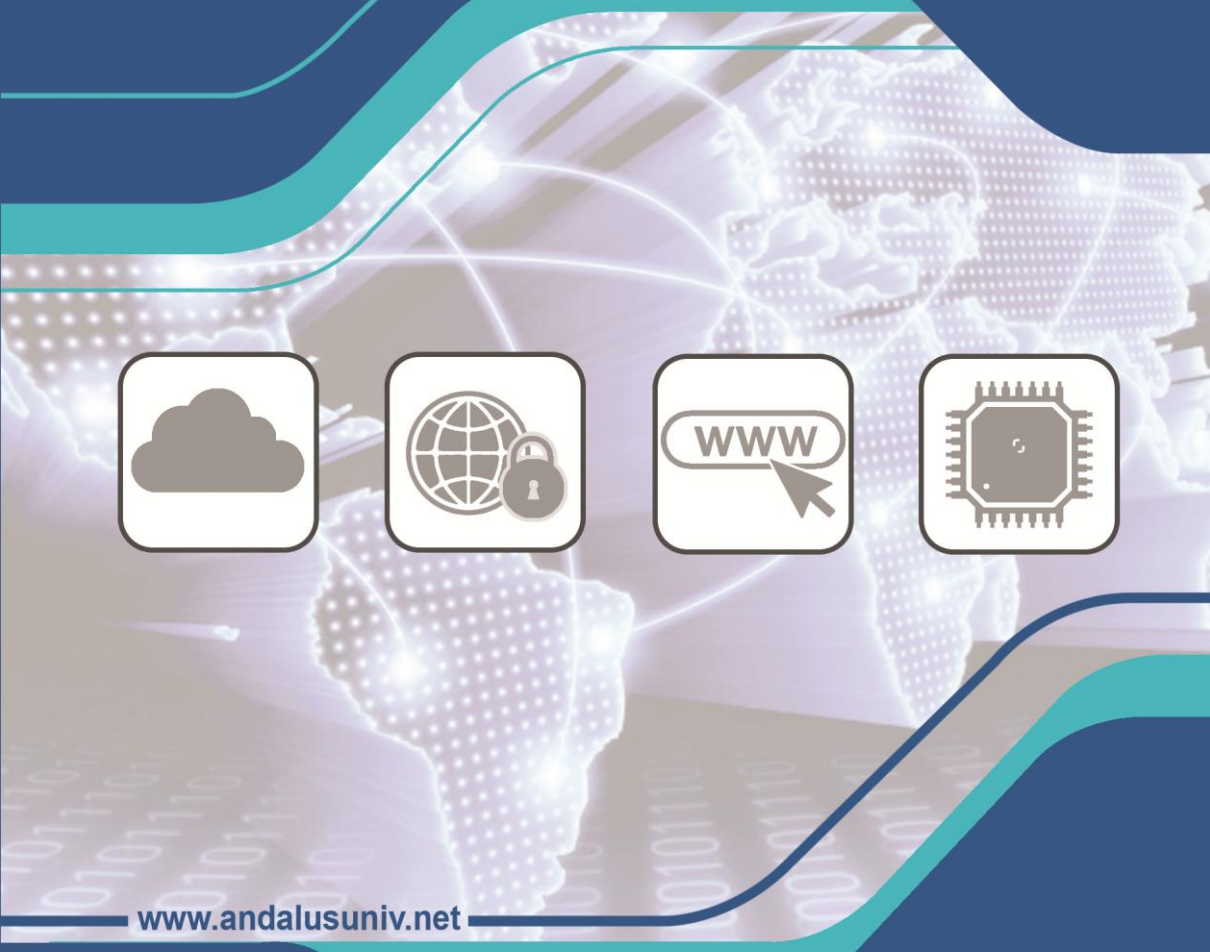
المـــؤتمر العـــلمي الأول فـي تقنية المعلومات والشبكات

Proceedings of 1st Scientific Conference on Information Technology and Networks

1-2 Novmber 2016



www.andalusuniv.net

جامعة الأندلس
للعلوم والتقنية
القيمة الكاملة للتعليم...

الجمهورية اليمنية-صنعاء-دارسلم
هـــاتف: 675567 1 967+
فاكـــس: 675885 1 967+
صندوق بريد: 37444

Proceedings of 1st Scientific Conference on Information Technology and Networks

1-2 November 2016

Organized by

Faculity of Engineering and Information Technology

AUST

Alandalus University for Science & Technology , Sana'a– Yemen

اللجنة الإشرافية (Steering Committee):

- ا.د/ أحمد محمد برقعـــان رئيس اللجنة، رئيس المؤتمر.
- ا.م. د/ فكـــري الحميـــدي نائب رئيس اللجنة، نائب رئيس المؤتمر.
- أ.د/ شـــرف الشهـــــاري. عضواً
- ا.م. د / عبدالله عباس مهدي. عضواً
- ا.م. د /يحيى عبدالرزاق قطران. عضواً
- ا.م. د /عبدالله عبدالرحمن بكير. عضواً
- د/ فـــؤاد علـــي اليريمـــي. عضواً
- د/رضـوان علـي المخـــــلافي. عضواً

اللجنة العلمية (Technical Committee):

- أ.د/ خليـــل الوجيـــة.
- أ.د/ غـالب الجعفـري.
- أ.م.د/ فكري الحميدي.
- أ.م.د/ ايوب الحمادي.
- أ.م.د/ شرف الحمدي.
- أ.م.د/ عـــادل ســـلام.
- أ.م.د/ فضـل باعلوي.
- د/ عبدالقادر العبادي.
- د/ عمـــــار الزهـــاري.
- د/ منير المخـــــلافي.
- د/ نشوان الذبحاني.
- د/ اسماعيل الاحمدي.
- د/ حسـن باســـــواد.

- د/ فـــؤاد اليريمـــي.
- د/ حســـن الشيـــخ.
- د/ نبراس العـــزعزي.
- د/ عبدالرحمن الهادي.

اللجنة الفرعية (Organizing Committees) :

❖ لجنة العلاقات العامة :

- م/ عبدالرحمن الخالدي. رئيساً
- أ/ فيصـــل الدعيـــس. عضواً
- أ/ أحمـــد النميـــر. عضواً
- أ/ بكـاريحـــي بكـــار. عضواً
- أ/ اشـــرف الزبـــيدي. عضواً
- أ/ صفـــوان الصلـــوي. عضواً
- أ/ أحمـــد العـــواضي. عضواً

❖ اللجنة الإعلامية :

- أ/ زينب العمـــــري. رئيساً
- أ/ محفوظ الميـــاسي. عضواً

❖ لجنة الاستقبال والتسجيل :

- د/ فـــؤاد اليريمـــي. رئيساً
- د/ حســـن الشيـــخ. عضواً
- د/ نبراس العـــزعزي. عضواً
- م/ موسى غرســـان. عضواً
- أ/ الهـــام السنباني. عضواً

❖ لجنة مطبوعـات المؤتمر:

- أ/ زيـــــاد المقبـولي. رئيساً
- م/ عمر سالم كعيتي عضواً
- م/ نـايـــف الحنـق عضواً

❖ **لجنة اعداد توصيات المؤتمر:**

- أ.د/ خليل الوجيـــه.
- أ.م.د/ فكري الحميدي.
- د/ احمـــد باحـــارثه.
- د/ عبدالقادر العبادي.
- د/ شـــرف الحمـــدي.
- د/ منـــير المخـــــلافي.
- د/ محمد القبـــاطي.

قائمة المحكميـــن (Reviewer List):

- أ.م.د/ فكري الحميدي. جامعة الاندلس للعلوم والتقنية.
- د/ عمـــار الزهـاري. جامعة ازال للتنمية البشرية.
- د/ فـــؤاد اليريمـــي. جامعة الاندلس للعلوم والتقنية.
- أ.د/ خليل الوجيـــه. جامعة ذمار.
- أ.م.د/ شرف الحمدي. جامعة صنعاء.
- أ.م.د/ فضـل باعلوي. جامعة صنعاء.
- د/ حســـن الشيـــخ. جامعة الاندلس للعلوم والتقنية.
- د/ نبراس العـزعـــزي. جامعة الاندلس للعلوم والتقنية.
- د/ انـــور سيـــــف. جامعة صنعاء.
- أ.م.د/ ماهـر السنباني. جامعة ذمار.
- د/ محمـــد الخولاني. جامعة حجة.
- د/ محمد القبـــاطي وزارة النفط.
- أ.م.د/ عـــادل ســـلام. جامعة عدن.
- د/ محمد الصـــارم. جامعة طيبة.
- د/ ابراهيـم البلطـــة جامعة صنعاء.
- د/ منـــير المخـــــلافي جامعة سباء.
- أ.د/ غـالب الجعفـري. جامعة صنعاء.
- د / عبداللطيف غلاب. جامعة العلوم والتكنولوجيا.

المحتويات

رسالة رئيس الجامعة (رئيس المؤتمر):

لقد دأبت جامعة الأندلس للعلوم والتقنية على تشجيع الابداعات والأنشطة الأكاديمية التي تتبناها الكليات والأقسام الأكاديمية في الجامعة، من خلال العديد من الأنشطة والتي على رأسها إقامة الندوات والمؤتمرات العلمية والترتيب المناسب لها واستضافة الباحثين والأكاديميين الذين يرفدونها بالأفكار الحديثة في مجال البحث.

وهذا العام 2016-2017م يتم افتتاحه بالمؤتمر العلمي الأول لتقنية المعلومات والشبكات والذي يشارك فيه العديد من العلماء والباحثين في هذا المجال من مختلف الجامعات اليمنية والشركات العاملة في المجال وكذا طلاب الدراسات العليا ومشاريع التخرج المتميزة للبكالريوس.

ونتمنى للمؤتمر النجاح والخروج لقرارات تطور رؤية مشتركة للجامعات في هذا المجال الحيوي، حيث سبق المؤتمر ورشة عمل حول تحديد أولويات البحث في مجال تقنية المعلومات برعاية وزارة التعليم العالي لكي تقر في هذا المؤتمر.

وختاماً نشكر للجنة المنظمة جهودها المبذولة لإنجاح هذا المؤتمر نأمل ان تحقق أهدافها العلمية والتوعوية، ويجني المشاركون فيها الفوائد المرجوة.

وفق الله الجميع لكل خير

أ.د. أحمد محمد برقعان
رئيس جامعة الأندلس

رسالة عميد الكلية (نائب رئيس المؤتمر):

مرحبا بكم في رحاب جامعة الاندلس للعلوم والتقنية في المؤتمر العلمي الاولى لتقنية المعلومات والشبكات 2016م (SCITN'2016). استمرارا لا يماننا بأهمية البحث العلمي في تطور الشعوب والمجتمعات ورقيها، وتحقيقا لرسالة الجامعة فأننا نخطو خطوات جادة في هذا المسار ولعل سابقا الندوة العلمية الاول عام 2014م وحاليا المؤتمر العلمي الاول (SCITN'2016) هي من ثمار هذا التوجه لتوفير البيئة المناسبة للمهتمين في البحث العلمي في مجال تقنية المعلومات والشبكات للالتقاء وتبادل الافكار والمشاكل العلمية والبحثية وطرح الحلول المناسبة . ان التطور السريع والهائل في مجال تقنية المعلومات وشبكات الاتصال في العالم وضع امامنا تحديات في مواكبة هذا التطور ومعرفة الجديد وعكسه على تطوير واقعنا ولايتأتا هذا الا بالبحث العلمي وربطة بالمؤسسات الوطنية وتحويل مشاكل هذه المؤسسات الى مشاكل بحثية يتم توجيه مؤسساتنا العلمية وطلابنا لحلها. هناك مجالات كثيرة تحتاج الى بحث ودراسة مثل Cloud Computing ، Green Computing ، Mobile Computing، Vehicular Networks ، Intelligent Transportation Systems، Web/Data Mining ، Cyber Security ومجالات اخرى كثيرة نحن كمتخصصين وباحثين وعلماء في حاجة الى لقاءات علمية سنوية لتبادل الافكار والبحث عن الحلول ، وعكسها في تطوير قدرات ومهارات ابناءنا الطلاب.

في الاخير احب ان اشكر رئاسة الجامعة واعضاء اللجنة العلمية واعضاء اللجنة الاشرافية واللجنة المنظمة وكذلك المشاركين لمساهماتهم في انجاح هذا اللقاء العلمي. مرة اخرى ارحب بكم الى جامعة الاندلس للعلوم والتقنية واتمنى ان تتحقق الفائدة المرجوة من هذا المؤتمر وان يتم الاستفادة منه وان تكون نقطة انطلاقه لمن يرغب في البحث العلمي مستقبلا.

د. فكري الحميدي

عميد كلية الهندسة وتقنية المعلومات

رسالة وكيل قطاع البحث العلمي (وزارة التعليم العالي):

يسعدنا في وزارة التعليم العالي والبحث العلمي ان نبارك و نثمن الجهود والأنشطة التي تقوم بها جامعة الأندلس للعلوم والتقنية من أجل المؤتمر العلمي وكذلك في سبيل تحديد أولويات البحث العلمي في مجال تقنية المعلومات والاتصالات .

ونحن اذ نتابع ونشرف على الآلية والاجراءات والخطوات التنفيذية المتبعة من قبل الجامعة للمؤتمر العلمي و في تحديد أولويات البحث العلمي في مجال تقنية المعلومات والاتصالات نؤكد أن ذلك يأتي في إطار توجه وزارة التعليم العالي والبحث العلمي نحو تعزيز دور البحث العلمي في عملية التنمية الاقتصادية والاجتماعية في الجمهورية اليمنية.

ومن هذا المنطلق فان تحديد جامعة الأندلس للعلوم والتقنية لأولويات البحث العلمي في مجال تقنية المعلومات والاتصالات يمثل مساهمة حقيقية وفاعلة في المشروع الوطني لتحديد أولويات البحث العلمي في الجمهورية اليمنية الذي يشمل جميع المجالات والحقول العلمية والذي تنفذه وزارة التعليم العالي والبحث العلمي بالتعاون مع الجامعات اليمنية.

إن مشروع تحديد أولويات البحث العلمي في الجمهورية اليمنية يعد تظاهرة علمية وطنية يمثل الخطوة الأولى على طريق بلورة رؤية وطنية متكاملة لدور البحث العلمي في النهوض بالتنمية الاقتصادية الشاملة.

كما ان مشروع تحديد أولويات البحث العلمي في الجمهورية اليمنية يمثل خارطة طريق أمام الجامعات والمؤسسات والمراكز البحثية والعلمية من أجل المساهمة الحقيقية والفاعلة في خدمة التنمية والمجتمع للدفع بعجلة التنمية الاقتصادية نحو الأمام .

وفي الختام اتقدم بجزيل الشكر والتقدير إلى جميع منتسبي وموظفي جامعة الأندلس للعلوم والتقنية نظير الجهود التي يبذلونها في سبيل نجاح المؤتمر وفي تحديد أولويات البحث العلمي في مجال تقنية المعلومات والاتصالات .

كما أشكر الخبراء والباحثين والمتخصصين في مجال تقنية المعلومات والاتصالات من مختلف المؤسسات الوطنية من الجامعات الحكومية والأهلية والمراكز البحثية والعلمية ومؤسسات القطاع العام والخاص ومؤسسات المجتمع المدني الذين شاركوا في الأعداد للمؤتمر وفي تحديد أولويات البحث العلمي في مجال تقنية المعلومات والاتصالات.

والله ولي الهداية والتوفيق ،،،

د. صادق الشراجي
وكيل قطاع البحث العلمي

A Framework for Name Matching in Arabic Language

Salah AL-Hagree
Faculty Computer Science and Information Systems, Thamar University, Yemen
s.alhagree@gmail.com

Maher Al-Sanabani
Faculty Computer Science and Information Systems, Thamar University, Yemen
M.sanabani@gmail.com

Abstract

An extensive research has been done for searching an effective algorithm for name matching that is play a vital and crucial role in many applications. Therefore, a many algorithms have been developed to measure the similarity of string but most of them designed mainly to handle Latin based languages. While, the name matching algorithms on Arabic context is rare because the dealing with Arabic context is a challenging task due to the characteristics and unique features of the Arabic language. Consequently, a framework for Arabic name matching has been proposed in this paper. The proposed framework takes a unique features of the Arabic language and the different levels of similarity for the Arabic letters such as phonetic, letter's form and keyboard similarities. Furthermore, the proposed framework has been considered transposition operation and enhanced states of insertion and deletion operations. The carried experiments in this paper have been shown the proposed framework gives more accurate results than the compared algorithms.

Keywords: Arabic Name Matching , Bigram, Matching Framework, Levenshtein Distance.

1. Introduction

The name matching is hot topic from beginning of computer science. It is a challenge for all community of scientists to devise more efficient algorithm. Therefore, a many name matching algorithms have been developed and used to cope with important topic. The name matching algorithms have been classified into two categories which are Exact matching algorithms [1],[2] and approximate (inexact) string matching algorithms [3],[4],[5],[6].

Discovery and matching of names, personal names or place names or company names or Scientific names is used in an increasing number of applications and it constitutes a central part of many applications such as Customer Relation Management (CRM), Customer Data Integration (CDI), Anti-Money Laundering (AML), Criminal Investigation (CI), HealthCare(HC), and Genealogy Services(GS). If only exact matching was available in these types of applications it would not be possible to deal with name variations, which unavoidably occur in the data and names

in real world data sets. Therefore, the exact matching techniques are not suitable for large and complex information system because they cannot be able to retrieve names with more than one acceptable spelling. In order to get more accurate results, an approximate name matching should be applied instead of exact matching. Therefore, the motivation in this paper is providing a matching algorithm for Arabic names that is considering on an approximate string matching algorithms which can be dealing with a technician errors allows in field of computer science. This type of matching algorithms have been implemented in many application such as computational biology "DNA" [7]spelling correction [8],[9] text retrieval [10],[11] [12]Handwriting Recognition, Linking database [13],[14] and Name recognition [15].

This paper is organized as follows. Section 2 illustrates some challenges of Arabic Language. Section 3 describes the related work in field of study. Section 4 demonstrates the proposed framework for Arabic name matching. Section 5 presents the experimental and results discussions. Finally conclusions and future work are presented in Section 6 .

2. Challenges of Arabic Language

Arabic language is one of the world's major language and used by nearly two billion people. Therefore, a wide belt of nations are spoken Arabic language which has several challenges , features and characteristics. The Arabic language has been categorized into three main types such as Classical Arabic (CA), Modern Standard Arabic (MSA) and Colloquial/Dialects Arabic (C/DA) [16]. In this work, MSA are dealt. Moreover, Arabic-like scripting is used in several languages, such Urdu (the official language of Pakistan), Persian (the official language of Iran) and Pashto (the official language of Afghanistan). In general, the Arabic alphabet has 28 basic letters . In some cases, researchers add the four Lam-Alif "لا" ligatures to form 36 Arabic letters (viz., the basic 28, Alif-Maqsura "ى", Ta-Marbuta "ة", and the Hamza "ء" letters), which results in a total of 40 letters [17],[18],[19]. Generally, the feature of word in Arabic language such as the length is shorter than the length of the word in Latin based languages. A lot of Arabic words from short roots can be driven .

Most Arabic writers are frequently confused about the placement of hamzah "ء" /ʔ/ (*glottal stop*) which has eight different forms (ـأ - ء - ـإ – ـا – ـؤ – ـآ – ـئ – ـا). Alif Maqsoura and Yaa (ـى – ـي) are often confused at the end of words, as one of them has two dots beneath and the other does not; the same with Taa Marboutah and Haa (ـة - ـه), where one of them has two dots above and the other does not have. The presence of the similar form letter such as: {ـفـ - ـغـ} , {ـعـ - ـهـ},{ـصـ - ـحـ},{ إ, أ, آ} ,{ر,ـر} ,{ـذ,ـذ},{ـتـ - ـنـ - ـيـ - ـبـ}.

Arabic name poses several issues such as:

- At first issue, the insertion or deletion cost of a blank has been defined to be equals to zero beneficial to segmentations which might occur in names. Thus if

a blank appears accidentally inside a name. For example, spelling the word " عبد الله" instead of " عبدالله " , "ابوبكر" instead of "ابو بكر" and "رئيسالجمهورية" instead of " رئيس الجمهورية "

- The second issue is that if Hamza letter /?/ (*glottal stop*) is placed after long vowels (Alif, Yaa, and Waw) in a name, the Hamza can be removed without any negative effect. Both names, with and without Hamza are correct. For example, spelling the word "اسما" instead of " أسماء " and "انباء" instead of the "انبا".
- The third issue is that if duplicate letter, a diacritic called "Tashdid" is used above the letter which is pronounced twice. Almost in all cases, especially in names, "Tashdid" is also dropped from the name such as "محممد" should be written as "محمّد" but it is written in form of "محمد" in personal names and the letter of "م" repeated just in pronunciation.
- The fourth issue spelling mistakes typographical errors were frequently made by Arabic writers according to certain characters. For example, " الإسلامي" / The Islamic with " إ" can be written "الاسلامي" with " ا".
- The fifth issue Editing Errors: These errors result when learners misspell a word by omitting, adding, replacing or duplicating a letter within a given word. For example, spelling the word " متبة" instead of the " مكتبة ", " محممد " instead of " محمد " , " كتتاب " instead of " كتاب" and " فاطمة " instead of " فاطمه ".
- The sixth issue An analysis of the characters subject on substitution and transposition error has led us to conclude that substitution and transposition can occur because of three levels of similarity as flowing [9],[19]:
  - The first levels ,the similarity of keyboard relates to the proximity between the Arabic keyboard keys, which means that instead of entering a specific key, the operator clicks on the key directly adjacent to the correct one. For example, spelling the word "صلاح" instead of the " صباح "
  - The second levels, the similarity of letter's form has to do with the similarity between Arabic characters: the similitude among certain Arabic character is a plain evidence. One might take as an example the similarity in calligraphy between: (ض,ص), (ذ,د), (ز,ر), (ق,ف), (خ,ج,ح), (ظ,ط), (غ,ع) et.
  - The three levels, the similarity of phonetic (ص,س) for example, spelling the word "عصفور" instead of " عسفور ", (ض,د) for example, spelling the word "دجاجة" instead of " ضجاجة " et.

3. Related work

A brief survey of the related work in this research direction is presented below.

In [20] Levenshtein introduces Edit distance algorithm which is used for (Pattern Matching) string processing. This algorithm measures the difference between two string sequences. Levenshtein distance counts the minimum number of single-character edits (Deletion, Substitution and Insertion) required to change one word into the other, where the cost of substitution is the same as the cost of deletion or insertion,

depends on binary codes. This work does not consider the transposition operation of two adjacent characters.

Damerau [21] improved Levenshtein algorithm by additional operation to check the distance between strings, that's it, a transposition of two adjacent characters. Damerau stated 'that this algorithm has corresponded to more than 80% of human misspellings'. By taking four string operations (Insertion, Deletion, Substitution, and Transposition) , this algorithm has also been used in biology to measure the variation between DNA and other applications. This work does not consider the multi states of transposition operation.

In [22] study aims to enhance Levenshtein algorithm for matching Arabic names depending on Matrix character's form similarity, Matrix phonetic similarity and Matrix letter proximity in keyboard depending on the law of Euclidian computing distance between letters. Its disadvantage is that, matching Arabic names returns 1 if and only if the strings are identical, and more of 0.10 if and only if the strings have no identical and cannot the allows in transposition errors as mentioned in Levenshtein algorithm. The pronunciation depends on Arabic language types such as modem standard or colloquial , classical ; for example in the Egyptian colloquial al Gaf is pronounced like al Alif 'ادري ' will be 'قدري ' and in Upper Egypt colloquial al Gaf is pronounced as Jim therefore ' 'قدري " will be pronounced as 'جدري '.

Furthermore, there are several studies that used Levenshtein distance for different applications such as: Spellchecking for Arabic [8],[9],[23],[24].

In [25] Performs Modified Damerau-Levenshtein Distance (MDLD) test on two input strings which is support block transpositions of multiple characters that have been applied on Oracle database. This work increase the time complex to $O(N^3)$. Furthermore, this work cannot handle possible error and spelling variations.

In [3],[5],[26] are used Q-grams which are substrings of length q of a longer string. Depending on q, q-grams are called unigrams, bigrams (or 2-grams), trigrams, and so on. For instance, all bigrams of the string computer are "co", "om", "mp", "pu", "ut", "te", "er". One way to compute the similarity between two strings is the Dice coefficient in Equation (1) [27]:

$$\mathrm{DICE}_{(\mathbf{s},\mathbf{t})} = \mathbf{2} * \frac{\mathbf{A}(\mathbf{a}) \cap \mathbf{A}(\mathbf{b})}{\mathbf{A}(\mathbf{a}) \cup \mathbf{A}(\mathbf{b})}, \qquad (1)$$

Where A(x) is the set of bigrams of a string x, i.e., the measure is obtained by dividing the number of bigrams common to the two strings by the average number of bigrams in each string. This work cannot handle possible error and spelling variations because this work handles word rather than character . Furthermore, stemming process is needed to get high similarity between two words.

In [10],[11] have been used n-grams algorithm in text retrieval for Arabic . These algorithms have been shown low the similarity between two words when they are using pure n-grams, for example the similarity measure of the two words "الاستفسارات" and "استفسر" would be 0.533. Therefore, they have been improved the results of the N-gram techniques by applying a two-step approach, in which the N-gram approach is combined with a stemming technique.

In [27] Kondrak suggests a hybrid method (N-DIST) that is mixing the components of Edit Distance and N–Gram algorithms and proposed a new similarity measurement. This measurement has been evaluated depend on Genetic cognates words of the same origin that belong to distinct languages. For example, German "vater", English "father" and Norwegian "far" constitute a set of cognates, Confusable drug names, and Translational cognates. This algorithm takes the advantage of Levenshtein and n-grams algorithms. Therefore, this algorithm has been increased the time complex to $O(N^3)$. This work does not consider the transposition operation of two adjacent characters.

In [28],[29] have been used N-DIST algorithm in different applications based on English language while Sanabani et al.[19] enhanced N-DIST algorithm for matching Arabic names that is called N-DIST-A. This algorithm takes a characteristics and a unique features of the Arabic language. Furthermore, it takes the different levels of similarity for the Arabic letters such as phonetic, letter's form and keyboard similarities. However, this N-DIST-A algorithm dos not take into account multiple state of transposition to correct errors.

The N-DIST-A algorithm has been considered in this paper. Therefore, the proposed algorithm in the proposed framework will be improve the N-DIST-A algorithm by including multiple stats for transposition and substitution operations and addition two states to insertion and deletion operations. This enhancement to handle a different possible variation (such as typographical and spelling errors) of same names.

4. The Proposed Framework

This section, introduces the proposed framework for Arabic name matching that is shown in Figure 1. The proposed framework includes four steps to compute the distance between two names.

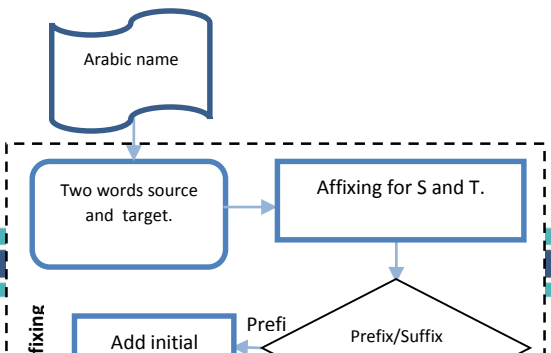


Figure 1. The proposed framework for Arabic name matching.

These steps are affixing step , join unigram to bigram and split bigram to unigram step and step to compute the weights of different costs for replacement, transposition, deletion and insertion operations, respectively; and estimate the costs of phonetic similarity, keyboard distance and character form similarity and the final step computes the distance minimum transformation cost for two names and percentage of their similarity.

The detail process to compute the distance of two Arabic names in the proposed framework are as follows :

Affixing Step: The affixing method aims to emphasizing the initial segments, which tend to be much more important than final segments in determining the similarity of names. A unique special symbol is defined for each letter of the original alphabet. Each name is augmented with a prefix composed of n-1 copies of the special symbol that corresponds to the initial letter of the name. For example, 'محمد 'is transformed into 'محمد - '. That begin by forming an (n + 1) X (m + 1) matrix MD initially containing all zeros, that is $MD_{(i,j)} = 0$ where $i = 0,1,2,\ldots,n$ and $j = 0,1,2,\ldots,m$.

Unigram Join and Bigram Split Step:

The goal idea is that assign values $MD_{(i,0)}=I$ where i=0,1,2,…,n and $x_{[i]}$= S.Substring(i, 2), also $MD_{(0,j)} = j$ where j = 0,1,2,…,m and $y_{[j]}$= T.Substring(j, 2). These transaction are shown in Figure 2.

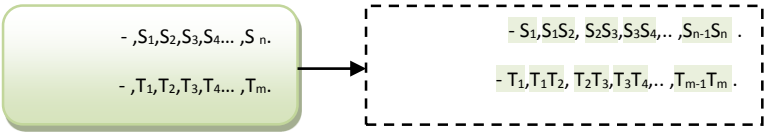


Figure 2. Join unigram to Bi-gram

While split Bi-gram to unigram is shown in Figure 3. This proses started to fill the values of $MD_{(i,j)}$ from second top row and going from left to right according to the following step:

$S_1=x_{[i-1]}$; $S_2=y_{[j-1]}$; $S_{11}=S_1$.Substring(0,1);$S_{12}=S_1$.Substring(1,1);$S_{21}=S_2$.Substring(0,1);$S_{22}=S_2$.Substring(1, 1).

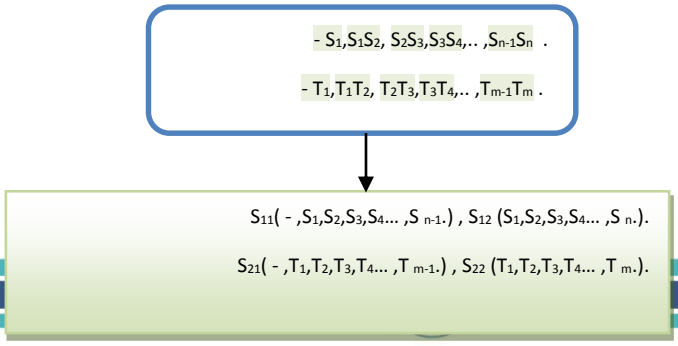


Figure 3. Split Bi-gram to unigram.

Computes The Weights Step :

The proposed framework has been introduced a new states for the substitution (replacement) cost and transposition cost operations as shown in Figure 4. These states are weighted as symbols w_1, w_2, w_3, w_4, w_5, w_6 and w_7 . These weights depend on three levels of similarity as mentioned sixth issue at section 2. are used to adapt to operational environment and get more accurate results on different situations. The cost of substitution and transposition can be note as C_{st} . The computed C_{st} can take value within an interval [0.0, 1.0] based on seven cases as shown in Figure 4.

Figure 4. Weighted cases substitution and transposition.

Inserting and deleting operations return the cost of inserting and deleting the letter (S_{11},S_{22} or S_{12},S_{21}) shown in Figure 5. depending on the conditions which are shown in Figure 1. inserting and deleting returns the value w_8 and w_9 respectively as the cost of inserting and deleting the letter (S_{11},S_{22} or S_{12},S_{21}). These weights handle the first, second and third issues which are mentioned at section 2. .

Figure 5. Weighted cases inserting and deleting .

Compute The Distance and Percentage of Their Similarity Step:

The distance between the name S and name T is compute using the Equation (2) [9],[22], [23].

$$\mathbf{MD}_{S,T}(i,j) = \text{Min}\begin{cases} \mathbf{MD}_{S,T}(i-1,j-1) + \mathbf{Cst} // \text{Substitution and Transposition} \\ \mathbf{MD}_{S,T}(i-1,j) + \mathbf{Cost}\text{ID} \quad // \text{deletion} \\ \mathbf{MD}_{S,T}(i,j-1) + \mathbf{Cost}\text{ID} \quad // \text{insertion} \end{cases} \quad (2)$$

The value MD $_{(n,m)}$ is the minimum transformation cost for two names which computed in dynamic programming method. The final step is determine the percentage of similarity between the names S and T. The value of their similarity has been computed by the following Equation (3) [5],[30],[31].

$$\text{Sim}(s,t) = 1 - \left(\frac{\text{MinimumDistnace}}{\text{Max}(|s|,|t|)} \right) \quad (3)$$

Where | S | is the length of source name and | T | is the length of target name.

5. The Experiment and Results

This section shows the experiments that have been carried in this paper to illustrate the proposed framework that is called **(HL-Big-A)** and compare it against the compared algorithm.

5.1. Data Preparation of Dataset

This section describes the names of Arabic that is used to test the proposed framework for Arabic name matching .

For more investigation a collection of datasets have been used in this experiment for testing the proposed **(HL-Big-A)** and compared algorithms. Because no standard collection of Arabic names exists, therefore, Four datasets have been extracted manually form that are named Dataset0 [19],[22], Dataset1 [32], Dataset2 [33] and Dataset3 [19] respectively. While Dataset4 [19] has been obtained from the employees database of Ministry of Education in Yemen. Each dataset contains some of Arabic names with different possible variation (such as typographical and spelling errors) of same names. A collection of all kind of variation have been considered in the variation of datasets.

5.2. Experimental Results

The first experiment has been carried based on Dataset0 which has 10 pairs of names shown in Table 1. the proposed framework **(HL-Big-A)** gives more accurate results than the compared algorithm **(N-DIST and N-DIST-A)** for all pairs in dataset0 as shown in Table 1.

Table 1.Comparison between algorithms in Arabic Dataset

No	String.		Compared Algorithm				Proposed framework	
			N-DIST		N-DIST-A		HL-Big-A	
	S	T	Dist.	Sim %	Dist.	Sim %	Dist.	Sim %
1	عمار	عممار	1.00	0.80	0.25	0.95	0.25	0.95
2	عصار	عسار	1.00	0.75	0.43	0.89	0.17	**0.96**
3	انباء	امباء	1.00	0.80	0.49	0.90	0.17	**0.97**
4	باهر	ناهر	1.50	0.63	0.58	0.86	0.67	0.83
5	سنا	ثناء	2.50	0.38	0.44	0.89	0.24	**0.94**
6	مراد	موراذ	2.00	0.60	1.25	0.75	1.62	0.68
7	أسماء	اسما	2.50	0.50	0.00	1.00	0.00	1.00
8	كامل	كمال	2.00	0.50	1.50	0.63	0.50	**0.88**
9	المرء	ألمر	2.50	0.50	1.00	0.80	0.50	**0.90**
10	عبدالله	عبد الله	1.50	0.81	0.00	1.00	0.50	0.94
Average(percentage similarity)			0.63		**0.87**		**0.91**	

Experiment has been carried with a variety of datasets to get the evidence of proposed framework ability. Four datasets are selected and applied on the proposed and compared algorithms as shown in Table 2. gives the evidence of proposed framework ability and shows the proposed framework gives more accurate results than the compared algorithm for all datasets .

Table 2. Result of testing the Proposed framework and Compared algorithm with different dataset.

Dataset	Compared Algorithm		Proposed framework
	N-DIST	N-DIST-A	HL-Big-A

Dataset1(11 pairs)	0.734	0.83	**0.881**
Dataset2(50 pairs)	0.811	0.88	**0.928**
Dataset3(10 pairs)	0.771	0.79	**0.880**
Dataset4(600 pairs)	0.738	0.87	**0.940**

6.Conclusion

In this paper a proposed framework has been introduced that is used a new methodology to improve the Arabic name matching accuracy.

The proposed framework include affixing, join unigram to bigram and split bigram to unigram and unique features of Arabic language of many weighted, including phonetic similarity, keyboard distance and character form similarity.

The main contributions of this paper is the methodology of handling the transposition. The transposition are handled differently at the different states of the name matching in order to improve. Another contribution is the intensive investigation of several novel features of Arabic language have been considered in the proposed framework such as hamzah /?/ (*glottal stop*) by putting rules for deletion and insertion operations. Furthermore, it gives different weights for replacement, transposition, insertion, and deletion operations. Therefore, the proposed framework gives more accurate results than the compared algorithms. In the future, the proposed framework will be extend to be able to deal token name Arabic .

7. References

[1] C. Charras , T. Lecroq, " Handbook of exact string matching algorithms", King's College Publications, 2004.

[2] Abdulrakeeb M. Al-Ssulami , "Hybrid string matching algorithm with a pivot", Journal of Information Science 2015, Vol. 41(1) 82–88, 2015.

[3] G. Navarro, "A guided tour to approximate string matching"," ACM computing surveys, vol. 33, no. I, pp. 31-88, . 2001.

[4] Hall, P. A. V., Dowling, G. R., " Approximate string matching. ACM Computing Surveys (CSUR) ", 12, 4, 381,402. 1980.

[5] Christen, P. ,"A Comparison of Personal Name Matching Techniques and Practical Issues", Technical Report TR-CS-06-02, Joint Computer Science Technical Report Series, Department of Computer Science, 2006.

[6] MPJ van der Loo, "The stringdist package for approximate string matching". The RJournal 6(1) 111-122, 2014.

[7] D. Cantone, S. Cristofaro, and S. Faro, "Efficient string-matching allowing for non-overlapping inversions," Theoretical Computer Science, 2012.

[8] H. Gueddah, A. Yousfi, M. Belkasmi, " Introduction of the weight edition errors in the Levenshtein distance ", International Journal of Advanced Research in Artificial Intelligence, Vol 1 Issue 5, pp 30-32, 2012.

[9] H. Gueddah, A. Yousfi, "The Impact of Arabic Inter-character Proximity and Similarity on Spell-Checking ", International Conference on Intelligent Systems: Theories and Applications ,pp 162,2013.

[10] S. Mustafa and Q. Al-Radaideh, "Using N-Grams for Arabic Text Searching", Journal Of The American Society For Information Science And Technology, vol.55, pp.1002–1007, 2004.

[11] S. Mustafa, "Character contiguity in N-gram-based word matching: the case for Arabic text searching", Information Processing and Management, vol.41, pp.819–827, 2005.

[12] Chowdhury SD, Bhattacharya U, Parui SK. "Online Handwriting Recognition Using Levenshtein Distance Metric". Document Analysis and Recognition (ICDAR), 12th International Conference;p.79,83, 25-28, 2013.

[13] Tarek El-Shishtawy , "Linking Databases using Matched Arabic Names ", Computational Linguistics and Chinese Language Processing ,Vol. 19, No. 1, pp. 33-54, 2014.

[14] B. W. Paleo, C. G. G. Hita, J. d. C. Lima, C. Ribeiro, and J. JambeiroFilho, "A Modified Edit-Distance Algorithm for Record Linkage in a Database of Companies," 2014.

[15] Juan Delgado, Fernando Galárraga, Walter Fuertes, Theofilos Toulkeridis, Cesar Villacís, Fidel Castro, "A Proposal of an Entity Name Recognition Algorithm to Integrate Governmental Databases" , 2016.

[16] Hamzah A. Alsayadi and Abeer M. ElKorany , "Integrating Semantic Features for Enhancing Arabic Named Entity Recognition" ,International Journal of Advanced Computer Science and Applications, Vol. 7, No. 3, 2016.

[17] Nizar Habash. "Introduction to Arabic Natural Language Processing". Morgan & Claypool Publishers, 2010.

[18] Baligh M. Al-Helali & Sabri A. Mahmoud "A Statistical Framework for Online Arabic Character Recognition", Cybernetics and Systems, 47:6, 478-498, 2016.

[19] Maher Sanabani, Salah Al-Hagree. ,"Improved An Algorithm For Arabic Name Matching". Open Transactions On Information Processing ISSN(Print): 2374-3786 ISSN(Online): 2374-3778.2015.

[20] V. I. Levenshtein, "Binary codes capable of correct-ing deletions, insertions, and reversals", Soviet Physics,pages 845–848, 1966.

[21] Damerau, F., "A Technique for Computer Detection and Correction of Spelling Errors", Communications of the ACM, 7(3), pp.171-176, 1964.

[22] Hesham H. , Ali EI-Bastawissy , Abdel Fattah , " AEDA: Arabic Edit Distance Algorithm Towards A New Approach for Arabic Name Matching", IEEE

international conference, IEEE Trans. Pattern Analysis and Machine Intelligence, 15(9):926–932,2011.

[23] Aouragh S, Gueddah, H., Yousfi, A.“ Adaptating the Levenshtein Distance to Contextual Spelling Correction”. International Journal of Computer Science &Applications;. p.127-133, (2015).

[24] Gueddah, H., Yousfi, A., & Belkasmi, M., “The filtered combination of the weighted edit distance and the Jaro-Winkler distance to improve spellchecking Arabic texts”. *International Conference of Computer Systems and Applications (AICCSA)* (pp. 1-6). IEEE, *2015*.

[25] Boehmer and Ton, is available on https://confluence.csiro.au/public/taxamatch/the-mdld-modified-damerau-levenshtein-distance-algorithm ,2008.

[26] Ullmann, J.R.: A binary n-gram technique for automatic correction of substitution deletion, insertion and reversal errors in words. The Computer Journal 20(2), 141– 147 (1977).

[27] Kondrak, G, “N-gram similarity and distance”, In M. Consens and G. Navarro (eds.), Proceedings of the String Processing and Information Retrieval 12th International Conference, Buenos Aires, Lecture Notes in Computer Science, 3772, Springer, Heidelberg, Germany, 115–126, 2005.

[28] Abdulhayoglu, M. A., & Thijs, B. “Matching bibliometric data from publication lists with large databases using n-grams”. In Proceedings of 14th international society of scien to metrics and informetrics conference (ISSI), Vienna, Austria, Vol. 2, pp. 1151–1158, 2013.

[29] Abdulhayoglu, M. A , Bart Thijs , Wouter Jeuris , “Using character n-grams to match a list of publications to references in bibliographic databases” , DOI 10.1007/s11192-016-2066-3,2016.

database”, Pattern Recognition 47/1096–1112, 2014.

[30] Gang Cheng, Fei Wang, HaiyangLv,Yinling Zhang , “A New Matching Algorithm for Chinese Place Names”, IEEE , 2011.

[31] Wael Hassan Gomaa, Aly Aly Fahmy, “Automatic scoring for answers to Arabic test questions”, Computer Speech & Language 28:4, 833-857,2013.

[32] Mahmoud, S. & AlMuhtaseb, H., “ Automatic Arabic Optical Text Recognition (AOTR) ”, Final Report, Research Project # IN060337, KFUPM, 2010.

[33] Sabri A.Mahmoud, IrfanAhmad, Wasfi G. Al-Khatib, Mohammad Alshayeb, Mohammad TanvirParvez, VolkerMärgner, GernotA.Fink, “ KHATT:An open Arabic offline handwritten text

Comparative Study on Cloud Portability and Interoperability using Semantic Representation

Rasha Ali Al-Arasi
Department of Computer Science
Sana'a University ,Yemen
Email: rasha.ali66@gmail.com

Dr. Anwar Saif
Department of Information Systems
Sana'a University ,Yemen
Email: anwarsaif.ye@gmail.com

Abstract

Cloud computing is a promising Information Technology(IT)paradigm which enables the internet's evolution into a global market of collaborating services. However, providers often offer their own proprietary applications, interfaces, APIs and infrastructures, resulting in a heterogeneous cloud environment. This heterogeneous environment lacks of a standardization that makes it difficult for users to promote to other cloud service providers. Many approaches tried to fill in the gab of portability and interoperability by providing semantic description for the cloud services. In this work we have made a comparative study between three approaches that introduce cloud services description based on semantic representation. This study will make a clear view of the advantages and disadvantages of these selected approaches.

Keywords: Cloud computing; Cloud Portability, Cloud Interoperability, Semantic Web

1. Introduction

Recently, cloud computing has become one of the most important evolutions in computer science. The National Institute of Standards and Technology (NIST)[1] defines cloud computing as: "cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction". The basic idea behind cloud computing is to make computing, both location and device independent. This enables a computing task, or data to be available at anytime, anywhere, and on any device can connect to a cloud service.

Cloud computing offers three types of service models namely Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS) and Software-as-a-Service (SaaS). In the case of IaaS, the cloud provides

processing, storage, memory, and other resources to the consumer. The user has no control over the underlying cloud resources but he controls the system resources, OS, and installed applications. In PaaS, the cloud allows a user to deploy applications however in this case he cannot control the underlying infrastructure or the operating

system. In SaaS, the user has limited control and is allowed to access only a fixed set of applications and services which are deployed by the service provider [2].These service models enable a user to utilize cloud resources according to his requirement without having to invest in deploying and managing resources on his own. Cloud computing has exponentially grown, with new cloud providers, platforms and services being offer in the IT market. Variety of cloud services, often providing non-uniform and incompatible interfaces, makes it hard for customers to decide how to develop, or even worse to migrate, their own application into the cloud. Lack of a shared standard for the description of such services can represent an obstacle to the development portability and interoperability of the service.

Interoperability refers to the ability of two or more systems or applications to exchange information and mutually use the information that has been exchanged, whereas portability refers to ability to move an entity from one system to another so that it is usable on the target system[3]. Interoperability and portability in cloud computing are illustrated in Figure 1.

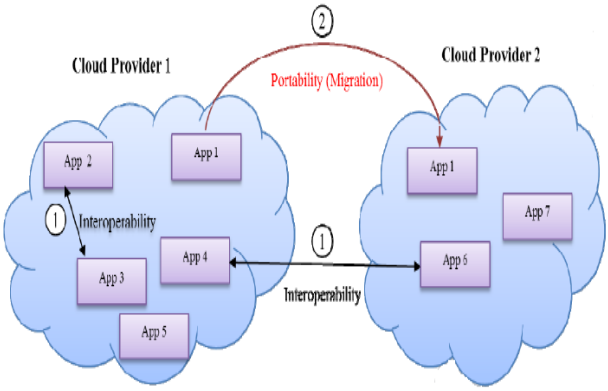


Figure 1. Interoperability and portability in cloud computing[4]

To overcome the shared standard obstacle, customers need a tool or technique to identify the architectural solution best answering their requests. In particular, an easily query-able categorisation of services to support the selection of the right cloud service, as well as a computable description of services enabling useful comparison and mapping among different provider's offers.

The semantic web coined by Tim Berners-Lee has experienced a growing attention from both academic and industrial areas with the objective of elevating the meaning of web information resources and making them easily readable by machines through certain key technologies[5]. Many technologies form the base of semantic technologies such technologies are Web Ontology Language (OWL), Web Ontology Language for Services (OWL-S), SPARQL Protocol And RDF Query Language (SPARQL) and Web Services Description Language (WSDL). OWL: is a semantic web language designed to represent knowledge about things, groups of entities and relations between them, whereas OWL-S: is designed to represent such kind of semantic information[6]. SPARQL allows to interrogate the knowledge base in order to dynamically discover equivalences among services or to compare resources and configurations and filter them according to a set of desired specifications.

Web Services Description Language (WSDL) is an XML format for describing network services as a set of endpoints operating on messages containing either document-oriented or procedure-oriented information[7].

Web semantic models are one of the contributory factors of solving interoperability and portability problems. These models are helpful in cloud computing in the following aspects:

- Functional and nonfunctional definitions, that is, the ability to define application functionalities and quality-of-service details in a platform-agnostic manner;
- Data modeling, including meta-data added through annotations pointing to generic operational models, which plays a key role in consolidating API's descriptions;
- Service description enhancement, in particular regarding service interfaces that differ between vendors even if the operation's semantics are similar.

The rest of the paper is structured as follows: Section 2 states the motivation; while section 3 addresses the comparative study. Section 4 presents the conclusion and future work.

2. Motivation

Cloud computing is having an enormous impact on how organizations manage their information technology resources, the current cloud computing landscape consists of a diverse set of products and services that range from infrastructure services (IaaS), to software service (SaaS). The variety of cloud services has led to proprietary architectures and technologies being used by vendors, increasing the risk of vendor lock-in for customers. This motives us to make highlights on the metrics that classify the cloud services description.

3.Comparative study

In this section we will explore the role of semantic web in cloud portability and interoperability from a wide range of literatures. In this section various approaches are screened and a detailed comparison between three of these approaches will be made.

The cloud service landscape is growing up very rapidly and there are different aspects of this evolution that need to be systematised in a formal way. Different compendiums have been written to try to describe the main cloud computing features and platforms [2] [8] [9]. Owing to the amount of heterogeneity, the task of manually compare functionalities and characteristics (from cloud service and appliance providers' websites and documentation) is not easy. Several classification attempts have been made in order to categories cloud services according to different criteria, like exposed functionalities or service models.

The OpenCrowd [10] consortium has proposed an online, freely navigable taxonomy, which categorises cloud services according to both their service model (IaaS, PaaS or SaaS) and application context. However, this taxonomy doesn't report the criteria followed to insert each service under a certain category. Indeed for this reason a lot of ontologies related to cloud computing were emerged. In [11] the authors identified four main categories of cloud computing ontologies according to their scopes: cloud resources and services description, cloud security, cloud interoperability and cloud services discovery and selection. They have presented a notable example of cloud service discovery system based on matchmaking. In the presented system the users can identify the cloud services required by means of three kinds of requirements: functional requirements (like programming language for PaaS service type), technical requirements (like CPU clock or RAM for IaaS service type) and cost requirements (like maximum price) as input parameters[12].

In context of interoperability at PaaS level [13] introduced a PaaS semantic interoperability framework, enabling heterogeneous cloud PaaS

systems and their offerings to overcome the semantic incompatibilities. It aims at resolving semantic interoperability conflicts raised during the deployment or migration of an application by defining three dimensions: PaaS entities, types of semantics, and levels of semantic conflicts.

We will study in detail these approaches[14] [15] [16]. The approach presented in [14] addresses semantic description of cloud resources, comprehending both a general and agnostic representation of such services and a focus on provider specific solutions, whereas the approach presented in [15] proposed a uniform, integrated, machine-readable, semantic representation of cloud services, patterns, appliances and their compositions. The authors in [16] proposed a scalable architecture for a semantic inter-cloud information exchange.

To achieve the interoperability and portability[14] proposed a semantic-based model consisting of three layers, each layer has been formalized through a group of ontologies. It is formalized by means of OWL, OWLS and SPARQL languages and inherits most of the elements from several initiatives such as NIST and Open Cloud Computing Interface (OCCI). The primary objective is to develop a common re- presentation base, formalised through a set of interrelated ontologies which can be used to annotate existing cloud services, together with the operations they expose and the parameters they exchange.

Figure 2 shows an accurate overview of the entire knowledge base, including the composing ontologies and their connections. The knowledge base is organised in three layers:

- The upper layer contains the agnostic service description ontology.
- The middle layer contains the cloud services categorisation ontology.
- The bottom layer contains two different groups of ontologies: cloud Provider ontology and cloud services OWL-S description.

Figure 2. Knowledge base[14]

In order to allow users to easily query the knowledge base and retrieve complete information about the desired services, the different ontologies have been connected through a set of properties. Figure 3 shows relationships in the knowledge base.

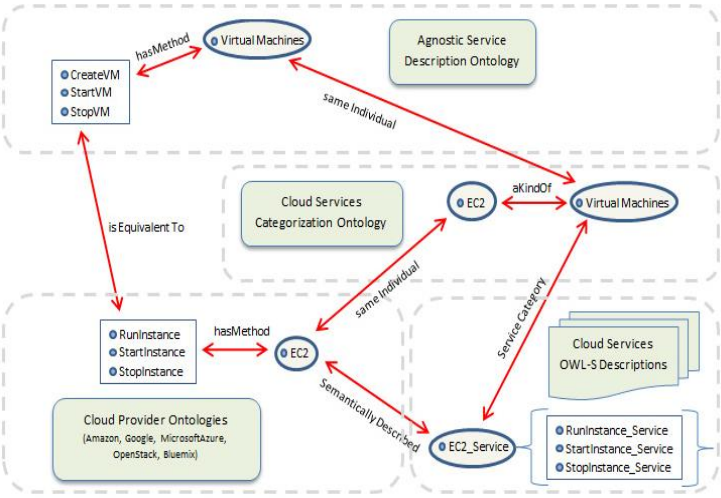


Figure 3. Relationships in the knowledge base[14]

In cloud ontology structure of these ontologies Figure 4

shows an agnostic service description ontology OWL structure which identifies abstract services, resources, methods and parameters which can be

used as a base to represent cloud provider specific concepts. This ontology is organised according to the same structure followed by the provider ontologies .

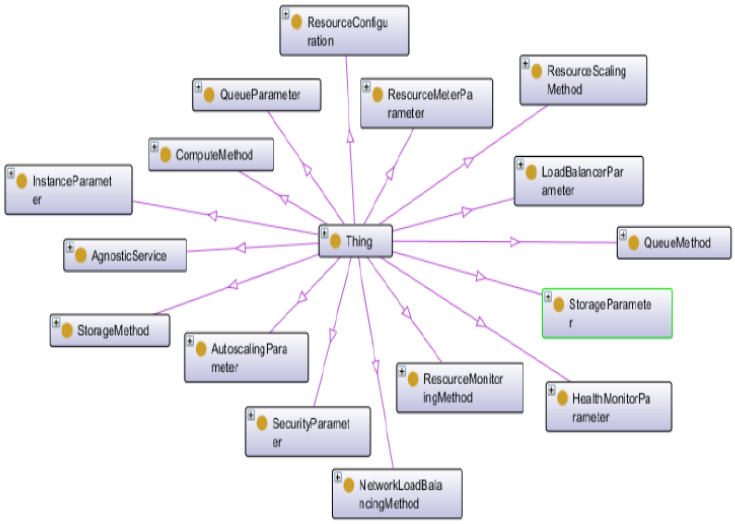


Figure 4. Agnostic service description ontology OWL structure[14]

The core of the categorisation ontology is represented by the service category class. Figure 5 shows a cloud service categorisation ontology OWL structure.

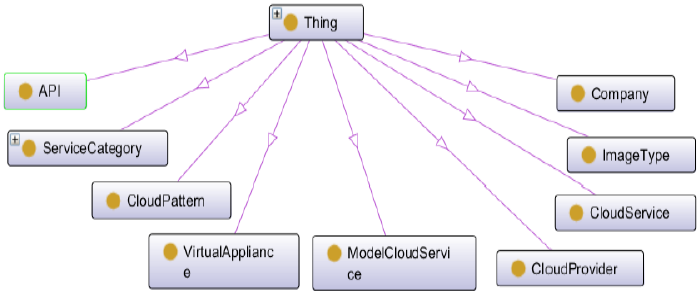


Figure 5. Cloud service categorisation ontology OWL structure[14]

In this approach the authors introduced a semantic based approach to describe cloud services, resources and appliances, together with their related operations and parameters. In particular, the authors leverage a multilayer ontological architecture through which were able to describe both agnostic and platform dependent concepts that are used to compare cloud platform offers. The main objective of such approach is to support the automatic

resolution of issues related to the portability and interoperability of applications from legacy systems to the cloud and/or across multiple cloud platforms. By using SPARQL queries, it is possible to automatise the service selection/replacement processes involved in the porting of the cloud or migrating it from a platform to another. However the limitation in this approach becomes obvious for being not able to completely avoid human intervention into the process, especially when we want to dynamically swap cloud services.

In context of portability and interoperability at SaaS level [15] presented a semantic representation consisting of a set of interrelated OWL ontologies describing cloud services, patterns, virtual appliances and their compositions. An automated reasoning by means of SPARQL queries, Semantic Web Rule Language (SWRL) rules and Description Logic (DL) is exploited to enrich this representation in order to assess the equivalence between services, appliances and patterns. The overall approach is a graphical based representation, structured into five conceptual layers. The graph represents concepts (graph nodes) and relationship (graph edges) at different levels. In each level relationships among concepts of the same level are represented in addition to inter-level relationships.

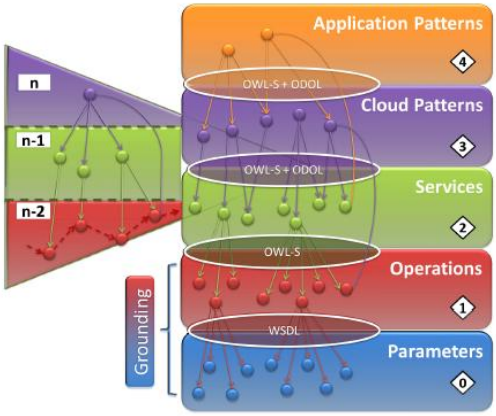


Figure 6. The Conceptual Layers[15]

The five conceptual levels are shown in figure 6 and listed below:

- Parameters Level.
- Operations Level.
- Services Level.
- Cloud Patterns Level.

- Application Patterns Level.

The semantic representation of cloud patterns, cloud services and virtual appliances supports the development of portable and potentially interoperable cloud solutions as well as p orting of existing applications to multiple clouds. This formalization enables the comparison of solutions provided by different vendors, despite of possible differences in their semantics, as long as they refer to a common shared representation. The use of inference rules enables the enrichment of the knowledge base each time a new cloud provider's offer is included. This approach supports portability since it starts from a non-cloud description of an application and drives the developer to a cloud service. It is potentially enables interoperability, by enabling the composition of cloud patterns composed of services offered by multiple providers. The limitation of this approach comes from the inability to support IaaS and PaaS cloud models and the necessity to manually write SPARQL queries to retrieve the needed information.

To achieve the interoperability at IaaS level, the authors in [16] proposed a scalable architecture for a semantic inter-cloud information exchange. It is based on an Inter-cloud ontology to semantically describe entities within the inter-cloud federated environment. In this approach SPARQL queries are executed in order to determine resource configurations and exchanges. Inter-cloud architecture is shown in Figure 7. It includes clouds, roots, exchanges (which mediate governance between clouds), and gateways (which mediate data exchange between clouds). Functional elements of the architecture include name spaces, presence, messaging, resource ontologies, and trust infrastructure.

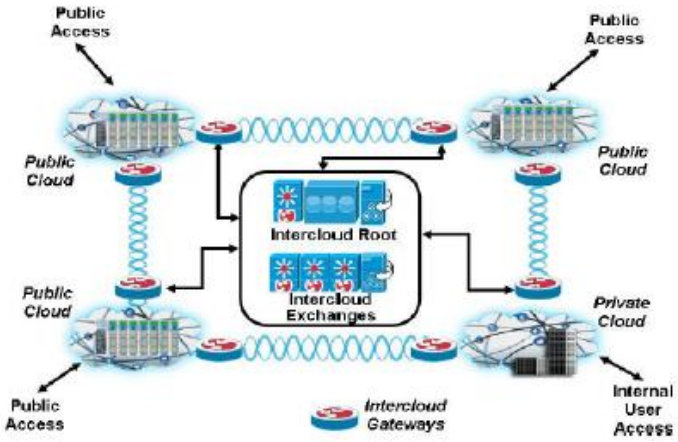


Figure 7. IEEE P2302, Inter-cloud Architecture[16]

Inter-cloud ontology used to semantically describe entities, within the inter-cloud federated environment. It represents the knowledge base on which SPARQL queries are executed in order to determine resource configurations and exchanges, according to consumers needs and resources availability. Everything in the inter-cloud environment can be seen as a resource, which can be accessed and consumed through an Application Programming Interface (API). Thus, the main concept of the inter-cloud ontology is represented by the resource class. Figure 8 shows a snapshot of the ontology tree. At the second level of the ontology, three classes are existed: the cloud resource class, the group class, and the interface class. The cloud resource class describes resources related to a cloud computing environment. Resources describing the current ontology are identified by three subclasses; compute, network and storage. The group class identifies the two main groups of resources which will be modeled by their ontology, here they are represented by the infrastructure and federation subclasses. The interface class is introduced to associate specific access methods to a group individual. Using these classes and relations makes it possible to describe infrastructure resources independent of the providers and also the resources offered by specific providers. The authors have shown that there is a demand for federated and interoperable clouds. In this regard, in particular, a scalable and semantic based information exchange is needed. The outlined IEEE P2302 Inter-cloud architectures take these requirements into account. However this approach is partial approach.

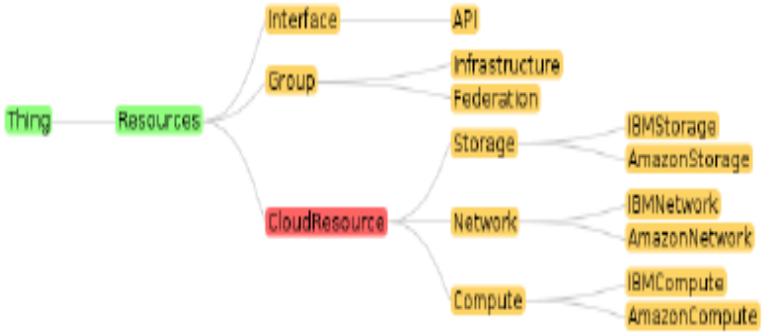


Figure 8. IEEE P2302, snapshot of the ontology organization[16]

To make a clear view of the differences between these approaches, we have made a comparison based on seven selected metrics as shown in Table 1. The semantic technology metric illustrates the technologies used in each approach. The completeness degree is satisfied in [14] and [15] approaches while it is partial in the [16] approach. The dependence metric illustrate whether these models are built from scratch or inspired from the already defined standards. The approach [14] considers the three cloud models while [15] and [16] approaches consider only SaaS and IaaS, respectively. The metric cloud service metric is used to assess the efficiency of each approach. The approaches are validated through cloud services (CS) or simulation tools (ST). The last metric shows the interoperability and portability support for each approach.

As specified in Table 1, it is reassuring to observe that all of the approaches have formalized their models in terms of semantic technology. .

Table 1. Illustrates the differences between the three approaches.

Metric	[14]	[15]	[16]
Semantic technology	OWL, SPARQL	OWL, SPARQL, DL, SWRL	OWL, SPARQL, SWRL
Development tools	N/A	N/A	Java JEEE, Jena
Completeness degree	Complete	Complete	Partial
Dependency on other models	OWL-S OCCI	ODOL WSDL OWL-S	Native

Cloud model	SaaS , PaaS, IaaS	SaaS	IaaS
Cloud services	CS:Windows Azure ,AWS,Open-Stack, etc.	CS: Windows Azure ,AWS	None
supports interoperability and portability	Both	Both	Interoperability

4. Conclusion and future work

The addressed approaches described how the semantic technologies can be used to fully represent description of cloud services. This representation enables users to easy select the cloud providers, and resolve the cloud interoperability and portability obstacles. Moreover, it gives the customer the ability to use incompatible services offered by different vendors at the same time. The comparative study addressed three of these approaches and highlights the major differences between them. In the future we are planning to come up with a new framework for semantic representation for the cloud services.

References

[1] P. Mell and T. Grance, "The NIST definition of cloud computing," 2011.

[2] C. V. Rajkumar Buyya and S. ThamaraiSelvi, "Mastering Cloud Computing," ed: McGraw Hill Education, 2013.

[3] B. Di Martino, G. Cretella, and A. Esposito, "Cloud Portability and Interoperability," in *Cloud Portability and Interoperability*, ed: Springer, 2015, pp. 1-14.

[4] R. Rezaei, T. K. Chiew, S. P. Lee, and Z. S. Aliee, "A semantic interoperability framework for software as a service systems in cloud computing environments," *Expert Systems with Applications*, vol. 41, pp. 5751-5770, 2014.

[5] P. Hitzler, M. Krotzsch, and S. Rudolph, *Foundations of semantic web technologies*: CRC Press, 2009.

[6] M. Dean, G. Schreiber, S. Bechhofer, F. van Harmelen, J. Hendler, I. Horrocks, *et al.*, "OWL Web Ontology Language Reference. W3C Recommendation (February 10, 2004)," ed, 2004.

[7] E. Christensen, F. Curbera, G. Meredith, and S. Weerawarana, "Web services description language (WSDL) 1.1," ed: Citeseer, 2001.

[8] D. Puthal, B. Sahoo, S. Mishra, and S. Swain, "Cloud computing features, issues, and challenges: a big picture," in *Computational*

Intelligence and Networks (CINE), 2015 International Conference on, 2015, pp. 116-123.

[9] F. F. Moghaddam, M. Ahmadi, S. Sarvari, M. Eslami, and A. Golkar, "Cloud computing challenges and opportunities: A survey," in *Telematics and Future Generation Networks (TAFGEN), 2015 1st International Conference on*, 2015, pp. 34-38.

[10] Available: http://cloudtaxonomy.opencrowd.com

[11] D. Androcec, N. Vrcek, and J. Seva, "Cloud computing ontologies: A systematic review," in *Proceedings of the third international conference on models and ontology-based design of protocols, architectures and services*, 2012, pp. 9-14.

[12] T. Han and K. M. Sim, "An ontology-enhanced cloud service discovery system," in *Proceedings of the International MultiConference of Engineers and Computer Scientists*, 2010, pp. 17-19.

[13] N. Loutas, E. Kamateri, and K. Tarabanis, "A semantic interoperability framework for cloud platform as a service," in *Cloud Computing Technology and Science (CloudCom), 2011 IEEE Third International Conference on*, 2011, pp. 280-287.

[14] B. D. Martino, G. Cretella, A. Esposito, and G. Carta, "An OWL ontology to support cloud portability and interoperability," *International Journal of Web and Grid Services*, vol. 11, pp. 303-326, 2015.

[15] B. Di Martino, A. Esposito, and G. Cretella, "Semantic representation of cloud patterns and services with automated reasoning to support cloud application portability," 2015.

[16] B. Di Martino, G. Cretella, A. Esposito, A. Willner, A. Alloush, D. Bernstein, *et al.*, "Towards an Ontology-Based Intercloud Resource Catalogue--The IEEE P2302 Intercloud Approach for a Semantic Resource Exchange," in *Cloud Engineering (IC2E), 2015 IEEE International Conference on*, 2015, pp. 458-464.

Development Challenges of the Internet of Things in Yemen

Dr. Mansoor Noman Marhoob Ali
Computer Science Department
College of Education, Arts and Sciences - Marib
Sana'a University, Sana'a -Yemen
amnm@mail.ru

Dr. Abdualmajed Ahmed Ghaleb Al-Khulaidi
Computer Science Department
College of Education, Arts and Sciences - Marib
Sana'a University, Sana'a -Yemen
alkhulaidi @mail.ru

Abstract

The Internet of Things (IoT) promises to offer a revolutionary, fully connected "smart" world. Yet the challenges associated with IoT need to be considered and addressed for the benefit of individuals, society, and the economy to be realized.

The broad scope of IoT challenges will not be unique to industrialized countries, but also developing regions will need to respond and realize the potential benefits of IoT. In addition, the unique needs and challenges of implementation in the underdeveloped regions will need to be addressed, including infrastructure readiness, technical skill requirements, and policy resources.

This paper presents development Challenges of the IoT in Yemen such as availability of internet at everywhere and at reasonable cost, shortage of IPv4 addresses and transition to IPv6, Power supply for sensors, acceptability among the society, and Security and privacy protection.

Keywords: IoT, IPv4, IPv6, NAT, 6LoWPAN, challenges.

1. Introduction

IoT is a new step in technological progress, new stage for the development of the Internet [1], and gives us great opportunities to be connected with the virtual world.

'IoT allows people and things to be connected anytime, anyplace with anything and anyone, ideally using any path/network and any service. [2]

IoT enhances the collection, analysis, and distribution of data, which can be transformed into all kinds of receivers.

The main reasons behind this interest of IoT are the capabilities that it offers. It promises to create the world around us where everything in this

world will be connected to each other via Internet and to communicate with each other with a minimum human intervention [1].

In recent time, an enormous amount of research and development works are carried out in different parts of the world to make IoT to be feasible. The main application domains for the IoT illustrated in Figure 1.

Cisco estimates the IoT will consist of 50 billion devices connected to the Internet by 2020.

The aim of IoT is to make our daily life easier and smarter and to be applied in all areas of life; such as industry, environmental monitoring, healthcare, military, and cities management. Imagine the world where billions of objects can communicate and share information. Data regularly collected, analyzed and used to initiate action, providing intelligence for planning, management, and decision-making. This is the world of the IoT, however, in Yemen, its scope develops slowly.

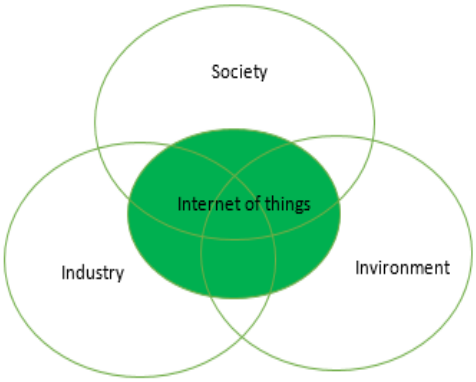


Figure 1. Internet of things application domains

The development of IoT is a step-by-step process, IPv6 address space to make it possible to connect unlimited number of devices with Internet [2].

There are a lot of difficulties in the field of IoT such as the problem of insufficient security, the less-protection of personal information, the complexity of communication, the variety of protocols, the lack of common standards, issues of power supply of connected devices, and the acceptability among the society.

Most of these difficulties are common for all countries. This paper presents the most important challenges facing the IoT in Yemen now, the main challenges of IoT in Yemen can be summarized as availability of internet at everywhere and at reasonable cost, shortage of IPv4 addresses and transition to IPv6, Power supply for sensors, acceptability among the society, security and privacy protection.

2. The Main Development Challenges of IoT in Yemen now

2.1. Availability of Internet at Everywhere and at Reasonable Cost

The success of the IoT depends on the availability of the Internet and its spread at everywhere [3], in addition, the Internet should be available with a reasonable cost, and this problem may not be very easy to be solved in Yemen.

According to [4] the number of Internet users in Yemen is still limited with reference to the population as not to exceed the proportion 25% of the total population. Over 75% of population in Yemen still unable to use internet (Figure 2).

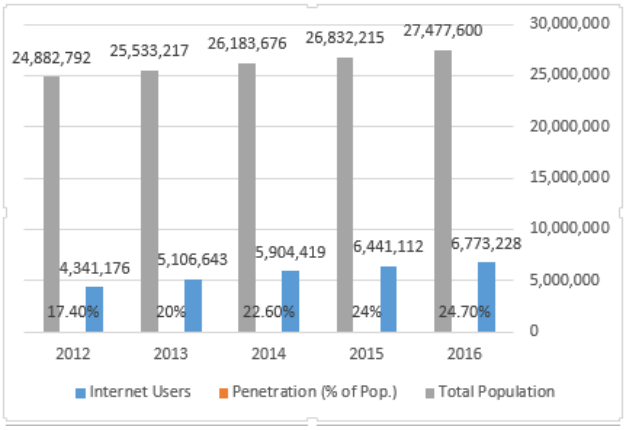


Figure 2. Internet users and population statistics in Yemen

In order to make internet available, there is a need of huge investment of providing infrastructure and resources.

2.2. Shortage of IPv4 addresses, and transition to IPv6

Currently in Yemen, the internet protocol (IPv4), IPv4 uses 32-bit (4-byte) addresses, IPv4 has only 4.3 billion (2^{32}) possible IP-addresses, IPv4 has been completely depleted by Feb 2011 [5]. Now in Yemen the problem of shortage of IPv4 addresses is actual, as solution for this problem, providers use Network Address Translation (NAT), but it is not good solution. When hiding behind a NAT too many devices and sites may decide that the load on them created by robots, and take measures to protect from such actions and can completely block the IP-address, or increase the intensity of the show captcha. Therefore, it will decrease the quality of access to the Internet for users. A similar problem exists for mail systems: the mails sent by the user in the NAT could be mistakenly classified as spam and did not reach the addressee.

The development of the IoT requires many new sensors with new unique IP-addresses, so, the transition to the IPv6 is very necessary, IPv6 uses 128-bit (16-byte) addresses, so the new address space supports approximately 340 undecillion (2^{128}) addresses. Currently the Ministry of Communications and Information Technology is working for the transition to IPv6. IPv6 simplifies network management by using automatic configuration, also provides new features enabling an easier configuration of devices, data streaming compliance, improved security, and effective peer-to-peer connections avoiding NAT barriers [5]. IPv6 has been designed to provide secure communications for users and mobility for all devices. In areas of network protocol security, IPv6 contains addressing and security control information, i.e., IPSec to route packets through the Internet. In IPv4, IPSec is optional. With IPv6, IPSec support is integrated into the protocol design and connections can be secured when communicating with other IPv6 devices. IPSec provides data confidentiality, data integrity and data authentication at the network layer, and offers various security services at the IP layer and above. IPv6 is the most suitable technology for the IoT, since it offers scalability, flexibility, tested, extended, ubiquitous, open, and end-to-end connectivity [5].Most sensor networks using wireless communications networks such as IPv6 Low Power Wireless Personal Area Networks (6LoWPAN) [6], Bluetooth, Wi-Fi, WiMAX, 3G and 4G networks, and a satellite network (e.g., GPS). Sensor networks in IoT use communication protocols based on IP (e.g., IPv6). "IPv6-based solutions are key enablers for the success of the IoT interoperability, acceptance and integration" [5]. 6LoWPAN is a key component in order to realize the IP-based integration of constrained devices, it is used in a multitude of projects, exploring a wide range of use cases such as smart buildings, smart environments, smart cities, etc. In all cases, constrained devices, forming 6LoWPAN networks, are used to collect information from the real world and this information is used to generate intelligence and make the world around us smarter [7]. 6LoWPAN is a communication standard that allows the low-power devices to communicate and exchange data via IPv6. There are many benefits of using IP-based connectivity to form the sensor access network:

- IP connects easily to other IP networks without the need for translation gateways or proxies.
- IP networks allow the use of existing network infrastructure.
- IP is open and free, with standards, process and documents available to anyone.

6LoWPAN works on the IPv6 protocol suite based on IEEE 802.15.4 standard. Hence, it has the characteristics of low-cost, low-rate and low-power deployment.

2.3. Power Supply for Sensors

Power consumption has the greatest challenge for sensors. Today sensors need to be able to sustain longer battery lifespan, especially in cases such as outdoor deployments, to shorten hardware maintenance and prevent breakdown of communication.

The most important factors for IoT power sources are:

- Wireless, smart self-charring capability.
- Environmentally friendly and cost-effective materials.
- Flexible shape and small size.
- Enhancement in energy and power density.
- Longer lifetime.

Sensors must operate fully autonomously, to fully realize, potential of IoT. Batteries are uses to supply the necessary energy required for sensor signal processing and communication. So there is a need to use environmentally friendly new types of rechargeable batteries. The rechargeable batteries will need to be charged from different renewable energy sources, wind and solar. The choice of wind energy may not be limited for sensor nodes due to its initial investment and availability of wind.

The development of new, efficient, and compact batteries, as well as new energy generation devices or energy harvesting will be the main power source for IoT devices [8], "Energy harvesting technologies [9] use power generating elements such as solar cells, piezoelectric elements, and thermoelectric elements to convert light, vibration, and heat energy into electricity, then use that electricity efficiently." These technologies will help advance the growth of the IoT.

Yemen is not industrialized country, but it must consider these requirements when importing IoT devices, especially that clean energy sources available in Yemen significantly and that should be exploited and utilized.

2.4. Acceptability among the Society

Most consumers are not ready to use the IoT technology because they do not have enough knowledge about it. Most companies in Yemen consider that the technology is not mature, the infrastructure of companies is not adequate, and implementing of IoT may be expensive.

There are also many concerns related to security and privacy [10], some applications handle sensitive information about people, such as their location and movements, or their health and purchasing preferences. Confidence in and acceptance of IoT will depend on the protection it provides to people's privacy and the levels of security it guarantees to systems and processes. The main reasons in Yemen are the low level of technical skills and high illiteracy rate among the society.

2.5. Security and Privacy Protection

IoT raises important questions and introduces new challenges for the security of systems, processes and the privacy of individuals. Privacy should be protected in the device, in storage during communication and at processing which helps to disclose the sensitive information.

To assure data confidentiality during the transmission of data, the most common approach is encryption.

Cryptographic and stenographic security measures in the information exchange process will result in a more secure and robust IoT infrastructure.

To secure the information exchange in the IoT existing encryption technology needs to be carefully reviewed [11], the encryption algorithms need to be faster and less energy consuming.

Security at both the device and network is critical to the operation of IoT.

This does not require the research for a solution that does not yet exist or proposing a revolutionary approach to security, but rather an evolution of measures that have proven successful in IT networks, adapted to the challenges of IoT and to the constraints of connected devices.

For privacy, the situation is more serious, the heterogeneity and mobility of 'things' in the IoT will add complexity to the situation, moreover, privacy-preserving technology is still in early stages. In the future, new standards and

technologies should address security and privacy features for users, network, data and applications.

3. Conclusion

This paper presented, the IoT as a technology that will change the world to make it smarter, and the development challenges of the IoT in Yemen such as availability of internet at everywhere and at reasonable cost, shortage of IPv4 addresses and transition to IPv6, Power supply for sensors, acceptability among the society, and Security and privacy protection. The transition to the IPv6 is very important step for development the IoT in Yemen now.

Security and privacy are probably the most challenging issues in the IoT, and they are very important topics for research in the field of IoT.

4. References

[1] C. Perera, A. Zaslavsky, P. Christen and a. D. Georgakopoulos, "Context Aware Computing for The Internet of Things: A Survey," *IEEE Communications Surveys & Tutorials, vol. 16, no. 1,* pp. 414-454, 2014.

[2] A. Jara, D. Fernandez, P. Lopez, M. Zamora and a. A. Skarmeta, "Lightweight MIPv6 with IPSec Support," *Mobile Information Systems, vol. 10, no. 1,* pp. 37-77, 2014.

[3] S. C. Mukhopadhyay, Internet of Things Challenges and Opportunities, Switzerland: Springer International Publishing, 2014.

[4] "Internet Live Stats - Internet Usage & Social Media Statistics," [Online]. Available: http://www.internetlivestats.com. [Accessed 27 Aug 2016].

[5] S. Ziegler, C. Crettaz, L. Ladid, S. Krco, B. Pokric, A. Skarmeta, A. Jara, W. Kastner and a. M. Jung, "IoT6 – Moving to an IPv6-Based Future IoT," in *The Future Internet*, Berlin Heidelberg, Springer Berlin Heidelberg, 2013, p. 161–172.

[6] A. Jara, Y. Bocchi and D. Genoud, "Mobility management in bluetooth low energy," in *IEEE 11th Consumer Communications and Networking Conference (CCNC)*, Las Vegas, 2014.

[7] I. Ishaq, D. Carels, G. Teklemariam, J. Hoebeke, F. Abeele, E. Poorter, I. Moerman and P. Demeester, "IETF Standardization in

the Field of the Internet of Things (IoT): A Survey," *JSAN, vol. 2, no. 2,* pp. 235-287, 2013.

[8] S. Chen, H. Xu, D. Liu, B. Hu and a. H. Wang, "A Vision of IoT: Applications, Challenges, and Opportunities With China Perspective," *IEEE Internet of Things Journal, vol. 1, no.4,* pp. 349-359, 2014.

[9] K. Sekine, "Energy-harvesting devices replace batteries in IoT sensors," 2016. [Online]. Available: http://core.spansion.com/article/energy-harvesting-devices-replace-batteries-in-iot-sensors/#.V7NXr1sr.. [Accessed 27 Aug 2016].

[10] L. Tan and a. N. Wang, "Future internet: The Internet of Things," in *3rd International Conference on Advanced Computer Theory and Engineering(ICACTE),* Chengdu, 2010.

[11] B. Aditya, P. Kumar and a. A. Doegar, "Internet of Things (IoT): IPv6 Implementation for Smart Vision of the World," *International Journal of Advance Foundation And Research In Science & Engineering (IJAFRSE) Vol. 1,* 2015.

[12] A. J. Jara, L. Ladid and a. A. Skarmeta, " The Internet of Everything through IPv6:An Analysis of Challenges, Solutions and Opportunities," [Online].Available: http://ipv6forum.com/iot/images/jowua-v4n3-6.pdf. [Accessed 28 Aug 2016].

Designing Mathematical Model and Stochastic Hybrid Random Method for a Cache System to Create a Hybrid Cache Algorithm for Web Proxy to Increase the Speed of Accessing and Processing Data.

Dr. Abdualmajed Ahmed Ghaleb Al-Khulaidi
Computer science department
College of Education, Arts and Sciences - Marib
Sana'a University, Sana'a -Yemen
alkhulaidi @mail.ru

Dr. Mansoor Noman Marhoob Ali
Computer science department
College of Education, Arts and Sciences - Marib
Sana'a University, Sana'a -Yemen
amnm@mail.ru

Abstract

This scientific article contains a mathematical model design and stochastic hybrid random method design for a cache system to get a hybrid cash algorithm to increase the speed of accessing, processing and saving data.. By this method , we can hybrid any kind of algorithms with each other. The simplicity of this method in hybridization is not only exclusive to web server but can be also used with databases, information systems, operating systems and CPU. Experiments were carried out to measure performance proficiency of hybrid cache algorithms by helping trace-driven simulation by comparing between our hybrid cache algorithm with famous hybrid cache algorithms(LRFU). Our hybrid algorithm contains three algorithms(GDS,LRU,LFU).Our hybrid algorithm was used in squid proxy that works with operation system Linux . This proxy server checks the data by the famous hybrid algorithm (LRFU) and our hybrid designed algorithm. The results of our hybrid designed algorithm indicate that Hit Ratio(HR) was 22% and Byte Hit Ratio (BHR) was 18% compared with famous hybrid algorithm(LRFU). This indicates that our hybrid cache designed algorithm is more efficient than famous algorithm(LRFU).

Keywords: Cache, Hit Ratio, Byte Hit Ratio ,Web Cache.

1. Introduction

When processing a big amount of data, a problem shows up, which is the slow access speed[1]. Caching is taking the data from the memory and saving it in the small cache which is much faster at delivering the data. In the contemporary time, raising performance and efficiency through cache algorithms has become an interested research topic, some scientists have committed some researches in this topic like.[2],[3],[4]: Sharieh A.,Castro M.,Dasarathan D., Chen Y.C., …etc. Hit ratio is a measurement for the cache's efficiency, it is referred to the availability of the necessary commands and data in the cache when the CPU directly searches for it, and when the CPU finds what it is looking for it is called a "hit", but when it does not find it it's called a "miss". In most modern devices the hit ratio nears 90%, this shows that the cache makes it faster. The access time to the memory in the computer can

be improved which leads to a higher hit ratio, this means that the CPU will be retrieving what it needs from the cache directly rather than accessing the main memory, and this suggests that we need better cache algorithms [8] ,[9] ,[10] . In computing, a cache is a hardware or software component that stores data that can be served faster; the data stored in a cache might be the result of an earlier computation, or the duplicate of data stored elsewhere. A cache hit occurs when the requested data can be found in a cache, while a cache miss occurs when it cannot. Cache hits are served by reading data from the cache, which is faster than recomputing a result or reading from a slower data store; thus, the more requests can be served from the cache, the faster the system performs.

1.1.Cache Memory

Cache memory can be called CPU memory. This random access memory (RAM) that a computer microprocessor can access more quickly than regular RAM [11], [13], [17], [18]. This memory is typically integrated directly with the CPU chip or placed on a separate chip that has a separate bus interconnect with the CPU. The cache contains copies of the RAM, when the CPU reads a word from the RAM it first starts to look for it in the cache or the cache memory and if it can be found it can be processed, but if it does not find it then it reads a block or a batch of words which gets written in the cache then the processing starts. The basic purpose of cache memory is to store program instructions that are frequently re-referenced by software during operations.

1.2.Web Proxy

Proxy is referred to a web agent, it consists of many devices and software that works in the network field, it works as an agent or as a middle man between the user and the other internet web servers and services and one of it's jobs is temporary caching of repeated requests from the user in the present time[1],[2]. The proxy is considered as special type of servers where it receives and processes the users' requests but it does not contain independent information rather bits and parts of the real servers' contents and acts as an agent for the real servers to display their contents.

1.2.1.Proxy Server Duties

As stated before, there are many types of proxies and they work in either of the following ways[20]:

1. Saving the most repeated requests and updating the currently saved contents, then delivering the cached contents to other users thus increasing the speed of replies and reducing traffic in the network.
2. Works as a firewall and traffic filter and manager the network since all the incoming or outgoing data of the network go through it.

The disadvantage of proxies is that sometimes the users who use the proxy will receive outdated data since the proxy replies to their requests using previously stored contents when the real contents have actually been dynamically altered in the real server without the proxy's knowledge; Some services in the network requests special proxies to reach the required real data which forces the local network to utilize other proxies in other networks to reach the desired data. Caching is a collection of frequently requested data by the users and it is stored locally, thus allowing a faster access to them by the users; it is seen as a small group of data for huge amount of information stored in the storage servers. A cache is always physically close from the network internet user, today, cache systems are available in many modern personal computers and servers, many modern web browsers store frequently visited web pages by the user in the memory or other storage devices which allows other users to load those pages faster later on when he visits them. Web caching is a collection of locally saved web pages including the images, videos, texts and other types of documents that can be accessed using the HTT Protocol [15]. The web browser makes sure to prolong, maintain and update the web cache. The web cache works as a program installed in the local private servers that archives and recalling the frequently requested data by the user. If the requested documents are found by the proxy in the local cache then it retrieves it and registers a 'hit', however if the requested documents were not found then proxy registers a 'miss' and requests the documents from the real servers and transports it to the requested user[5]. Through the action of retrieving the data from a local server we get a faster response and reduce the time in the network and also increase the view range for every user. The proxy server allows the control of the contents being transferred so that bad pages would not be sent to the requester, it also frees space for new documents by deleting old and unused documents, this process occurs dynamically through using the modern LRU 'least recently used' algorithm. Web caching is the temporary storage of remote web objects on a local server. Advantages of this technique range from reduced access latencies to reduced server load and bandwidth consumption[16].

2. Algorithms Existing Cache

In computing, cache algorithms (also frequently called cache replacement algorithms or cache replacement policies) are optimizing algorithms — that a computer program or a hardware-maintained structure can follow in order to manage a cache of information stored on computer. When the cache is full, the

algorithm must choose which items to be discarded to make a room for the new ones[16],[19],20].

2.1.Belady's Algorithm

The most efficient caching algorithm would be to discard the information that will not be needed for the longest time in the future. This optimal result is referred to as Belady's optimal algorithm or the clairvoyant algorithm. Since it is generally impossible to predict how far in the future information will be needed, generally this is not implementable in practice.

2.2.Least Recently Used (LRU)

This cache algorithm keeps recently used items near the top of cache. Whenever a new item is accessed, the LRU places it at the top of the cache[4]. When the cache limit has been reached, items that have been accessed less recently will be removed starting from the bottom of the cache. This can be an expensive algorithm to use, as it needs to keep "age bits" that show exactly when the item was accessed. In addition, when a LRU cache algorithm deletes an item, the "age bit" makes changes on all the other items.

2.3.Most Recently Used (MRU)

This cache algorithm removes the most recently used items [6]. A MRU algorithm is good in situations in which the older an item is, the more likely it is to be accessed. Discards, in contrast to LRU, the most recently used items first.

2.4.Pseudo-LRU (PLRU)

For CPU caches with large associativity (generally >4 ways), the implementation cost of LRU becomes prohibitive[4][7]. In many CPU caches, a scheme that almost always discards one of the least recently used items is sufficient. So many CPU designers choose a PLRU algorithm which only needs one bit per cache item to work. PLRU typically has a slightly worse miss ratio, has a slightly better latency, and uses slightly less power than LRU.

2.5.Random Replacement (RR)

Randomly selects a candidate item and discards it to make space when necessary. This algorithm does not require keeping any information about the access history. For its simplicity, it has been used in ARM processors[7]. It admits efficient stochastic simulation [8].

2.6.Least-Frequently Used (LFU)

Counts how often an item is needed. Those that are used least often are discarded first. This cache algorithm uses a counter to keep track of how often an entry is accessed. With the LFU cache algorithm, the entry with the lowest count is removed first. This method isn't used that often, as it does not account for an item that had an initially high access rate and then was not accessed for a long time.

2.7.Adaptive Replacement Cache (ARC)

Constantly balances between LRU and LFU, to improve the combined result[10] [21]. ARC improves on SLRU by using information about recently-evicted cache items to dynamically adjust the size of the protected segment and the probationary segment to make the best use of the available cache space. Developed at the IBM Almaden Research Center, this cache algorithm keeps track of both LFU and LRU, as well as evicted cache entries to get the best use out of the available cache.

2.8.Clock with Adaptive Replacement (CAR)

Combines Adaptive Replacement Cache (ARC) and CLOCK. CAR has performance comparable to ARC, and substantially outperforms both LRU and CLOCK. Like ARC, CAR is self-tuning and requires no user-specified magic parameters[6].

2.9.Multi Queue (MQ)

The MQ cache contains multiple LRU queues, Q0, Q1, ..., Qm-1. Blocks stay in the LRU queues for a given lifetime, which is defined dynamically by the MQ algorithm to be the maximum temporal distance between two accesses to the same file or the number of cache blocks, whichever is larger [9] 12]. If a block has not been referenced within its lifetime, it is demoted from to Qi to Qi−1 or evicted from the cache if it is in Q0. Each queue also has a maximum access count; if a block in queue Qi is accessed more than 2i times, this block is promoted to Qi+1 until it is accessed

more than 2i+1 times or its lifetime expires. Within a given queue, blocks are ranked by the recently of access, according to LRU.

2.10.Greedy Dual-Size (GDS) Algorithm

Greedy Dual-Size Algorithm incorporates locality with cost and size to decide which document to evict[14],[15]. This replacement policy can enhance hit ratio. Greedy Dual-Size algorithm takes into consideration factors such as locality, size and latency/cost. The algorithm is also a modified enhanced version of LRU. It is concerned with the case when pages in a cache have the same size, but incur different costs to fetch from a secondary storage. The algorithm associates a value, H, with each cached page p. Initially, when a page is brought into cache, H is set to be the cost of bringing the page into the cache (the cost is always non-negative). When a replacement needs to be made, the page with the lowest H value, minH, is replaced, and then all pages reduce their H values by minH . If a page is accessed, its H value is restored to the cost of bringing it into the cache. Thus, the H values of recently accessed pages retain a larger portion of the original cost than those of pages that have not been accessed for a long time. By reducing the H values as time goes on and restoring them upon access, the algorithm integrates the locality and cost concerns in a seamless fashion [17][18]. To incorporate the different sizes of the document, we extend the Greedy Dual algorithm by setting H to cost/size upon an access to a document, where cost is the cost of bringing the document, and size is the size of the document in bytes. We called the extended version the Greedy Dual-Size algorithm. The definition of cost depends on the goal of the replacement algorithm: cost is set to 1 if the goal is to maximize hit ratio, it is set to the downloading latency if the goal is to minimize average latency, and it is set to the network cost if the goal is to minimize the total cost. At the first glance, Greedy Dual-Size would require k subtractions when a replacement is made, where k is the number of documents in cache.

3. The Known Cache Algorithms Hybrid Method (LRFU)

To give an example of this, we have to look at the build of the hybrid algorithm that is made by fusing the LRU (Least recently used) algorithm and the Frequently used algorithm to make an LRFU algorithm[22]. In the LRFU algorithm for every object in the cache a rating is calculated for the combined recency and frequency (CRF). To calculate the recursion, CRF recursion, the following function is used:

$$F(x) = \left(\frac{1}{p}\right)^{\lambda x} \qquad (1)$$

Where $p \geq 2$ is a variable which is equals near 2.

λ= control variable.

Where : X between 2 of the neighboring elements that were called. The rating is calculated for every element and the least rated element gets deleted separately from the cache . In it has been proven that when $\lambda = 0$ then all of

the CRF results will be exactly the same as LFU.

The performance of the LRFU was studied based on the

trace driven simulation example by designing hybrid methods for caching proxy servers as a temporary storage for web pages, this makes it the future requests get a faster reply. The temporary cache memory in the browser stores websites' data when visiting them, when the web resources is visited or required again the browser loads it from the saved cache which is much faster than connecting to the real server and downloading it again, but the resources in the cache can be old and obsolete and needs to be deleted so new data will get cached. To delete the temporary cached data in Firefox you need to do the following:

Open the menu by clicking on the button, then choose 'History'.

1. Click on the 'Clear Recent History'.
2. Choose the data you want to delete from the form, set it to the start to delete everything and tick the 'Cache' option.
3. To commit the action press the 'Clear Now' button.

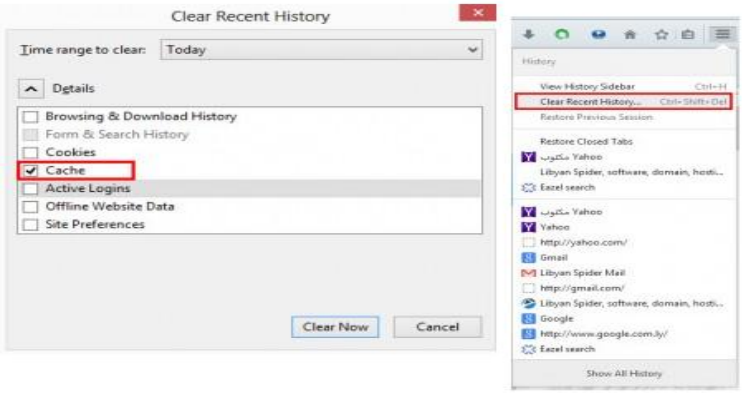


Figure.1. “Delete the temporary cached data in Firefox”

4. The Mathematical Model of Designing the Hybrid Cache Method for Web Proxy to Increase the Speed of Accessing and Processing Data

Let's look at the mathematical model for a hybrid cache method under the issue of combining two cache algorithms.

We'll assume that an information system possesses objects

that are represented as N={1,2…n} (2)

The system contains a cache system with cache memory:

M={1,2…m} (3)

The cache system can be found where there is a request for an object.

w=(r1,r2,…, r_t, r_T), $r_t \in N$ (4)

Where Object-r_t belongs to a certain N which is a request from the cache in a certain time 't'

We symbolize the set of the subsets where the object N is present in the Cache memory(M) as:

$\mathcal{M}_m$

$$\mathcal{M}_m = \{S | S \subseteq N \wedge |S| \leq m\} \quad (5)$$

When making a decision on solving a problem in a cache system you can use two algorithms, we'll assume that the following cache system:

Cache $A_1$and A_2 where they are:

$$A_1 = (Q_1, q_{10}, g_1), \quad (6)$$

$$A_2 = (Q_2, q_{20}, g_2), \quad (7)$$

Where Q_1 and Q_2 is a set of management algorithms for A_1 and A_2. Where q_{10} and q_{20} are the initial state for managing the algorithm for A_1 and A_2.. The process flow for A_1 and A_2 which based on the cache's state and the management algorithms state Q_1 and Q_2 would be the g_1 and g_2 algorithms. A new state for the cache and the algorithms for Q_1 and Q_2 also A_1 and A_2 of management

$$g_1 : \mathcal{M}_m \times Q_1 \times N \rightarrow \mathcal{M}_m \times Q_1; \quad (8)$$

$$g_1(S,q_1,x)=$$
$$\begin{cases} (S,q_{11}), & if \quad x\in S; \\ (S\cup\{x\},q_{11}), & if \quad x\notin S\wedge|S|<m; \\ (S\cup\{x\}\{y\},q_{11}), if\ (x\notin S)\wedge(|S|=m) \\ \wedge(y\in S)\wedge\big(y=d_1(S,q_1)\big); \end{cases} \quad (9)$$

$$d_1: \mathcal{M}_m \times Q_1 \to N \qquad (10)$$

This is in the state of A_1 algorithm

And for in the state of A_2 algorithm:

$$g_2: \mathcal{M}_m \times Q_2 \times N \to \mathcal{M}_m \times Q_2; \qquad (11)$$

$$g_2(S,q_2,x)=$$

$$\begin{cases} (S,q_{21}), & if \quad x\in S; \\ (S\cup\{x\},q_{21}), & if \quad x\notin S\wedge|S|<m; \\ (S\cup\{x\}\{y\},q_{21}), if\ (x\notin S)\wedge(|S|=m) \\ \wedge(y\in S)\wedge\big(y=d_2(S,q_2)\big); \end{cases} \quad (12)$$

$$d_2: \mathcal{M}_m \times Q_2 \to N \qquad (13)$$

The hybrid algorithm A for the A_1,A_2 algorithm will be called the organized set

$$A = (Q_1, Q_2, q_{10}, q_{20}, g), \qquad (14)$$

The process flow of the hybrid algorithm g

$$g: \mathcal{M}_m \times Q_1 \times Q_2 \times N \to \mathcal{M}_m \times Q_1 \times Q_2 \quad (15)$$

Then:

$$\big(S_1, q_{11}, q_{21,}\big) = g(S,q_1,q_2,x). \quad (16)$$

$$g_1(S,q_1,x)=$$

$$\begin{cases} (S,q_{11},q_{21}), & if \quad x\in S; \\ (S\cup\{x\},q_{11},q_{21}), \\ if \quad x\notin S\wedge|S|<m; \\ (S\cup\{x\}\{y\},q_{11},q_{21}), \\ if\ (x\notin S)\wedge(|S|=m)\wedge \\ (y\in S)\wedge\big(y=R(S,q_1,q_2)\big); \end{cases} \quad (17)$$

Where a certain Object-X has a request in a time 't'.

S – The state of the cache memory at the time 't'.

q_1- The state of the managing algorithm A_1 at the time 't'.

q_2- The state of the managing algorithm A_2 at the time 't'.

q_{11}- The state of the managing algorithm A_1 at the time 't+1'.

q_{21}- The state of the managing algorithm A_2 at the time 't+1'. And in the event of a hybrid algorithm R it will be the following: $R: \mathcal{M}_m \times Q_1 \times Q_2 \to N$ (18)

In the state R of the cache memory and a state q_1,q_2 for the managing algorithm and the algorithms A_1,A_2 it is possible to choose an Object-Y to show state. The cache memory S. From this model we can see that the main problem of hybridization is summarized by choosing R which allows the selection of the management state q_1,q_2 for the algorithm A_1,A_2 . To manage the influential power of the algorithms A_1,A_2 we provide the normal management concept. We'll assume a variable without an influential power that represents the algorithms A_1,A_2 for deciding the decision R. $\lambda \in [0,1]$. The variable will be the definition of the variable of managing the influential power for the algorithms A_1,A_2. In the state of [0,1] symbolized by **I**:

It will become **I**=[0,1] and R's syntax will be:

$$R: \mathcal{M}_m \times Q_1 \times Q_2 \times I \to N, y = R(S, q_1, q_2, \lambda) \quad (19)$$

Then R will have the following attributes:

$$R(S, q_1, q_2, 0) = d_2(S, q_2) \quad (20)$$

$$R(S, q_1, q_2, 1) = d_1(S, q_1) \quad (21)$$

Where: d_1,d_2 will reflect :

$$d_1: \mathcal{M}_m \times Q_1 \to N \quad (22)$$

$$d_2: \mathcal{M}_m \times Q_2 \to N \quad (23)$$

In this state showing the process flow of the hybrid algorithm A for A_1,A_2 is:

$$g: \mathcal{M}_m \times Q_1 \times Q_2 \times I \times N \to \mathcal{M}_m \times Q_1 \times Q_2 \quad (24)$$

Thus the general state will be:

$$(S_1, q_{11}, q_{21}) = g(S, q_1, q_2, \lambda, x) \quad (25)$$

Where λ the influential power management variable – A_1,A_2 in 't'.

$(S, q_1, q_2, \lambda, x) =$

$$\begin{cases} (S, q_{11}, q_{21}), \quad if \quad x \in S; \\ (S \cup \{x\}, q_{11}, q_{21}), \quad if \quad x \notin S \wedge |S| < m; \\ \quad (S \cup \{x\}\{y\}, q_{11}, q_{21}), if\ (x \notin S) \\ \wedge (|S| = m) \wedge (y \in S) \wedge \big(y = R(S, q_1, q_2, \lambda)\big); \end{cases} \quad (26)$$

$$R: \mathcal{M}_m \times Q_1 \times Q_2 \times I \to N \quad (27)$$

To achieve the goal of the article we suggest acquiring the hybrid algorithms through random strategy method.

5.Algorithms Hybridization Method

Let's look at the idea using an example of hybridizing two cache algorithms. When solving the selection f the random object we will either use A_1 or A_2 algorithm. The control variable is the probability of choosing algorithm A_1 . We'll assume at time 't' the probability choosing the algorithm A_1 in solving a problem is: $\lambda, 0 \leq \lambda \leq 1$ .Then to solve this equation using the A_2 algorithm it must be chosen with a probability of 1 - ʎ, we compare the random value '4' with the time [0,1] using the distribution law, if the value '4' in the time 't' is less than 'ʎ 'then we will use the algorithm A_1, this means that outcome of the random value: $d_1(S, q_1)$. If the value '4' in the time 't' is less than xx then the algorithm used to solve this will be A2, that means that the outcome of the random value : $d_2(S, q_2)$. We will call the suggested method for hybridization as 'Stochastic Hybrid Method'. The hybrid algorithm A was built on A_1,A_2 's bases.

$$A = (Q_1, Q_2, q_{10}, q_{20}, \lambda_0, g) \quad (28)$$

The initial value for the control variable in the hybrid algorithm A .$\lambda_0 \in I = [0,1]$. The process flow of the algorithm g:

$$g: \mathcal{M}_m \times Q_1 \times Q_2 \times I \times N \to \mathcal{M}_m \times Q_1 \times Q_2 \quad (29)$$

$$g_1(S, q_1, q_2, \lambda, x) = \begin{cases} (S_1, q_{11}) = g_1(S_1, q_1, x);\ (S_1, q_{21}) \\ \quad = g_2(S_1, q_2, x); \\ \quad if\ 0 \leq \zeta < \lambda; \\ (S_1, q_{21}) = g_2(S_1, q_2, x); \\ \quad (S_1, q_{11}) \\ \quad = g_1(S_1, q_1, x); \\ \quad if\ \lambda \leq \zeta \leq 1; \end{cases} \quad (30)$$

Where : S_1 — cache memory state at time 't+1'.

X = r_t – Object request in time 't'.

Where:

$$x \in S_1$$

$$\zeta \in I$$

Random number which it's probability equals to the distribution law at the time I=[0,1].

6.Hybrid Cache Algorithm for Proxy Server

If we look at the foundations of the LFU and LRU, LRU was designed based on LRU-Queue, in this situation the Hit Ratio is calculated using the following algorithm:

$$CV_{LRU} = \frac{1}{(T - T_1)} \qquad (31)$$

Where:

T is the real time.

T_i is the last time of arrival of the object.

The LFU algorithm was built based on the count of requests of the object and is calculate with:

$$CV^i{}_{LFU} = N_i \qquad (32)$$

The GDS was specially developed for web systems and servers, the Hit Ratio is calculated with:

$$CV^i{}_{GDS} = \frac{C_i}{S_i} \qquad (33)$$

Where:

C_i is the access value for the object

S_i is the size of the object.

Hybrid algorithm processing flow:

1. The object with the lowest value of the Hit Ratio will be replace.

2. The value of the Hit Ratio for each object in the cache memory decrease depending on the object with the lowest value of the Hit Ratio.
3. Repeat the first and second step until the cache has been freed ,then it allows to save temporary the new object.
4. The new object is cached and the hit ratio is calculated using the previous formula:

$$CV^{i}_{Hybrid} = \frac{P_i.C_i}{S_i} \qquad (34)$$

Where:

C_i is the access value for the object .

S_i is the size of the object.

P_i is the requests of the object depend on **:**

T , **T_i** , **N_i** .

Where:

T is the real time .

T_i is the last time of arrival of the object.

N_i count of requests of the object **.**

Let N be the total number of requests (objects).

δ_i= 1, if the request i is in the cache, while δ_i= 0, otherwise Mathematically, this can be expressed as follows:

$$HR = \frac{\sum_{i=1}^{N} \delta_i}{N} \qquad (35)$$

$$BHR = \frac{\sum_{i=1}^{N} b_i \, \delta_i}{\sum_{i=1}^{N} b_i} \qquad (36)$$, where b_i is the size in bytes of the i_{th} requested object.

7. Using the Mathematical Model for the Trace-Driven Simulation by Using Hybrid Cache Algorithm

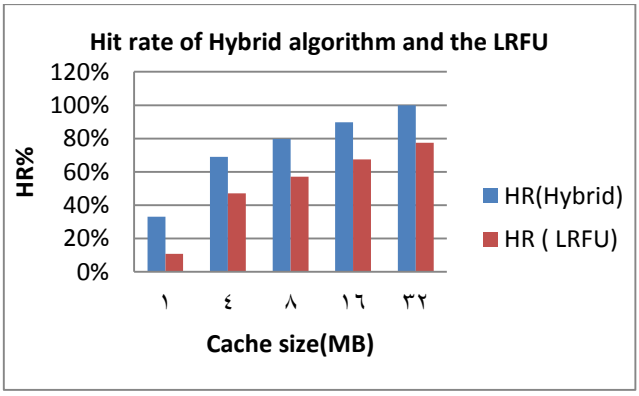
We did the trace-driven simulation to compare the performance of our proposed hybrid cache algorithm with LRFU cache replacement algorithm. Hybrid cache algorithm has used in Squid proxy which is open source and work in operating system environment Linux to improve the temporary storing and replace the web pages in proxy. Squid proxy is open source web proxy and program code and it is available for development and work as a medium between web browsers, for users and servers

containing data. Also it manages servers and users and produces caching. The proxy server keeps the repeated data in users request and storage at speeding the work and reducing to pass in the network. To fix the proxy on Google chrome, we follow:

1. Enter Google chrome settings.
2. Press show advanced settings.
3. Press change proxy settings.
4. Internet setting page will appear ,then press connection then Lan setting.
5. Tick true in the square proxy server then press advanced.
6. Write proxy server then tick true on the square use the same proxy server then press ok.

"Table 1. Hit rate of Hybrid algorithm and the LRFU".

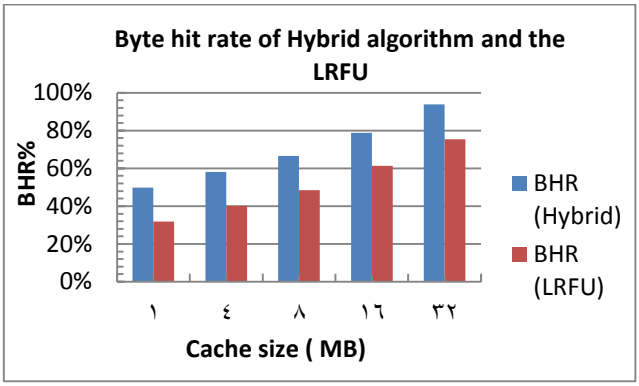
Total Cache set size is(in MB)	1	4	8	16	32
HR(Hybrid)	0.32996936	0.69046616	0.79714118	0.89816377	0.99930842
HR (LRFU)	0.10780965	0.47172469	0.57036901	0.67414663	0.77494165



"Figure.2. Hit ratio (HR) for the proxy trace"

"Table 2. Byte hit rate of Hybrid algorithm and the LRFU".

Total Cache size set is(in MB)	1	4	8	16	32
BHR(Hybrid)	0.498694 8	0.5804562 3	0.6665129 2	0.7873636 6	0.9388083 6
BHR (LRFU)	0.319245 6	0.4023471 8	0.4843876 5	0.6141368 9	0.7536817 6



"Figure.3. Byte hit ratio (BHR) for the proxy trace"

Hit Ratio (HR), Byte Hit Ratio (BHR) are the most widely used metrics in evaluating the performance of Web caching. HR is defined as the percentage of requests that can be satisfied by the cache. BHR is the number of bytes satisfied from the cache as a fraction of the total bytes requested by user. Total Cache size set is 1MB,4MB, 8MB,16MB,32MB. The results of our hybrid cache algorithm indicate that Hit Ratio(HR) was 22% and Byte Hit Ratio (BHR) was 18% compared with famous hybrid cache algorithm(LRFU). This indicates that our hybrid cache designed algorithm is more efficient than famous hybrid cache algorithm(LRFU).

8.Conclusion

In this scientific article, a mathematical model design and stochastic hybrid random method design for a cache system to get a hybrid cash algorithm to increase

the speed of accessing, processing and saving data. The hybrid cache method can hybridize any type of cache algorithms with each other. The simplicity of this method in hybridization is not only exclusive to web server but can be also used with databases, information systems, operating systems and CPU. Experiments were carried out to measure performance proficiency of hybrid cache algorithms by helping trace-driven simulation by comparing between our hybrid cache algorithm with famous hybrid cache algorithm(LRFU). The results of our hybrid designed algorithm indicate that Hit Ratio(HR) was 22% and Byte Hit Ratio (BHR) was 18% compared with famous hybrid algorithm(LRFU). This indicates that our hybrid cache designed algorithm is more efficient than famous algorithm(LRFU).

9. References

[1].Sirshendu Sekhar Ghosh and Dr. Aruna Jain, "Web Latency Reduction Techniques: A Review Paper", IJITNA vol.1 No.2 pp20 – 26, September, 2011.

[2].Sirshendu Sekhar Ghosh, Dr. Aruna Jain, "Hybrid Cache Replacement Policyfor Proxy Server", International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 3, March 2013 .

[3]. M. Dehghan, L. Massoulie, D. Towsley, D. Menasche, and Y. Tay, "A Utility Optimization Approach to Network Cache Design," in Proc. of IEEE INFOCOM 2016, 2016, to appear, arXiv preprint arXiv:1601.06838.

[4].C. Fricker, P. Robert, and J. Roberts, "A versatile and accurate approximation for LRU cache performance," in Proceedings of the 24th International Teletraffic Congress, 2012,p. 8.

[5].N. C. Fofack, P. Nain, G.Neglia, and D.Towsley, Performance evaluation of hierarchical TTL-based cache networks,"Computer Networks,vol. 65, pp.212 – 231, 2014.

[5]. JyotiPandey, AmitGoel, Dr. A K Sharma "A Framework for Predictive Web Prefetching at the Proxy Level using Data Mining" IJCSNS,VOL.8 No.6, pp303-308, June 2008.

[6]. Shaul Dar, Michael J. Franklin, BjörnÞórJónsson, DiveshSrivastava, and Michael Tan. Semantic Data Caching and Replacement. VLDB, 1996.

[7]. A.Balamash, M.Krunz, "An Overview of Web Caching Replacement Algorithms",IEEE Communications Surveys & Tutorials, Second Quarter, 2004.

[8]. Nimrod Megiddo and Dharmendra S. Modha. ARC: "A Self-Tuning, Low Overhead Replacement Cache". FAST, 2003.

[9]. Yuanyuan Zhou, James Philbin, and Kai Li. The 23Multi-Queue Replacement Algorithm for Second Level Buffer Caches. USENIX, 2002.

[10]. Eduardo Pinheiro , Ricardo Bianchini, Energy conservation techniques for disk array-based servers, Proceedings of the 18th annual international conference on Supercomputing, June 26-July 01, 2004, Malo, France.

[11]. Cheng Li, Philip Shilane, Fred Douglis and Grant Wallace. Pannier: A Container-based Flash Cache for Compound Objects. ACM/IFIP/USENIX Middleware, 2015.

[12]. P. Cao, E.W. Felten, A.R. Karlin, K. Li, "A Study of Integrated Prefetching and Caching Strategies", In Proc. Of the ACMSIGMETRICS Conference on Measurement and Modeling of Computer Systems, pp. 188-197, 1995.

[13] N. Beckmann and D. Sanchez, "Jigsaw: Scalable software-defined caches," in PACT-22, 2013.

[14] Q.Yang, J.Z.Huang, N.Michael, "A Data Cube Model for Prediction-Based Web Prefetching", Journal of Intelligent Information Systems, 2003,vol.20 (1),pp.11-30.

[15].W.-G. Teng, C.-Y. Chang, and M.-S. Chen, "Integrating Web caching and Web prefetching in client-side proxies," IEEE Transactions on Parallel and Distributed Systems, vol. 16, no. 5,pp. 444–455, 2005.

[16]. C. Bouras, A. Konidaris, and D. Kostoulas, "Predictive prefetching on the Web and its potential impact in the wide area." World WideWeb, vol. 7, no. 2, pp. 143–179, 2004.

[17]. N. Beckmann and D. Sanchez, "A cache model for modern processsors," MIT, Tech. Rep. MIT-CSAIL-TR-2015-011, 2015.

[18] Bhawna Nigam and Dr. Suresh Jain, "Analysis of Markov Modelon different Web Prefetching and Caching schemes", IEEE, 2010.

[19]. L. Shi, B. Song, X. Ding, Z. Gu, and L. Wei, "Web prefetching control model based on prefetch-cache interaction," in Proc. First International Conf. on Semantics, Knowledge and Grid SKG '05,2005.

[20]. Jyoti Pandey, Amit Goel, Dr. A K Sharma "A Framework forPredictive Web Prefetching at the Proxy Level using Data Mining"IJCSNS,VOL.8 No.6, 303-308, June 2008.

[21]. F.J. González-Cañete, E. Casilari, A. Triviño-Cabrera "A contenttypebased evaluation of Web Cache replacement policies" IADISInternational Conference Applied Computing 2007.

[22]. L .Donghee, H. N. Sam, C. Jongmoo, L. M. Sang, "Implementation and Performance Evaluation of the LRFU Replacement Policy", Dep. Of Computer Engineering, Seoul National University, Seoul 151-742 Korea.

Toward a model for Cloud Computing Banking in Yemen

Fursan Mohammed A. Thabit
PCTE college, Punjab Technical University, India
enrfursan@gmail.com

Abstract

Cloud computing is one of nowadays technology trends which provides a new way to manage the different information systems, such as banking systems. The transactions between the banks' accounts must be managed with a high-level security and a high-level performance with a real-time system management if the system needs for an emergent system recovery. All of these features can be provided easily and on-demand using the cloud computing. In addition, like any other telecommunication technologies, it needs some specific components which must be connected in a standardized manner to achieve the best results. Yemeni banks have parts of this needed infrastructures. However, they hadn't started their own cloud computing banking yet due to the fear of losing the high security offered by the local networks providers. In this research, we are going to highlight a roadmap of developing a model for cloud computing banking in Yemen. This will contribute in upgrading the technologies used in the managements of the information systems of banks in Yemen.

Keywords: Cloud Computing, Banking Management, on-demand services, Scalable infrastructure, Cloud computing banking, Yemeni banks.

I. Introduction

Cloud computing today encompasses every vertical in the market across sectors. Organizations are adopting innovative cloud apps to support their everyday business operations. Cloud technology offers secure deployment options that can help banks develop new customer experiences, enable effective collaboration and improve speed to market, all while increasing IT efficiency. Banks that take advantage of cloud computing are better positioned to respond to economic uncertainties, interconnected global financial systems and demanding customers [1]. Before moving to the cloud, banks must consider issues around data confidentiality, security, regulatory compliance,

interoperability of standards, and quality of services [2]. The banks in Yemen tried to solve these issues using the traditional networks depending on the internet providers and some data center infrastructures. Cloud adoption can improve banking speed of performance, cost avoidance, simplification of the service catalogue and revenue growth. Cloud adoption also add benefits like: respond to growing competition, expand the bank's brand, and avoid risks and manage security [3].

There are many researches which study the banking as an online banking, electronic banking, mobile banking and the internet banking. Those pieces of research study all those fields from different perspectives. The cloud computing banking is a new technology which benefits the internet technology and many other technologies to provide the security, ease of use, availability and scalability. This new technology achieved various successes especially with the systems which need to have high availability. Some of Yemeni banks started in the last few years offering some infrastructures those can be used to create a cloud computing system. However, no one of those banks had used the cloud computing yet. Moreover, the boards of Yemeni banks don't trust in cloud computing as a secure technology, although they can have their own cloud computing systems.

There're some reasons to have a cloud computing banking [4-7]:

- **Performance:** using the HPC and the HTC, cloud computing performance is better than the performance of any other communication technology. Traditional network performance can provide a very little increase of the performance.
- **Availability**: the cloud computing services is available anytime everywhere. Most of traditional management systems in Yemeni banks are available in the working hours only.
- **Security:** The local networks are safer than the wide-range networks in general. But according to the solutions provided by many researchers and security companies, cloud computing can provide various solutions for this matter. There are two main cloud computing models (public and private cloud computing). The private cloud computing is more secure than the public cloud and therefore it will be a good choice for the banking systems.
- **Energy efficiency:** Cloud computing can save power more efficiently than any other communication technology. This is due to the developed algorithms and researches in greening and enhancing the cloud computing.
- **On-demand self-service:** provided by cloud computing providers to enable the cloud resources provisioning on demand whenever they are required. In on-demand self-service, the user accesses cloud services

through an online control panel. The traditional system management can't do all of that physically or with the same quality of services. On the other hand, traditional systems need to pay everything it needs, where the cloud computing resources can be just rented.

- **Broad network access:** which are the resources hosted in a private cloud network (operated within a company's firewall) that are available for access from a wide range of devices, such as tablets, PCs, Macs, and smartphones. These resources are also accessible from a wide range of locations that offer online access. The traditional systems can use these resources, but of the range of accessed devices is very little, then the resources may not work in the same efficient way.
- **Resource pooling:** describe a situation in which providers serve multiple clients, customers or "tenants" with provisional and scalable services. These services can be adjusted to suit each client's needs without any changes being apparent to the client or end user. This feature can't be provided by a very limited number of resources hosted in the same geographical place.
- **Rapid elasticity:** ability to provide scalable services. Capabilities can be elastically provisioned and released, in some cases automatically, to scale rapidly outward and inward commensurate with demand. To the consumer, the capabilities available for provisioning often appear to be unlimited and can be appropriated in any quantity at any time. Scaling the resources is limited to be offered using the limited resources of the traditional systems. The resources, used in the banking management systems by many banks in Yemen, are less than needed especially when there is a work load.
- **Measured Service:** Cloud systems automatically control and optimize resource use by leveraging a metering capability at some level of abstraction appropriate to the type of service (e.g., storage, processing, bandwidth and active user accounts). Resource usage can be monitored, controlled and reported, providing transparency for the provider and consumer. This feature is not provided by any banking system in Yemen till this moment.

According to the all above, we are going to develop a new model that will try to provide all these features as possible in the cloud computing banking to be adopted as the new technology to use and to provide the banking services in Yemen. The rest of this research contains related works of the cloud computing banking, the proposed model for the cloud computing banking, the discussion of that model, and finally the conclusion.

II. Related work

According to Jiang, et al. [8], "Cloud computing is an Internet-based application platform, which provides enterprises with shared resources including hardware, software, data storage, and customizable business applications on demand." Cloud computing is a result of four technologies convergence: Hardware Technology, Internet Technology, Distributed Computing Technology and the System Management. Cloud computing enhances business functions for everything from advanced data analytics to compliance activities and HR [9]. Nowadays it covers various fields of our daily lives, especially in commercial side. The most important part of commercial processes is the money exchanging which at the end will depend on accounts in Banks for the both sides of the exchanging transaction.

Cloud computing today encompasses every vertical in the market across sectors. Organizations are adopting innovative cloud apps to support their everyday business operations. Cloud technology offers secure deployment options that can help banks develop new customer experiences, enable effective collaboration and improve speed to market, all while increasing IT efficiency. Banks that take advantage of cloud computing are better positioned to respond to economic uncertainties, interconnected global financial systems and demanding customers [1]. Before moving to the cloud, banks must consider issues around data confidentiality, security, regulatory compliance, interoperability of standards, and quality of services [2]. The banks in Yemen tried to solve these issues using the traditional networks depending on the internet providers and some data center infrastructures. Cloud adoption can improve banking speed of performance, cost avoidance, simplification of the service catalogue and revenue growth. Cloud adoption also add benefits like: respond to growing competition, expand the bank's brand, and avoid risks and manage security [3].In Capgemini perspective [10], when a bank moves into cloud computing, there are two primary challenges that must be addressed: The first is the security. The confidentiality and security of financial and personal data and mission-critical applications is paramount. Banks cannot afford the risk of a security breach. The second is the regulatory and compliance. Many banking regulators require that financial data for banking customers stay in their home country. Certain compliance regulations require that data not be intermixed with other data, such as on shared servers or databases. As a result, banks must have a clear understanding of where their data resides in the cloud. Financial institutions must select the right service, deployment, and operating models to address security and compliance concerns. In the initial phases of cloud computing adoption, it is expected that banks will own and operate the cloud themselves with service providers taking increasing ownership and control of the cloud infrastructure as cloud computing matures and more rigorous controls become available [10].

According to Mbelli, at al. [11], management of big data becomes more important as cyber banking is becoming more evolved and becoming more and more digital. The storage of this big data will rely on technologies such as cloud. The issues of the security of these cloud technologies may include things like the vendor lock-in threats, what level of security to apply, business continuity if there is a breach of the cloud security, etc. The top vulnerabilities to South African cyber banking are [11]:

- Less or inadequate maintenance, checking and analyzing of the security incidents audit logs.
- Inadequate or insufficient security applied at application software level.
- Less control of admin privileges.
- Less monitoring and control of access to cyber systems.
- Insufficient or Inadequate incident response capabilities.
- Inadequate or insufficient security assessment and awareness.
- Poor or no policy and standard management in place.

Sahil stated that, Banks may have various reasons for migrating to the cloud, but the main reason applications. A pivotal stumbling block for huge investments in new technologies has always been the capital expenditure needed for advance infrastructure. With cloud computing, various financial institutions only have to budget for functional expenses and wage for the services they use. This makes it effortless and more cost effective to test new applications on the cloud versus prevailing conventional infrastructures [1]. The matter is that not all the banks converted to the cloud computing. In addition, there are very few countries at the whole world that they started adoption the cloud computing.

Ghule, et al. [12] stated that, Cloud computing can help financial institutions improve performance in a number of ways: Cost Savings and Usage-based Billing, Business Continuity, Business Agility and Focus, and Green IT. Goel stated that The majority of today's cloud computing infrastructure consists of time-tested and highly reliable services built on servers with varying levels of virtualized technologies, which are delivered via large data centers operating under service-level agreements (SLAs) that require 99.99% or better uptime. Commercial offerings have evolved to meet the quality-of-service requirements of customers and typically offer such service-level agreements to their customers. From users', i.e., bank's perspective, the cloud appears as a single point of access for all their computing needs. Regardless of the geographic location these cloud-based services are accessible as long as an Internet connection is available. Open standards and open-source software have also been significant factors in the growth of cloud computing

III. Cloud Computing in banking systems

According to Al-Ariqi [20], there are two types of commercial local banks in Yemen, private and state banks. Those banks are found in the main cities only, i.e. Sana'a, Aden and the capital cities of each other governorate. The foreign banks hasn't the same problems of the local ones because the used techniques for management are absolutely different. Both of the local and foreign banks have many branches in different places in Yemen [20]. Those branches must be connected via a network. According to our interview, with one of the network experts who build and configure the banks infrastructure in Sana'a, the local banks use traditional networks. They are really have data centers, firewalls and online and offline servers. All of these equipment are installed randomly and don't follow any standards of data center or server installation. The networks' providers are local providers. The firewall is at most a hardware set and in rare they use a software firewall. The servers are also installed locally and the banks' databases are not connected to the internet. When we ask about the reason, the answer was that they can't trust in the security of the internet connection.

Although the banks systems are highly diffused in many different places in Yemen, the manual trading is the main and sometimes the only way for the money exchange. This means deficiencies in the basis of the work of the Yemeni banking sector. In addition, the performance of the operating banks in Yemen degrades day per day. The ATM is not provided by many Yemeni banks [20]. The banks those provided ATM in past, nowadays stop their ATM.

The internet banking is not a used service in any Yemeni local bank according to our study in the last month. Those banks websites are just for ads and their contents aren't changed. Some of the foreign banks provide this service. The cloud computing banking is not provided yet in any bank, both local and foreign ones. The local banks in Yemen has a weak competitiveness in the face of the regional and international foreign banks [20].

Yemeni government sector systems have no trust in internet technology. This is due to some problems can be summarized in the following:

- The e-government is just a proposal and Yemeni government has no infrastructure to manage the electronic transactions.
- Piracy, sabotage and sometimes destruction of the international networks makes the government reluctant to launch the e-government.
- Some people can't trust in others.
- Some ones in banks' boards can't believe that the technology can do everything effortless. Some others love destruction of the success projects even though those projects will meet the market needs.
- The Administrative corruption.

According to all above, we aim to develop a model for the cloud computing banking in Yemen. This model must ensure high security to be adopted by the

banks' boards. In addition, the model will enhance the banking services to be available anytime, everywhere and for any user. We propose this model will provide the banking services whatever the load-balancing rate is. This will be helpful in the times of payment of salaries, distribution of quotas of stock gains and so on. Using the security provided by the private clouds to make the banks' board trust the cloud computing. The result of their trust may lead to adopt the proposed model in banking. We aim also to provide the banking management services on-demand. This can be provided according to the service level agreement between the providers and the banks' boards.

We aim to develop a model for the cloud computing banking in Yemen. First, we will study the impact factors of banking usage and the real demands in the management of banks. Then will suggest the type of the cloud (public, private, or hybrid) and we will specify the needed equipment and standards according to each type of these types. After that we will suggest the needed standards and circumstances to adopt the resulted model in order to have a cloud computing banking management system. The final step will be to develop a sample example for the model and to test it to prove that it is applicable.

Statistically, we will make some questionnaire which will target the clients, employees and the banks' managements. We will visit the banks to have a real observation every time we can have the allowance from the management. Then empirically we will develop the new cloud computing banking model and will test it with a sample dataset with the agreement of one or two banks in Sana'a and Aden. The final results after that can be generalized for the all other banks in the different provinces.

IV. The proposed Model

We proposes the model shown in figure 1. This model consists of the three levels of the cloud computing architecture, IaaS, PaaS, and SaaS. The banking system components are distributed among these three to gain the essential features of the cloud computing. The model suggests a small private internal cloud for the bank. Then the management has the choice, either to use the public or the private cloud computing. The private cloud will be based on a public infrastructure to enable the cloud providing its features. The PaaS and SaaS must be owned by the bank management and we recommend the VM and its VMM too. In the public cloud, if the bank privacy and policy standards allow, the IaaS, PaaS and SaaS can be all public. The small private internal cloud must ensure the synchronization of data between the internal database and the cloud database. Some secure tunnels can be provided to ensure security.

V. Discussion

As we mentioned above, there will be two suggestions for the cloud computing banking, either to be fully public with a small private internal cloud, or to be fully private with a small private internal cloud except the IaaS layer. The first suggested to provide better performance and more high level cloud computing features. But it may be less secured than the private one. It is cost for implementation will be less than the private one because all the layers and their services can be rented. On the other hand, to save the security the model must be chosen to be private. The private cloud will be in two places. The internal cloud will be in the local in the bank and its branches. All the needed resources must be bought previously by the bank's board. The global cloud will depend on a public infrastructure to provide the hardware management and scalability. There must be some VMs to be easy to manage, easy for data migration and to provide security for the data on the public infrastructure. The platform can be rented or bought. The SaaS must be owned by the bank's board to ensure a high system and data security.

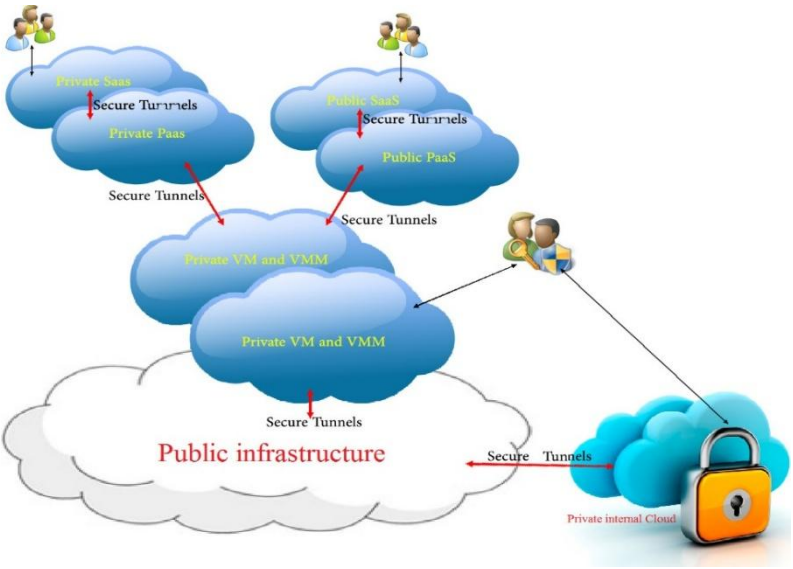


Figure 1: The proposed model

VI. Conclusion

This paper was written to propose a model for the cloud computing banking in Yemen. This model can be used to manage and to provide the banking services using the cloud computing. We assume that can be adopted by any other banks is in the same environment or in any other environment that has the same features. As a future work, this model must be tested, extended and enhanced.

VII. References

[1]. S. Patani, S. Kadam, and P. V. Jain. Cloud computing in the Banking sector: A survey. International Journal of Advanced Research in Computer and Communication Engineering, 3(2), (2014), pp: 5640-5643.

[2]. P. Gupta, A. Seetharaman, and J. R. Raj. "The usage and adoption of cloud computing by small and medium businesses". *International Journal of Information Management*, *33*(5), 2013, pp: 861-874.

[3]. A. A. Adelola. "New Media and the marketing strategies of Nigerian banks".

[4]. B. Nedelcu, et al. "Cloud Computing and its Challenges and Benefits in the Bank System", *Database Systems Journal*, 4(1), 2015.

[5]. K. Chandrasekaran. "Essentials of cloud computing*". CRC Press*, 2014.

[6]. T. Dillon, C. Wu, and E. Chang. "Cloud computing: issues and challenges". *24th IEEE international conference on advanced information networking and applications*, IEEE, 2010, pp:27-33.

[7]. K. Hwang, J. Dongarra, and G. C. Fox. "Distributed and cloud computing: from parallel processing to the internet of things". *Morgan Kaufmann*. 2013.

[8]. M. Jlelaty and Y. Monzer. "Factors in cloud computing adoption". 2012.

[9]. R. Buyya, J. Broberg, and A. M. Goscinski. "Cloud computing: Principles and paradigms". (Eds.). *John Wiley & Sons.* (Vol. 87). (2010).

[10]. S. Sriram. "Cloud Computing in Banking: What banks need to know when considering a move to the cloud", 2011.

[11]. T. M. Mbelli, and B. Dwolatzky. "Cyber Security, a Threat to Cyber Banking in South Africa: An Approach to Network and Application Security". *Cyber Security and Cloud Computing (CSCloud), 2016 IEEE 3rd International Conference*, IEEE, 2016, pp. 1-6.

[12]. S. Ghule, R. Chikhale, and K. Parmar. Cloud Computing in Banking Services. *International Journal of Science and Research Publications*, *46*, (2014), pp: 1-2.

Towards an Arabic-English Machine-Translation Based on Semantic Web

Neama Abdulaziz Bin-Dahan
Department of Computer Science
Sana'a University, Sana'a, Yemen
neama.abdulaziz@gmail.com

Fadl Mutaher Ba-Alwi
Department of information systems
Sana'a University, Sana'a, Yemen
fadlbaalwi@gmail.com

Ibrahim Ahmed Al-Baltah
Department of Information Technology
Sana'a University, Sana'a, Yemen
albalta2020@gmail.com

Ghaleb H. Al-gapheri
Department of Computer Science
Sana'a University, Sana'a, Yemen
drghalebh@yahoo.com

Abstract

Communication tools make the world like a small village, and as a consequence people can contact with others who are from different societies or who speak different languages. This communication cannot happen effectively without Machine-Translation because they can be found anytime and everywhere. There are a number of studies that have developed Machine-Translation for the English language with so many other languages, except the Arabic it hasn't been considered yet. Therefore, the aim of this paper is to highlight a roadmap for our proposed translation machine to provide an enhanced Arabic-English translation based on Semantic and to illustrate its work.

Keywords: Machine-Translation, Ontology, Semantic Web, F-measure.

i. Introduction

Nowadays, technology makes our life much easier through the mitigation of the daily hardness. Different people from different societies can now communicate with each other easily and without needing to ask any other persons to be a translator during their business transactions or their conversations. Students can also study any foreign language, either online or by travelling to another country to get a certificate.

Languages mean culture that is why when we talk about Arabic language, we must talk about the Holy Quran. The Holy Quran is the richest Arabic document with the vocabularies, and Arab people consider it as the huge reference of keeping their language safe [1-3]. Therefore, people need to understand the Holy Quran to know as perfect as possible. However, Arabic language's structure is so complex. In addition, the vocabularies of the Arabic language depend on the derivation and morphology. English language, on the other hand, is much easier, and there are many people who can speak English like if they were native. It is also an essential subject in schools and there are many institutes and centers that teach it [4].

According to Al-Raheb, et al, [5] English language has just one rule to construct the sentence, i.e. subject (S) then verb (V) and after that the sentence complement (C) or the object (O). The sentence in the Arabic language, on the other hand, can have one of four different rules to construct the word, i.e. either S+V+(C,O), V+S+(C,O), V+(C,O)+S, (C,O)+S+V or S+(C,O)+V [5]. The homographs in both languages are also a huge problem while translation, because translation may give us another meaning that was not intended. Thus, knowing homographs is a very important task while translation [6]. Using a machine that can specify the intended meaning of the vocabularies used to build the sentence is highly needed.

We notice that, the semantic web provides us a strong means in order to build a translation machine. That translation machine can give us a more efficient translation than that provided in the online translator or any other types of translators or statistical machines. It is effective according to the coherent architecture that was developed to make analysis for the sentence more than once. First is to do a natural language analysis. Then to make morphological analysis beside the context analysis to know the real meaning that the source means. After that is to use the statistical analysis to get the final translation that the machine will send to the target. This issue still found in the Arabic language, which is probably the main reason for the absence of complete and efficient translation machines in the state of the art. This is in turn provided an opportunity for researchers to enrich this topic further. From another point of view, most of current researches have focused on the parser technique only. Because of that, this study intends to do a semantic analysis in addition to the statistical analysis. The semantic analysis can be divided into a morphological analysis and ontology to specify the correct meaning in the other language.

The researchers demonstrate that the morphological analysis should be done first before the translation. Then, using the normal statistical Machine Translation (MT) could be quietly enough to get the right meaning of the sentence.

The aim of this paper is to propose a translation machine to provide an Arabic-English translation. The reminder of this paper is organized as follow: Section 2 presents the related works in some details. The proposed machine model is demonstrated in section 3. Section 4 presents the discussion of this study. Section 5 concludes this study.

ii. Related work

Today's machine translations (MTs) have many different parts to give us the most efficient translation it can. Thus, there are so many sets of research that contribute to developing these parts. We are going to talk here about some developments in MTs, parsers and Ontologies.

1. Machine-Translation (MT)

The statistical method was introduced to be a new approach for MT [7]. Statistical MT was stated by IBM researchers in [8]. They thought that MT is old as the first generation of the computers. They also stated that the translation must depend on many factors, the most essential one was the whole original text itself. In contrast, they treated the words without recognizing the connection between words or even recognizing the sentence structure. Hutchins stated that the MT system may be

designed to satisfy the following criteria: to deal with single words, to get restricted input text structure, to have pre-edited input texts with any grammatical category, and without caring about the ambiguity of words or any other operators during translation. They also classified the MT to many groups according to the type of aiding (machine-aided MT, human-aided MT, computer-aided translation), input or output edition (pre-edited or post-edited), number of targeted languages (bilingual or multilingual), translation approach (direct translation, interlingua, or transfer) and syntactic structure analysis (predictive analysis, phrase structure or dependency grammar). They also pointed to the importance of semantic MT, but it was just a survey and there were no implementations for their vision [9]. Chan, et al [10], integrated a state-of-the-art Word Sense Disambiguation system into a state-of-the-art hierarchical phrase-based MT, "Hiero" [10]. In contrast, they demonstrate only one way for the integration without introducing any rules that compete against existing rules [11]. Gupta et al, [11] introduced 16 features that were extracted from the input sentence and their translation and then showed a quality score based on Bayesian inference produced from their sample training data, but they didn't develop a new English-Hindi MT and their work was just an analysis study for an existing MT [11].

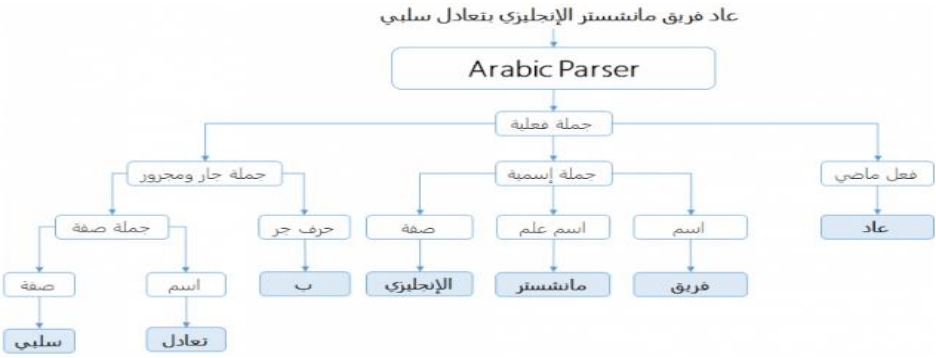


Figure 1: Microsoft's ATKS Parser analysis for Arabic sentence

2. Arabic Parsers

Natural Language parser is a machine that can understand the sentence parts and serve us in the translation using the MT. There were many researchers who have worked on this type of machine, but, here, we are going to explore some of the most modern Arabic Parsers which made a real differentiation in Arabic translation and in Arabic natural language processing.

An example of the parsers is Microsoft ATKS as shown in figure 1. This parser was used by Alagha, et al in [1-3] to simplify the process of converting the Arabic sentence into the RDF triples. Arabic Parser was proposed in [12-15]. English language has just one rule to construct the sentence, i.e. subject (S) then verb (V) and after that the sentence complement or object (C,O). However, Al-Raheb, et al [5], stated that, Arabic is a free word order: SV(C,O), VS(C,O), V(C,O)S [5]. In our vision, the structure of the Arabic sentence can be more various than that they stated.

In addition, according to Green, et al [16], their parser was similar to another Treebank in gross statistical terms, annotation consistency remains problematic [16]. Tounsi, et al [17], thought that the best-known Arabic statistical par ser, at that time, was developed by [17]. Thus they tried to enrich that parser's output with more abstract and deep dependency information. On the other hand, the categories they added to Bikel's parser resulted in substantial data-sparseness [18]. Green, et al [16], introduced the Stanford gold parser and showed that, in their paper, Arabic parsers is poorer in thoughts and still much lower than English ones. Their proposed parser, according to Stanford University, is demonstrated as one of the most effective and found parsers [16]. Al-agha, et al [1], developed a new parser which has many functionalities over than others [1]. Zaghouani, et al [19], added new specified post-editing rules of correction, according to seven categories: spelling errors, word choice errors, morphology errors, syntactic errors, proper name errors, dialectal usage errors, and punctuation errors [19].

3. Ontology-Based Machine-Translation

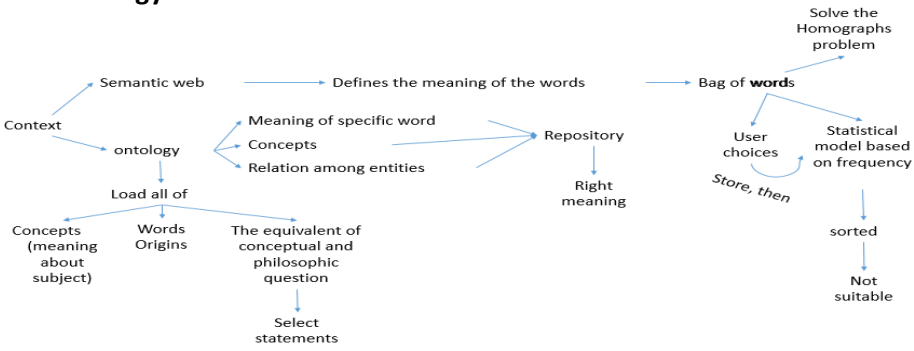


Figure 2 Dialog Context system analysis

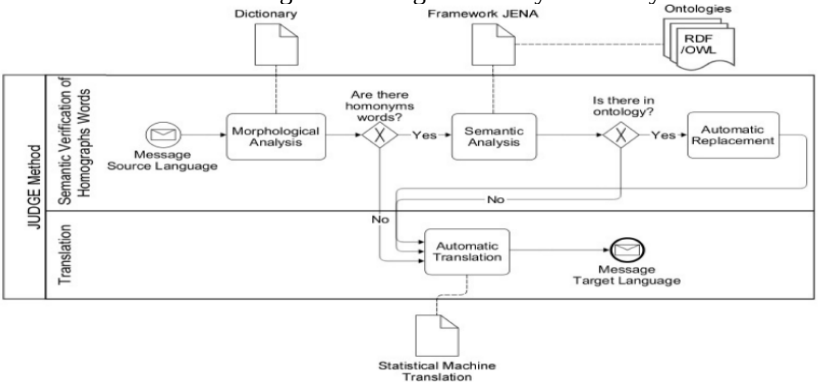


Figure 3: Portuguese Method Judge

Shi, et al, developed a world model for Chinese-English MT using an Ontology-driven [20]. Seo, et al [21], presented a syntactic and semantic method for English-Korean MT using Ontology for web-based MT [21]. Mousallem, et al [6], solved the ambiguity in dialog conversation using an ontology-based MT [6]. However, all of them can only serve

for their proposed languages. They also mentioned that the instant translators like google, bing and more others are statistical translators and have no semantic processing in their works [6]. Alagha, et al [1], presented a domain-independent approach to translate Arabic NL queries to SPARQL by getting benefits from linguistic analysis [1]. It was just to translate the Arabic questions to SPARQL, not to other human-language. The systems developed in [2,3] take questions expressed in Arabic and returns the answers drawn from an ontology-based knowledge base though the ontology file was not prepared as the system needs. While the translation is done to the SPARQL level only, it would not work as an MT between human-languages and it would be like the previous works that didn't translate the Arabic question sentences to any other human-languages.

The main solution of the Portuguese-English MT was to provide an efficient way to process the homographs [6]. And where the last dialog systems were depending on the statistical MTs. Thus, they proposed a semantic analysis before the traditional translation of the statistical MT. Figure 2 shows the mental plan, from which their proposed model was invented. Then they developed a model which was named the Method Judge as shown in figure 3. It consists of two essential parts. The semantic verification is to show the right meaning of the ambiguous words or the homographs which is followed by the automatic replacement of the vocabularies meanings. There will be a morphological analysis connected with the source language dictionary to specify the right meaning of the homograph words. There is also a semantic analysis to get the right meaning of that homograph from the language ontology. Then the resulted words will be replaced automatically with their effective meanings. The second part was the automatic translation of the remaining parts of the sentence sent in the dialog system. This type of translation will be done statistically using a cluster algorithm, like the way used in the instant translators such as google and bing translators.

According to Mousallem, et al [6], both languages have the same order of the sentence parts, there was no need to have two different parsers to process the natural language. The parser, in that research, was merged in the morphological analysis model, in which the sentence will be divided into its main parts and they stated here that the homographs may be either a noun or a verb. Thus, the automatic replacement will never generate any error. Even though the auxiliary verbs in one of the proposed languages (English or Portuguese) has a different meaning or usage, the meaning will be efficiently specified because they both have the same sentence structure. The researchers test their model with five different online translators (Google, bing, Worldling, Gengo and Systran) and the results were very interesting [6].

Al-agha, et al [1-3], proposed a machine to translate the Arabic questions to SPARQL. They used the ontology for the linguistic and the semantic analysis and to eliminate the ambiguity of homographs [3], although the system does not make intensive use of sophisticated linguistic methods. They also defined a Natural Language (NL) interface for questions formulated in natural language to return answers on the basis of a given which was that those systems were not able to retrieve precise answers to questions, but only to retrieve a set of relevant documents using the given keyword-based query [1-2].

iii. The proposed Machine Model

We can extend the Method Judge mentioned in [6] with the parsers idea in a model that can be illustrated in figure 4.

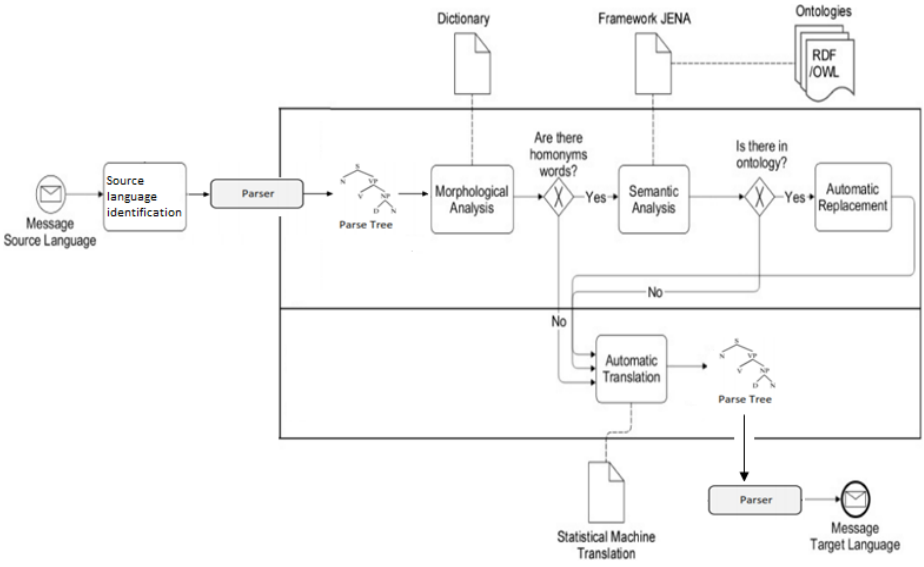


Figure 3: The proposed model

The Main method judge was explained in the last point. The extended MT system will be built from many other toolkits. The model assumes that, after choosing the source language, the machine must be able to detect the language from context before the source sentences are passed through a parser. The parser will specify the parts-of-speech in that language. We can after that decide which one of these words can be considered as a homograph. The homographs will be analyzed using the morphological analysis. Using the language ontology and the context of the source text, the correct translation will be chosen. The semantic analysis will be used to make sure which meaning of the available meanings of that homograph word is the most closer to the context. After that, a simple statistical MT will be done. Here must be a toolkit that must make sure the precision of the translation of the target language. Thus, we propose that toolkit will be the parser of the target language.

iv. Discussion

The extended model will adopt the Method Judge as the heart of the MT. It must ensure the precision of the translation between the Arabic language and English language according to two standards: the precision of the sentence grammar and

structure, and the precision of the words meanings which the MT writes during the translation.

Naturally, the Arabic Parser is different from the English one. Therefore, we assume that there must be a source language identification process that can be triggered automatically even though the user didn't till the machine what is the source language. The parser of the source must not be the parser of the target. Then the source language sentences will be passed through a Natural language processing toolkit. After that, the Method Judge must start working to give us the precise translation according to its morphological and semantic analysis. One of the included ontologies at a time will be chosen according to the source language. The other may not be used because the rest of the translation can be made by the statistical translation effectively. Once it is ready, the aggregated result will be sent to the target language parser to correct the structure of the resulted sentence before showing the final result to the end user.

Like the translation in all the MT papers which was proven mathematically and the results were measured using the F-measure, the results of the proposed machine must be tested using that measure, which can be calculated as the following [1]:

$$\text{Precision} = \frac{number\ \ of\ \ correctly\ \ translated\ \ \ queries}{number\ \ of\ \ queries\ \ \ generated\ \ \ by\ the\ system}$$

$$\text{Recall} = \frac{number\ \ of\ \ corre\,ctly\ \ translated\ \ \ queries}{number\ \ of\ \ testing\ \ \ queries}$$

$$\text{F-measure} = \frac{2 \times Precision\ \ \times Recall}{Precision\ + Recall}$$

v. Conclusion

Semantic Web technologies help us to solve many problems in trading, learning and communication. Building a statistical Machine-Translation was inefficient to give us the correct and right meaning of the sentence and its words in the context. Semantic Analysis with all of its partitions gives us a perfect solution to get a better translation in comparison with some statistical ones. There were many successive semantic MTs that translate between English language and many other languages. The ones which depend on specifying the meaning of the homographs from the context were much more powerful. We are trying to develop a semantic MT between Arabic and English languages using the facilities provided by the semantic web. This machine will depend on the latest extended Arabic ontology, or set of ontologies, and also an English one to provide an instant translation from Arabic to English and vice versa. In this research we highlight a proposed model for a MT. We also assumed that this MT will provide an enhanced translation in comparing with the results of the existent instant translators.

As a future work, if the ontologies are not as comprehensive as needed, we suggest making an extension from all the available ontologies and to add more terms if there's still a need. We also suggest develop this MT to support more languages. The MT must also be optimized to translate long texts and Arabic artistic texts.

References

[1].I. AlAgha, and A. Abu-Taha. "AR2SPARQL: An Arabic Natural Language Interface for the Semantic Web." *International Journal of Computer Applications*, 125.6, 2015

[2].I. AlAgha, and A. Abu-Taha. "An Ontology-Based Arabic Question Answering System" *Central library of Islamic Univeristy of Ghaza*, 2015

[3].I. AlAgha.. "Using Linguistic Analysis to Translate Arabic Natural Language Queries to SPARQL." *arXiv preprint arXiv:*1508.01447, 2015

[4].A. Alqudsi, N. Omar, and K. Shaker. "Arabic machine translation: a survey." *Artificial Intelligence Review,* 42(4), 2014, pp. 549-572.

[5].Y. Al-Raheb, A. Akrout, J. van Genabith, J. Dichy. "DCU 250 Arabic Dependency Bank: An LFG Gold Standard Resource for the Arabic Penn Treebank" The Challenge of Arabic for NLP/MT at the British Computer Society, UK, 2006, pp. 105-116.

[6].D. Moussallem, and C. Ricardo. "Using Ontology-Based Context in the Portuguese-English Translation of Homographs in Textual Dialogues." *arXiv preprint arXiv*:1510.01886, 2015.

[7]. W. Weaver. "Translation." *Machine translation of languages,* 14,1955, pp. 15-23.

[8]. P. F. Brown, J. Cocke, S. A. D. Pietra, V. J. D. Pietra, F. Jelinek, J.D. Lafferty, and P. S. Roossin. "A statistical approach to machine translation." *Computational linguistics,* 16(2), 1990, pp. 79-85.

[9]. J. Hutchins. "Machine translation: History and general principles." *The encyclopedia of languages and linguistics*, 5, 1994, pp. 2322-2332.

[10]. Y. S. Chan, H. T. Ng, and D. Chiang. "Word sense disambiguation improves statistical machine translation." *Annual Meeting-Association for Computational Linguistics* 45(1), 2007, p. 33.

[11]. R. Gupta, N. Joshi, and I. Mathur. "Analysing quality of english-hindi machine translation engine outputs using Bayesian classification." *International Journal of Artificial Intelligence and Applications*, 4(4), 2013, pp. 165-171.

[12]. E. Ditters. "A formal grammar for the description of sentence structure in modern standard Arabic." *EACL 2001 Workshop Proceedings on Arabic Language Processing: Status and Prospects*. 2001.

[13]. E. Othman, K. Shaalan, and A. Rafea. "A chart parser for analyzing modern standard Arabic sentence." *Proceedings of the MT summit IX workshop on machine translation for semitic languages: issues and approaches*. USA, 2003

[14]. A. Ramsay, and H. Mansour. "Towards including prosody in a text-to-speech system for modern standard Arabic". *Computer Speech and Language*, 22, 2007, pp. 84-103.

[15]. Z. Zabokrtsky, and O. Smrz. "Arabic syntactic trees: from constituency to dependency". *Proceedings of the tenth conference on European chapter of the Association for Computational Linguistics,* European Chapter of the Association for Computational Linguistics EACL 2003, pp. 183-186.

[16]. S. Green, and C. D. Manning. “Better Arabic parsing: Baselines, evaluations, and analysis.” *Proceedings of the 23rd International Conference on Computational Linguistics*, Association for Computational Linguistics, 2010, pp. 394-402

[17]. D. M. Bikel. “Intricacies of Collins' parsing model.” *Computational Linguistics*, 30(4), 2004, 479-511.

[18]. L. Tounsi, M. Attia, and J. van Genabith. “Parsing Arabic using treebank-based lfg resources.” 2009.

[19]. W. Zaghouani, N. Habash, O. Obeid, B. Mohit, H. Bouamor, and K. Oflazer. “Building an Arabic machine translation post-edited corpus: Guidelines and annotation”, *International Conference on Language Resources and Evaluation (LREC 2016).* 2016.

[20]. C. Shi, and H. Wang. “Research on ontology-driven Chinese-English machine translation.” *2005 International Conference on Natural Language Processing and Knowledge Engineering*. IEEE, 2005

[21]. E. Seo, I. S. Song, S. K. Kim, and H. J. Choi, “Syntactic and semantic EnglishKorean machine translation using ontology.” *Advanced Communication Technology, 2009. ICACT 2009. 11th International Conference on*. IEEE, 3, 2009.

Nearest but Available: Proposed System for Enhancing User Location Awareness.

Sameer Yaseen
Department of Computer Science, Yemen Academy for Graduate Studies
Sameer.yaseen8@gmail.com

Abstract

Our choices in everyday transactions are based on nearest places. When we are hungry we look for the nearest restaurant. When we are driving and want to park we just look for the nearest park. From people past experience of places they can identify the nearest place, but they cannot check the place availability. In this application we provide a solution for service availability. We build a web application based on user location and using Google Map API, PHP, and MySQL database to help in finding the nearest and available services around user. This system has been tested and it shows a promised results.

Keywords: Location based services; location based search, Google Map API, GPS, and GIS

1. Introduction

People usually prefer nearest places for taking services. Location Based Services (LBSs): are applications or services that provide services depend on devices locations [1]. Searching nearest services is possible today through many applications [2], [3], [4]. Looking for available services take long time and effort, since availability is a sophisticated problem. No single solution fit all service providers. Each service provider has its own measurements for availability. For example ATM machines availability measurement is the amount of money still inside the machine. Restaurants availability measurements depend on the working hours and number of customers that could be served at the same time. So we need a method for checking whether the customer is going to take the service looking for or not before moving to the place. As illustrated in figure 1 there are many service providers around us. Some of those service providers are available, and some of them are not. Each color represent service category. None available services also represented in a different color. So nearest place is not always the best place for getting services. We need to know place availability for better user location awareness. Here are some researches related to this problem. Nearest Available Parking lot Application (NAPA) is an application based on GPS system to help in finding the nearest parking lot in the campus for professors, students, and visitors [2].

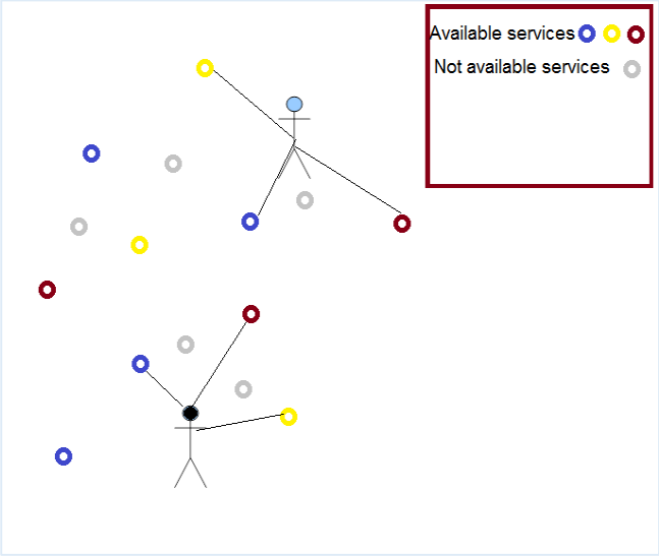


Figure 1: Services around us

In NAPA: central server used to store parking lots and their corresponding locations. The server responds to cars requests with nearest available parking lot and reserve spaces when used. When a car leaves the parking the server releases its reservation space. This project is limited with parking lots inside a university. Google Map nearby search [5] provides a search for nearest places around the user. This type of search provides some sort of availability based on place working hours. It shows the service provider's current status such as working now or close along with its working hours. WEi is an IOS app under development [6]. In this app user gender, location information, and age are shared to other users. When searching around places the app get that places number of users, gender, and ages. WEi is good however its user centric. It would not be accurate. For example, if some users are hanging out around the place, they will be counted as if they are taking the service. In addition, it will be difficult to check place working hours. Furthermore some places like banks have many services inside, while this type of apps fit one service at place. Similar research for this problem is what is done by Waqar Mirza M et al [7]. They used multi agent approach to solve nearest availability problem. They divide their system to three agents. The first agent is the interface to users which called User Interface and Query Agent (UIQA). Users send their requests to this agent. The second agent called Data Management Agent (DMA) which provides updated information about the searched service. The last agent called Crawling and Data Fetching Agent (CDFA). The system works as follow: UIQA forward user request to DMA, and the DMA search the database if the information in the database up to date then the DMA respond with the information from the database if the information is not up to date then the DMA send request to CDFA to crawl the service provider and respond with latest information. The previous systems either limited for a specific service or they will not work for all scenarios of availability. For example in some cases when service provider's about

closing. If we depend on these types of search the system will show the service provider as available; however the user will arrive after service provider closed his service. In our system we solve this problem by subtracting the time required to reach the service provider from the service provider remaining time to close and give suggestion to the user if the desired service could be taken or not. Another concern is the number of customers who are recently taking the service. The nearest bank branch to you maybe full with customers. If we depend on the previous systems it will shows the bank branch as available; however the user might not get their service due to number of customers. Here we will present a method for checking the nearest searched service availability percentage and give suggestion to user about the best place to take the service. In our system we are going to use PHP and MySQL along with Google Map API to solve this problem. We are going to build a system that will save people time and will help them for taking good services.

2. Nearest problem

Given a set of services that reflects service providers sp. So we need to divide these services into different categories depends on service type. For example Restaurants as category, Parking is another category and so on. As illustrated in table 1 each service in the service table have category identifier, latitude, and longitude.

Column	Type	Null	Default
service_id	int(10)	No	
user_id	int(10)	No	
service_name	varchar(60)	No	
catogry_id	int(10)	No	
Address	varchar(80)	No	
service_details	Text	Yes	NULL
Service_Lat	float(10,6)	No	
Service_lng	float(10,6)	No	
Max_customers	Int(10)	No	
availablity_page	varchar(250)	Yes	NULL

Table 1: Service table

Given a user location user-lat, user-lng, and search-radius for limiting search result in specific range. We need to locate the nearest services providers in the search-radius. So we need to calculate the distance between user location and each services in the category using haversine formula:

$$a = \sin^2(\Delta\varphi/2) + \cos\varphi_1 \cdot \cos\varphi_2 \cdot \sin^2(\Delta\lambda/2)$$
$$c = 2 \cdot \operatorname{atan2}(\sqrt{a}, \sqrt{(1-a)})$$
$$d = R \cdot c$$

Where φ is latitude, λ is longitude, **R** is earth's radius (mean **radius** = **6,371km**)[8],[9].Then order them according to distance as in the below SQL query:

$query = **sprintf("SELECT** address, service_name, service_lat, service_lng, **(6371 * acos(cos(radians(**user-lat**)) * cos(radians(** service_lat **)) * cos(**

radians(service_lng **) - radians(**user-lng**)) + sin(radians(**user-lat**)) * sin(radians(**service_lat **)))) AS distance FROM services where** catogry_id =**$category HAVING distance <** search-radius **ORDER BY distance LIMIT** 0 , 20**");**[10].

3. Availability problem

Availability problem has many cases depend on type of service. Some services are limited by working hours, and or number of customers (NOC) that can be served at the same time. Each service provider has its own (threshold) for NOC that could be served at the same time. When the NOC is greater or equal to threshold then the service provider is not available. Also, for the working hours, each business has its own official working hours that may differ from other businesses. When the person look for service provider out of its official working hours it should appears as unavailable. So service providers have to provide their working hours for the system, and the threshold of customer which could be served at the same time. Here we ask the service providers to provide their working hours and threshold of customer in the registration process. Also some service providers have advanced availability information which may differ depend on type of service such as current customers, average waiting time, and average service time. Some services are limited by spaces, so the service provider will share the reserved space and unreserved space it has. The system will show unreserved space as availability percentage. Some services are limited by money like ATM machines. ATM machine will share the amount of money still inside. If service provider has advance options then we ask him to provide its URL file that would be used to obtain advance options. Figure 2 illustrate the proposed system.

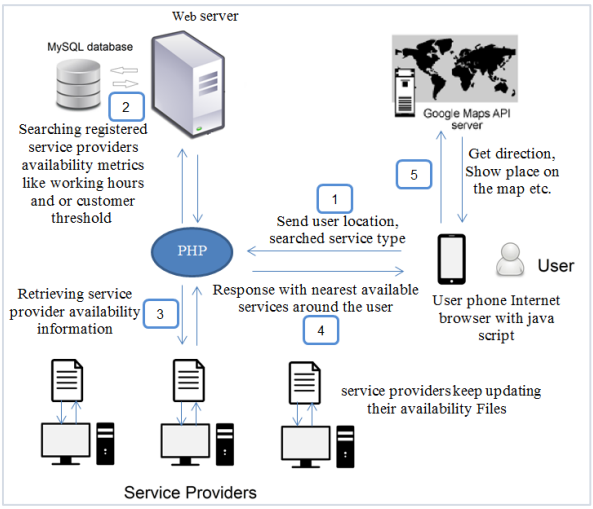


Figure 2: Proposed system.

The user selects type of service to search available services around. The system gets user location using geo location [11]. The system then search service providers around

the user. The system filters the searched result according to working hours as in the below algorithm..

Algorithm hour's availability:

```
For all element ∈ nearest result do {
    If user-time between service
    Provider start-time and end-time
    Then {
    If travel time < remaining service time Then {
    Proposed the result to user;
    } else {
    Service-provider Availability=false;
    }
    } else {
    Service-provider availability=false;
    }
    }
}
```

The system then proposed the result to user along with advance options. From the search result the user has the ability to fetch each service provider's advanced availability options separately.Travel time will be collected from Google Map API [12]. There are many options in Google Map API, but the request query for getting travel time must include user location, desired service location, driving mode, and departure time. Google Map API will response with several elements one of them called rows which contain duration element.

4. Implementation

Based on the research that has been done, it ended up with the creation of actual web application, which could be accessed by several devices that has supported browsers with enabled JavaScript.

4.1. Application design

As illustrated in figure 3, the web application consists of the main page which offers the login and register ability for service providers. The main page also offer search nearest available service for unlogged users.

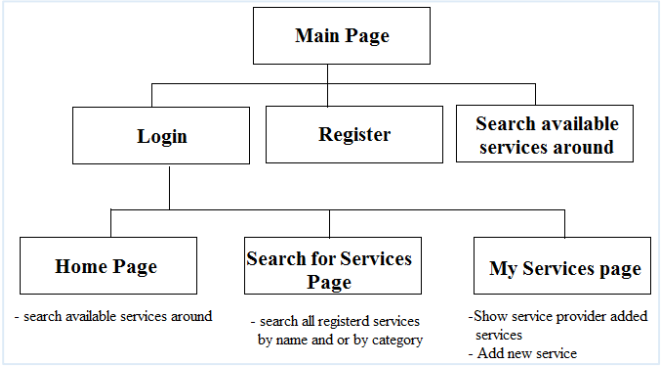


Figure 3: Application design

The home page offers search nearest available services for service providers. Figure 4show the home page and links to other pages after logging in.

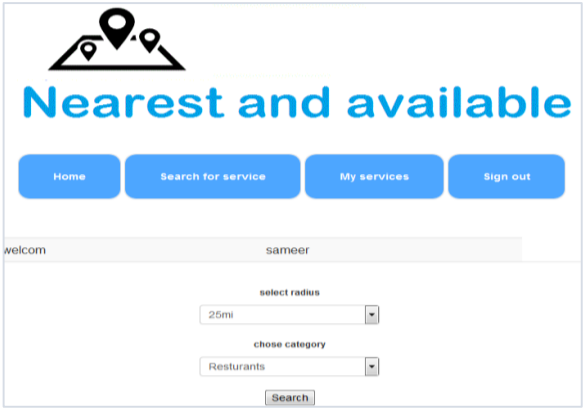


Figure 4: Home page

Service provider have the ability to explore its services from my services page, and add as many services as needed as illustrated in figure 5. Search for services page is advance search for services providers to search all registered services by service category and or by service name.

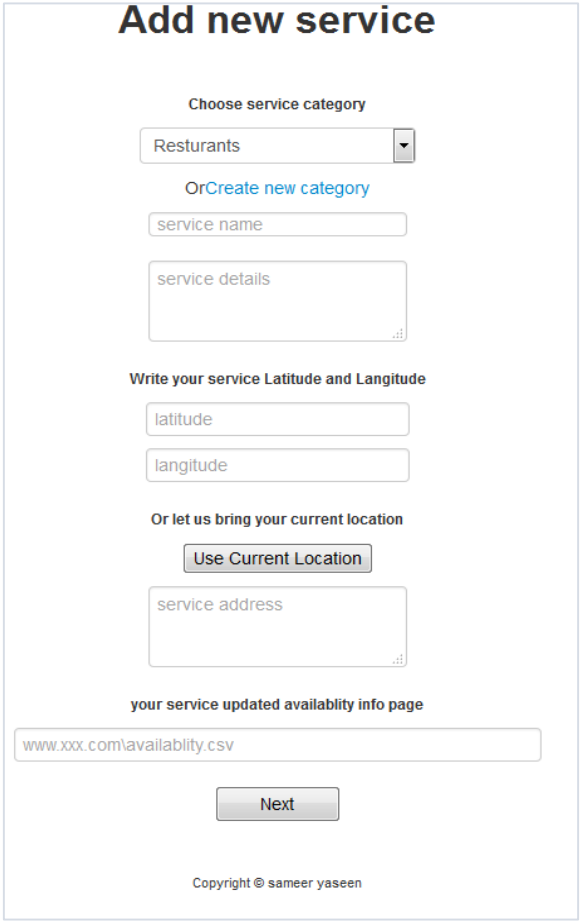


Figure 5: Add new service form

4.2. System test

The system has been tested at Azzahra Enterprise network. As illustrated in figure 6 wamp server 2.4 installed along with php version 5.4.12, and MySQL version 5.6.12 on a laptop with windows 7 64 bit operating system which used as a web server to hold application requests from user devices.

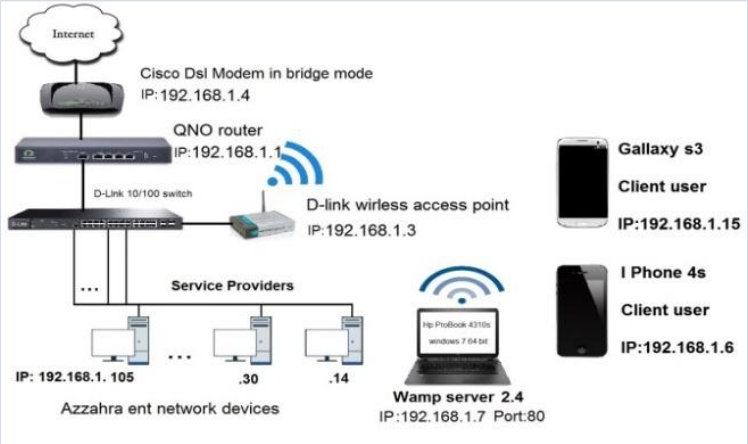
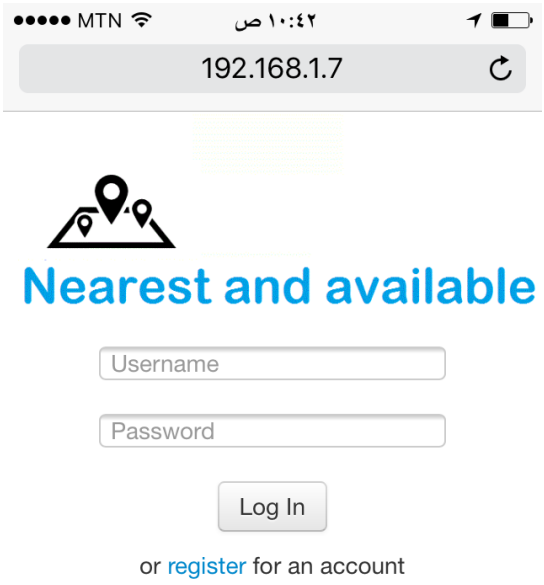


Figure 6: System testbed

Three desktop computers used as service providers. IPhone 4s , and Galaxy s3 used as clients for application test. As shown in figure 7 an http connection instantiated through a browser with the server URL 192.168.1.7. The server then responded with application main page which include embedded java script method for obtaining user location when needed.



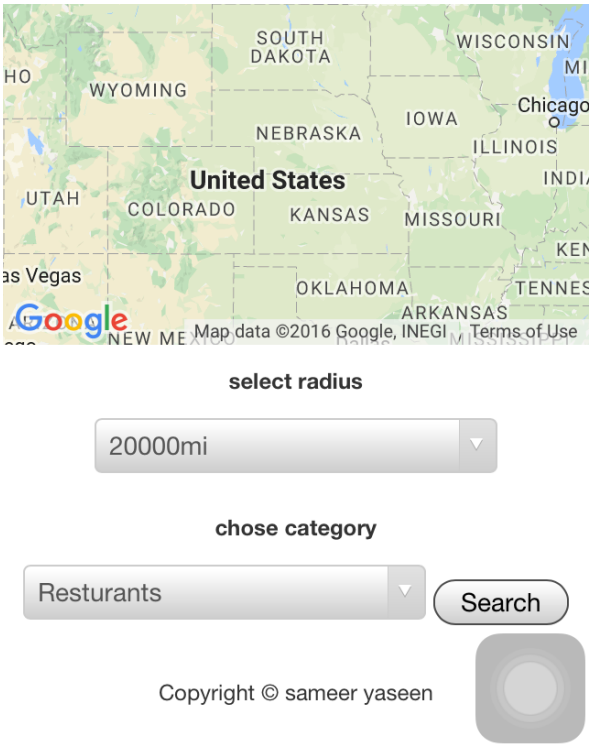


Figure 7: Login page

The user then select radius and chose the category and click search button. The browser send user location along with selected category and search radius. The server use the SQL query stated in nearest problem and then filter this query according to hour's availability algorithm stated on availability problem and then proposed the result to the user as shown in figure 8.

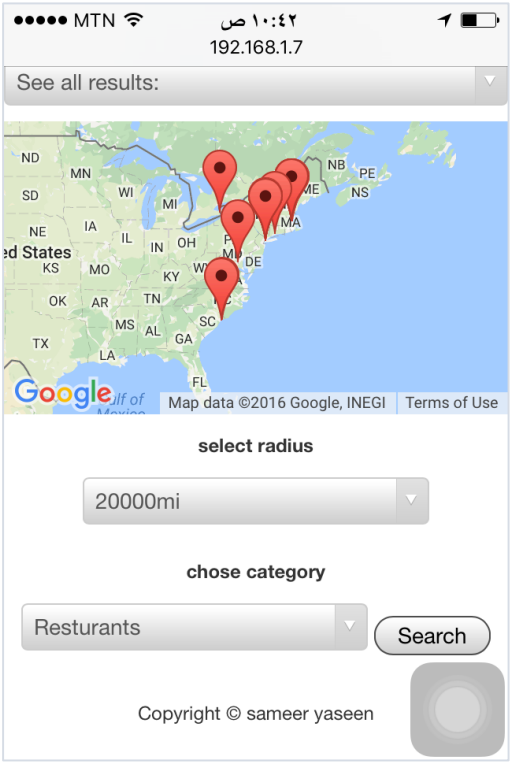


Figure 8: Searching result

As shown in figure 9 the user then have the ability to interact with each result separately. After user select service from the list , two buttons appeared one for draw direction of selected service and the other for fetching advance options if service provider have advance options.

4.3. Test result

Based on the test that has been done, the application works fine in both client devices, and it should works well in other devices; however the user have to use supported browser for geolocation method. Here are group of supported browsers from w3c[11]. During the test the clients could hardly ever establish a connection with the server while other company devices using network resources, so unused company devices communications prevented during the test.

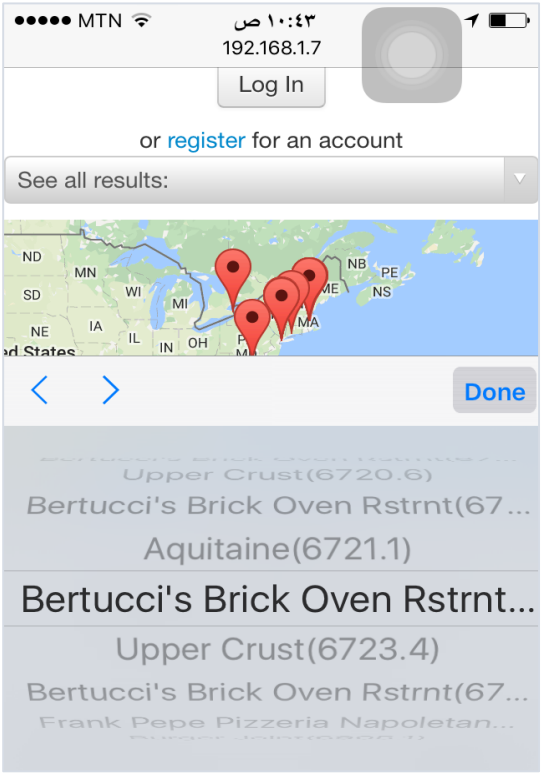


Figure 9: User interaction with search result

5. Future work

Service availability is a good measurement for enabling user to have better understanding of places around. However; to make sure that the user is going to get the service looking for, the system has to offer reservation ability for services. Reserving service is our future work. As seen, the queuing systems are expanded dramatically in most businesses places. Queue systems allow persons to select the desired service inside the place and prints small sheet of paper which include person turn number to get the service, and number of service applicant before him. In our future work we will allow the user to interact with service provider's queue system and reserve service remotely through our application. When the user arrives, he will be asked to show its reservation number to get the service.

6. Conclusion

Looking for services around us is very easy today because many apps provide this service; however looking for available services is a big challenge. Our time is very important and need to be saved in looking for available services. Our system provides this ability.

6. References:

[1] Amit Kushwaha, and Vineet Kushwaha,"Location Based Services using Android Mobile Operating System", *International Journal of Advances in Engineering & Technology*, Mar 2011.© IJAET.

[2] Hae Don Chon, D. Agrawal, and A. El Abbadi, "NAPA: Nearest Available Parking lot Application", *IEEE*, Department of Computer Sci., California Univ., Santa Barbara, CA, USA,2002.

[3] O. A. Ibrahim and K. J. Mohsen, "Design and Implementation an Online Location Based Services Using Google Maps for Android Mobile," *International Journal of Computer Networks and Communications Security (CNCS)*, vol. 2, pp. 113-118, 2014.

[4] M. Singhal and A. Shukla, "Implementation of location based services in Android using GPS and Web services," *IJCSI International Journal of Computer Science* Issues, vol. 9, pp. 237-242, 2012.

[5] https://www.google.com/maps, cited at 24/3/2016

[6] Krishna Raj Bagale,"Development of a Geolocation App in IOS Platform and its Market Analysis", Bachelor's thesis, Information Technology, Oulu University of Applied Sciences, spring 2016.

[7] Waqar Mirza M, Ana María Martinez-Enriquez, Aslam Muhammad, Afraz Z. Syed, and Arslan Lodhi," Intelligent system for searching nearest service using multi-agent approach",2013.

[8] https://en.wikipedia.org/wiki/Haversine_formula, cited at 13/10/2016.

[9] http://www.movable-ype.co.uk/scripts/latlong.html , cited at 13/10/2016.

[10] http://www.plumislandmedia.net/mysql/haversine-mysql-nearest-loc/ ,cited at 13/4/2016.

[11] http://www.w3schools.com/html/html5_geolocation.asp, cited at 13/4/2016.

[12] https://developers.google.com/maps/documentation/distance-matrix/intro , cited at 22/4/2016.

Comparative Study Between Three Approaches in Ontology Mapping

Suad Mohammed Othman
Computer Faculty Sana'a
University Sana'a, Yemen
Suad.m.othman@gmail.com

Dr.Fadl M. Ba-Alwi, Ph.D
Computer Faculty Sana'a
University P. O. Box 1247
dr.Fadlbaalwi@gmail.com

Abstract

Due to an increased use of ontology applications in different communities a growing number of ontologies are created by different organizations and individuals. Further, ontology users or developers want to integrate or use other ontologies to solve the problem. Furthermore user or developers want to find the correspondences between concepts in multiple ontologies. However the heterogeneous of data makes integrate and use multi ontologies not efficient. Ontology mapping is required for combining distributed and heterogeneous ontologies. Ontology mapping is important for different domains such as semantic web, data integration, information exchange, data warehousing and information retrieval to solve problems of heterogeneous data. In this paper we aim to compare between three approaches: KSMS, HSOMap and fuzzy ontology mapping systems based on some criteria's such as Matching techniques and aggregation techniques. This study displays the review of three approaches to help us to select the best approach for ontology mapping.

Keywords*:* Semantic web, ontology mapping, ontology matching, element and structure level matching.

Introduction

Increase in web and distributed computing leads access to a large number of independently created sources. For the effective utilization of the heterogeneous data, various processing techniques are required to resolve the information heterogeneity [10]. Information heterogeneity comprises of syntactic, structural and semantic heterogeneities. There is a need for more sophisticated techniques to solve the syntactic, structural and semantic heterogeneity. Various approaches have been suggested as solutions for

solving semantic heterogeneity [11].Ontologies are the main component of semantic web to solve the semantic heterogeneity problem.

Ontology is a formal, explicit specification of a shared conceptualization [9]. Ontology matching is a process that takes two ontologies (one source ontology and one target ontology) and finds semantic relationships between ontology entities using matching algorithms [1].
Ontology mapping is a high-level specification of the semantic relationships between the elements of two ontologies [2], and the relationships are expressed by equivalence, mismatch, overlapping, more general and less general [3]. Ontology matching is a required task for ontology mapping. The result of ontology matching is a numeric value between 0 and 1, and ontology mapping is true/false.

The main types of ontology mapping approaches are structure methods, terminological based methods, instance based methods, semantic methods, hybrid methods and ontology mapping using background knowledge[10]. Each of these categories can be subjected to separate analysis but they are not entirely independent of each other. Some system uses multiple methods for mapping such as [5] that use RDR that is a machine learning technique that represents the exception rules in a concise manner in order to keep all general rules consistent [12].

In almost approaches for ontologies mapping basic step to do ontology mapping this step is: Entity extraction by using Jena. Second step Pre-processing that applies text processing techniques such as tokenization, stop words removal, synonym, stemming and translation. Third step Calculation of similarities using proposed method and finally aggregation of similarities and post processing.

In the following sections we comparative between three different approaches for ontology mapping : Knowledge based Schema Matching System (KSMS) [4], second approach Harmony Search based Ontology Mapping (HSOMap) [6], and fuzzy ontology mapping [7] based on some criteria's that used in different basic step in ontology mapping operation .

The Rest of this paper is organized as section 2 provides related work; section 3 presented discussions about three different approaches for ontology mapping based on input, output, preprocessing techniques, matching techniques, aggregation techniques and performance criteria's. and section 4 conclude the paper, and finally my opinion.

Related work

There is a lot of research in ontology mapping field such as:

KSMS uses Hybrid-RDR [5] approach that combines both machine learning and incremental knowledge engineering approaches for matching entities at the element level using ontology features.in this paper researcher use Element level matching for ontology mapping but element level matching does not only give proper results for ontology mapping as it only considers matching names of entities not considers any meaning.

Incremental knowledge engineering approach [8] Censored Production Rules (CPR) based Ripple-Down Rules (RDR) for schema mapping. The approach uses the features created by the combination of string similarity metrics and text processing techniques for creating rules. However, the limitation of the approach is that it is time-intensive to create rules for mapping entity pairs one by one at the element level.

KSMS system [4] that allows users to correct and validate the matching results automatically. Proposed approach is based on Hybrid Ripple-Down Rules (RDR) that combines machine learning and knowledge acquisition approaches. This approach combines decision tree, J48 and incremental knowledge engineering approach, Censor Production Rules (CPR) based Ripple-Down Rules (RDR). The advantages of this approach: only one classification model of decision tree is used so it does not generate any over fitting problem and rules are not pre-defined and can be reused for other domains. Third, this approach not need time consuming. Finally it does not need a specific threshold value to filter result. Therefore, users do not need to gather proper knowledge about the domain. At the structure level matching, KSMS uses Similarity Flooding algorithm to match the hierarchical structure of a full graph. The final mapping result is produced by applying aggregation function such as Harmony-mean on the results of both levels.

HSOMap[6] approach that proposed symbiosis between harmony search algorithm and combinatorial methods for ontology mapping. this approach effectively finds a near-optimal mapping for two input ontologies. In this approach, proposed use of many kinds of rating functions, which are also called base matchers to evaluate the similarity of entities. Each base matcher captures the similarity between entities from a different perspective and is able to exploit the available side information about the entities effectively by used combinatorial technique. Also, a novel weighted harmonic-mean method is proposed to aggregate different metrics into a single similarity metric among all pairs of entities from two ontologies. After obtaining the combined similarity metric between ontological entities, a discrete harmony search algorithm is proposed to extract the best alignment.

Jie,Bo-Ju etal[7] proposed fuzzy approach for all ontological concept pairs, the four individual similarities are calculated by concept similarity

algorithms based on fuzzy knowledge. This author proposed new algorithms exploiting concept similarities on for individual levels including: concept name, properties, instances and structure .As for each individual similarity algorithm, its harmony, reliability and f-measure are worked out respectively. These similarities are aggregated and the mappings between ontologies are extracted by using all methods of SVM and Neural network. But in this system not use combine similarity methods that can improve ontology mapping.

Discussion

There are many systems that developed to solve ontology mapping. In this section we compare between three different approaches for ontology mapping these approaches are:

Knowledge based Schema Matching System (KSMS) [4]

Second approach is Harmony Search based Ontology Mapping (HSOMap) [6] that proposed symbiosis between harmony search algorithm and combinatorial methods for ontology mapping, third approach fuzzy ontology mapping [7].

Table 1 illustrate basic different between this three approaches based on some criteria's that is the most widely used for comparative between ontology mapping approaches and consider as basic operations that effective in ontology mapping processes such as : input criteria that represent the basic for any processes in ontology mapping the input can be OWL or RDF or XML file. Text possessing technique that used in preprocessing that involves algorithm which parses the input ontologies and extract element or ontology's entity and performance measures such as: precision, Recall, F-measure that calculate by formulas [6]:

$$\text{Precision} = \frac{\#correct\_found\_mappings}{\#found\_mappings} \quad (1)$$

$$\text{Recall} = \frac{\#correct\_found\_mappings}{\#existing\_mappings}. \quad (2)$$

$$\text{F-measure} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (3)$$

Table 1: illustrate basic different between three approaches.

Systems criteria's	HSOMap	KSMS	Fuzzy
input	OWL	OWL	OWL
Output or cardinalit-y	1:1 alignments	1:1 alignments	1:1 alignments
Text processing techniques	Stemming, Tokenization, Abbreviation, Synonym Removing stop word	Tokenization, Abbreviti-on, synonym	Stemming, Tokenization, Edit distance
Matching techniques	Structural, Lexical.	Element, Structural.	Element, Instance, Structural, Property.
terminological	WordNet	WordNet	WordNet, Vector distance
aggregate	weighted harmonic-mean method, Sigmoid function	Harmony-mean	Sigmoid function
Performnce	Precision= .77 recall=.75 F-measure= .77	Precision =.89 recall =.81 F-measure= .84	precision=.85 recall=.84 F-measure=.84

These systems were used 1:1 cardinality mapping that can be speed but not cover all probabilities. To cover many probabilities one suggestion using n: m cardinality mapping instead of using 1:1 cardinality.

In figure (1),(2),(3) and (4) : illustrate architecture different between three approaches Knowledge based Schema Matching System (KSMS) architecture show in figure 1 and the Hybrid-RDR that used in KSMS show in figure 2 . Harmony Search based Ontology Mapping (HSOMap) architecture show in figure 2 . fuzzy ontology mapping architecture show in figure 3.

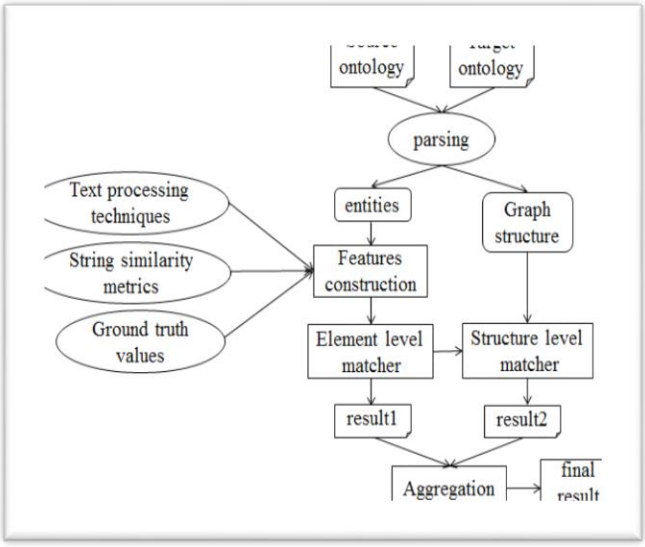


Figure 1. **KSMS Architecture**[4]

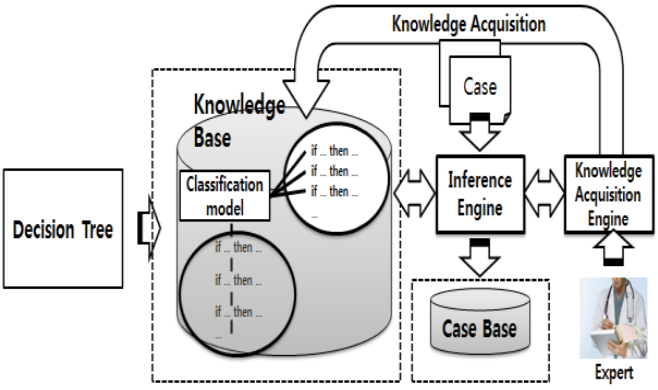


Figure 2 Hybrid-RDR Architecture[5]

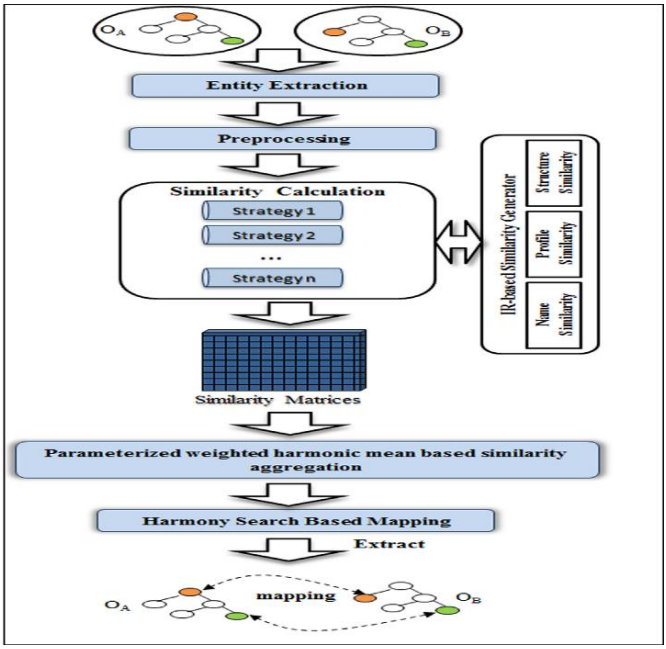


Figure 3. **HSOMap Architecture**[6].

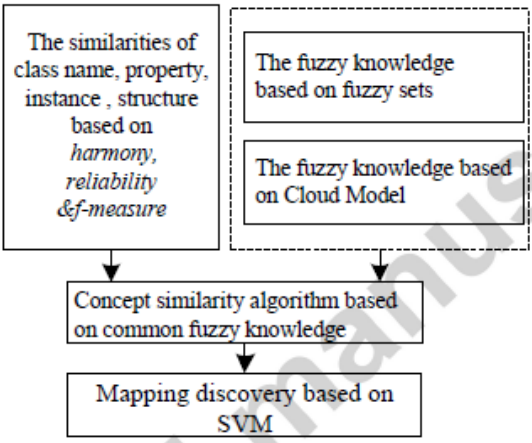


Figure 4. **fuzzy system Architecture**[7].

Conclusion

Ontology mapping is involved with many fields like Machine learning, database, linguistics and also necessary for tasks such as search, integration, query answering

and data translation. Based on the importance of this field, we have reviewed the three techniques in ontology mapping and made comparisons among them. Ontology mapping still a challenging area so we need design more scalable algorithm to be able to map large scale ontologies.

My opinion

Ontology mapping can improve by using domain knowledge and machine learning such as Ripple-Down Rule (RDR) that can give us a lot of advantages to solve some challenges in ontology mapping and used algorithms such as harmony search to discover optimal solution we believe if combine between two approaches HSOMap and KSMS in some steps can improve Performance .

.

References

[1] Faria, D., Pesquita, C., Santos, E., Palmonari, M., Cruz, I.F. and Couto, F.M. The agreement maker light ontology matching system. In On the Move to Meaningful Internet Systems: OTM 2013 Conferences, 2013, Springer.

[2] Cate, B.T., Dalmau, V. and Kolaitis, P.G. Learning schema mappings. ACM Transactions on Database Systems (TODS), 2013. 38(4): p. 28.

[3] Shvaiko, P. and Euzenat, J. A survey of schema-based matching approaches. In Journal on Data Semantics IV, 2005, Springer, p. 146-171.

[4] Sarawat A, Yang Sok Kim , Byeong Ho Kang. Adapting A Knowledge-based Schema Matching System for Ontology Mapping. ACE '16 Canberra, ACT Australia , 2016 ACM 978-1-4503-4042-7/16/02

[5] Anam, S., Kim, Y.S., Kang, B.H. and Liu, Q. Schema Mapping Using Hybrid Ripple-Down Rules. In the Thirty-Eighth Australasian Computer Science Conference, ACSC 2015, CRPIT: Sydney, Australia, p. 17-26.

[6] Rana Forsati, Mehrnoush Shamsfard .Symbiosis of evolutionary and combinatorial ontology mapping approaches . 0020-0255/© 2016 Elsevier , Information Sciences 342 (2016) 53–80

[7] Jie Liu, Bo-Ju Zheng, Li-Ming Luo, Jian-She Zhou, Yuan Zhang, Zheng-Tao Yu. Ontology Representation and Mapping of Common Fuzzy Knowledge. © 2016 Elsevier , Procedia Computer Science 00 (2016) 1–19

[8] Anam, S., Kim, Y. and Liu, Q.: Incremental Schema Mapping. Knowledge Management and Acquisition for Smart Systems and Services. 2014 Springer International Publishing.

[9] A Semantic Web Primer, third edition, Grigoris Antoniou and Frank van Harmelen, 2012

[10] Kaladevi Ramar and Geetha Gurunathan, Technical review on ontology mapping technique, Asian Journal of Information Technology 15 (4): 676-688, 2016

[11] Stuckenschmidt, H. and F. Van Harmelen, Information Sharing on the Semantic Web. (2005) Springer,New York, ISBN: 9783540205944, Pages: 276.

[12] F. M. Ba-Alwi and K. K. Bharadwaj, "Automated discovery of hierarchical ripple-down rules (HRDRs)", Proc of In the Twenty-third IASTED International Conference on Artificial Intelligence and Applications (AIA 2005), Innsbruck, Austria, February 14-16, 2005.

Survey of Semantic Annotation of Arabic Text

Marwa Al-hadi
Department of Computer Science,
Sana'a University,
Sana'a, Yemen
eng_marwa.alhadi@yahoo.com

Fadl Ba-Alwi
Department of Computer Science,
Sana'a University,
Sana'a, Yemen
Dr.Fadlbaalwi@gmail.com

Ibrahim Al-Baltah
Department of Information
Technology, Sana'a University,
Sana'a, Yemen
albalta2020@gmail.com

Ghaleb Al-Gaphari
Department of Computer Science,
Sana'a University,
Sana'a, Yemen
drghalebh@gmail.com

Abstract

Adding additional attributes to the semantic web is called semantic annotation, which is helpful in searching about a specific topic in the web. Arabic researches revolution that invoke to looking for a method that help Arabic researchers to access to Arabic resources in beneficial way. Semantic annotation is one of these methods, and many tools have been proposed to achieve this task. This study surveys semantic annotation tools. Therefore, this study performs as comparative analysis between some of mostly used Arabic and English tools for semantic annotation. We highlight some reasons that might cause the state of the art to pay less attention to annotate Arabic resources comparing to English resources. Furthermore, this study suggests to create a semantic annotation tool that support Arabic language and its rules, lexical analyzing, natural language processing automatically.

***Keywords*-Semantic Annotation; English Semantic Annotation tools; Arabic Semantic Annotation tools .**

INTRODUCTION

Notation is an old approach for writing notes on paper .It's like using highlighter for concentrate on specific words. It has been used by human-being from thousands of years before the revolution of computer and technology enhancement. Figure 1 demonstrates an example of the informal annotation [1].

Industry revolution, computer invention and Berners-Lee efforts have shifted the computer resource and interaction from human base into machine base through semantic web contribution. Therefore, filling the gap of syntactic web such as HTML into a semantic web was an argent need. Thus, semantic concepts make machine to machine interaction possible. The semantic concepts involve around the definition of domain classes, relationships among them, and individuals.

As a matter of fact, W3C suggests enhancing semantic web description by Semantic annotation.

Semantic annotation, which add extension attributes to web documents that increase the reusability. It improves the accuracy of semantic analysis by clearly determine the input, output, and operations of web resources. If semantic annotation is not used then web search is done in general knowledge base. The feature of full description increase the way of using semantic annotation. Consequently, it increases the ability to search for specific document.

Rich representation of semantic annotation of English text and Latin text is more interested than Arabic text. Semantic annotation research has less concentrate on Arabic language comparing to another language especially in semantic annotation part. Therefore the survey in this paper will mainly focus on semantic annotation of Arabic text. Furthermore, it aims at providing a good understanding of the state of the art on the topic, and to analysis and compare some annotation tools.

The reminder of this paper is organized as follow. Section II introduces semantic annotation. Section III shows related work about sematic annotation tools in brief with specific criteria of comparison. Section IV presents discussion of platforms. Section V contains conclusion and future work.

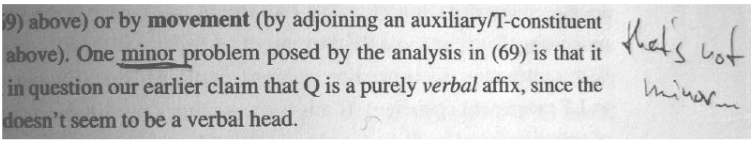

9) above) or by **movement** (by adjoining an auxiliary/T-constituent above). One minor problem posed by the analysis in (69) is that it in question our earlier claim that Q is a purely *verbal* affix, since the doesn't seem to be a verbal head.

Figure 1. Informal annotation obtained from[1]

Semantic Annotation

"The idea of having data on the Web defined and linked in such a way that it can be used by machines not just for display purposes, but for automation, integration and reuse of data across various applications" is called semantic web [2].

Enriching semantic web with full meaning using the concept of semantic annotation [3].As it has mentioned in [4], there are three types of annotation

mechanisms, which are Textual, Link, and Semantic .Adding comments to objects called Textual annotation, made linking of objects to contents is called link annotation and consist of semantic information for machine readable is called Semantic Annotation.

Semantic annotation have multiple definitions which define as the process of having various concepts such as things, places, people and assigning supplementary information to it in a given content[2]. And as the process of attaching metadata or ontology to text fragments in order to facilitate accessing of knowledge and retrieval process[5].

Previously, annotation process in details started from upload information structure of the database from the database owner to produce the sever side. Then annotator generate client side annotation in form of ontology. Mapping done between database side and client ontology which demands for annotator publishes the client ontology on the web[6].More specifically semantic annotation includes three steps to perform as follow: first: Identify of entity which generate meta-data using lexical analysis ,natural language processing to extract useful information from a given content. Second: Check entity dis-ambiguous by correction spelling, abbreviations management. Third: Annotation build by linking semantic meta-data into entities in document represent as RDF(Resource Description Framework) and OWL(Web Ontology language) [2].In developed way the process of semantic annotation way will be as a form of services using WSDL(Web services description language) that associate semantic with web services elements in two ways: First: Matching semantic with WSDL represented by finding semantic between elements. Second: Mapping semantic with WSDL represented by transformation rules[7] .

There are multiple tools that represent semantic annotation specially in English resources such as KIM, OntoMat, and Annotea. And limited number of Arabic annotation tools such asAraTation, Zaidi Gate, and AutoMatic Arabic Semantic Annotation Tool(AMASAT)[7].

RELATED WORKED

From hand writing notes on a paper using highlighting, summarization into a manual annotation tools until semantic annotation tools of machine-understandable .This work present a survey about some of the most famous tools that represent English text compare to limited tools of Arabic text representation.

Depends on most previous researches with additional effort to represent the tools as follow :

- Annotea[8,9,10]is first implemented by W3C's as Amaya. It is one of tools that represents semantic annotation which separate annotation such as comments, notes, explanations from contents of web documents .Annotation documents attached to contents documents. It's helpful for easily finding material when searching. One of the biggest challenge to it is supporting Arabic language. Multiple data from different severs that make a delay on its works rather than central point server .

- KIM(Knowledge and Information Management) [11,12] which generation semantics links between documents, domain models, and linked data .It's friendly graphic user interface with big data linking. KIM have the ability to analyze text and aware references to entities. Then matching it to the references using unique URI and specific description in knowledge base. The annotated done depends on URI of entity. Working with multiple server represent a big challenge for KIM and trying to enhance the concept of load balancing to be used .

- GATE [13,14]its java application comes for text annotation and huge data organization. Its base for multiple language such as English and Arabic. Its work as IE process, text analysis and content management. KIM works based on Gate and others Arabic have a vision of Gate but doesn't have available real one .

- AraTation(An Arabic Semantic Annotation Tool) [15] this tool works as text processing by normalize and clean text especially Hamza"ء" attention, words extraction to be present it as concepts and finally semantic annotation by mapping ontology to instances. This experiment done in Arabic news domain. This research is base for many of multiple researches. The biggest challenge that faced is

Arabic morphologies with multiple Arabic sentence format. But it's not available as a tool for modified and enhancement .

- Zaidi Gate[16]used for Arabic text extraction using named entities. By adding additional rule to Jape(Java Annotation Pattern Engine) to enable reorganization of terms. Still in work and need a lot of adding rules to Arabic language .Experiment done on Quran book specially on specific and limited text. If researchers thinking to enhance this experiment it will take a lot of time to represent all rules that present. Because Arabic language have a deep meaning and rules. We can adding [17] to this tool which will help in making summarization semantically.

- AutoMatic Arabic Semantic Annotation Tool(AMASAT)[18] for Arabic resources which using URL and corresponding ontology as an input and produce external annotation as an output using Span attribute. Using arability domain and unavailable of this tool it reduce the future work. One of the biggest challenge for using this tool is large documents with different concepts and relations difficult to represent. Demands enriching documents with additional lexical resources.

Even we have limited number of Arabic Annotation tools but it's still under research and not produce as available for future as a tool for used. We conclude and discuss each paper in a comparative way you can see that in Table 1 which configure comparative represent of different tool, represent challenges , advantages and dis-advantages .And according to Table 1 there is a huge gap for semantic annotation of web documents or resources specially text. Which mostly not available or we could say we have some annotation tools but we don’t have a free or available tool for use and enhance.

TABLE I. SURVEY SUMMARY

Platform	Language	Input format	Output format	Advantages	Disadvantages	Challenges	Availability
Annotea.	English	RDF	RDF	finding related material easily	Central point aggregation on server	Support Arabic Annotation	available
KIM	English	RDF	RDF, OWL	Free tool Faster in information retrieval, indexing and evaluation .	Complex work when it working with multiple server	Automatic load balancing on multiple server Working for large scale distribution system	available
AraTation	Arabic	URL of document	RDF	Colorful on annotation and represent on ontology and its base for multiple other researches.	Limited domain	Arabic Morphology Performance and time response	Not available
Zaidi Gate	Arabic	Document	Only annotate document in color	Help in information retrieval machine translation and automatic indexing.	Limited domain	Vast number of information manipulation	Available only at Leeds University
AMASAT	Arabic	RDF, OWL and N-TRIPLE.	RDFa and RDF	Produce Automatic semantic annotation document in high performance	Applied only of arability ontology or documents	Additional lexical resources. Represent a large documents with different concepts and relations	Not available

IV. discussion

In this study, we investigated that even there are multiple platforms that represent English text such as Annotea and KIM therefore, our finding that Annotea could be used for Arabic semantic annotation on simple sentences representation depends on its nature of finding easily related materials. Previous studies[11,12] have found on KIM that its complex work with multiple servers, but using the load balancing algorithm as a suggestion such as round robin could distribute balancing between servers and enhance its works. In comparison to Arabic semantic annotation platforms such as AraTation, Zaidi Gate, AMASAT which still as unavailable in general. We assume to upgrade AraTation performance and response time using manipulate Arabic rules language. In Addition for future we can crossbred between AraTation and AMASAT to produce better performance platform with enriching Zaidi Gate information retrieval and auto-indexing.

V. conclusion and future perspectives

Perfect definition of a semantic web is appearing in meta-data construction which represent in a semantic annotation. Although of availability of semantic Annotation in English language, it's still less attention of Arabic language resources even we have some but mostly not available. This work represents a comparative survey between multiple platforms of Arabic and English semantic annotation. For future work First, we suppose of creation a semantic annotation tool that supports Arabic language and its rules, lexical analyzing, natural language processing automatically. That's helping to support Arabic researchers to find Arabic resources using specific concepts. This tool will be and should be available for future and other development. Also, it will help in text summarization and translation. Second, we consider to hybrids between AraTation and AMASAT that produce automatic semantic annotation of Arabic resources with high performance.

REFERENCES

[1] M. Nagarajan, “Semantic annotations in web services,” in Semantic Web Services, Processes and Applications, Springer, 2006.
[2] "Web Naming and Addressing Overview (URIs, URLs, ...)", *W3.org*, 2016. [Online]. Available: https://www.w3.org/Addressing/.
[3] Yongxin Liao, Mario Lezoche, Herv´e Panetto, Nacer Boudjlida,"Why, Where and How to use Semantic Annotation for Systems Interoperability". 1st UNITE Doctoral Symposium, Jun 2011, Bucarest, Romania. pp.71-78, 2011
[4] Bechhofer, S., Carr, L., Goble, C., Kampa, S. and Miles- Board, "The Semantics of Semantic Annotation". In: Proceedings of the 1st International Conference on Ontologies, Databases, and Applications of Semantics for Large Scale Information Systems, 2002.
[5] Cunningham, H., Tablan, V., Roberts, I., Greenwood, M. A., & Aswani,"Information extraction and semantic annotation for multi-paradigm information". In Current
Challenges in Patent Information Retrieval (pp. 307-327). Springer, 2011.
[6] S. Handschuh, R. Volz and S. Staab, "Annotation for the Deep Web", IEEE Intelligent Systems, vol. 18, no. 5, pp. 42-48, 2003.
[7] S. Albukhitan, A. Alnazer and T. Helmy, "Semantic Annotation of Arabic Web Resources Using Semantic Web Services", Procedia Computer Science, vol. 83, pp. 504-511, 2016.
[8] Kahan, M. Koivunen, E. Prud'Hommeaux and R. Swick, "Annotea: an open RDF infrastructure for shared Web annotations", Computer Networks, vol. 39, no. 5, pp. 589-608, 2002.
[9] "annotea.org home page", Annotea.org, 2016. [Online]. Available: http://www.annotea.org
[10] S. McIlraith, D. Plexousakis and F. Van Harmelen, The Semantic Web - ISWC 2004. Berlin: Springer, 2004.
[11] B. Popov, A. Kiryakov, D. Ognyanoff, D. Manov and A. Kirilov, "KIM – a semantic platform for information extraction and retrieval", *Natural Language Engineering*, vol. 10, no. 3-4, pp. 375-392, 2004.
[12] "KIM Platform - Semantic Web Standards", *W3.org*, 2016. [Online]. Available: https://www.w3.org/2001/sw/wiki/KIM_Platform.
[13] GATE User Guide [Online]. Available: http://gate.ac.uk/sale/tao/
[14]H. Cunningham, D. Maynard, and K. Bontcheva, Text processing with gate. Gateway Press CA, 2011.
[15] L. M. Bin Saleh and H. S. Al-khalifa, “AraTation :
An Arabic Semantic Annotation Tool,” Proc. Int. Conf. Inf. Integr. Web-based Appl.Serv., pp. 447–451,ACM, 2009.

[16] S. Zaidi, M.-T. Laskri, and A. Abdelali, “Arabic Collocations extraction using Gate,” Int. Conf. Mach. Web Intell., pp. 473–475, 2010.
[17] F. Ba-Alwi, G. Gaphari and F. Al-Duqaimi, "Arabic Text Summarization Using Latent Semantic Analysis", *British Journal of Applied Science & Technology*, vol. 10, no. 2, pp. 1-14, 2015
[18] S. Al-Bukhitan, T. Helmy and M. Al-Mulhem, "Semantic Annotation Tool for Annotating Arabic Web Documents", *Procedia Computer Science*, vol. 32, pp. 429-436, 2014.

TSVCC:Tracking Service in Vehicular networks based on Cloud Computing Platform as a Service

Salah Mohammed Abduljalil
Faculty of engineering and Information technology,
University of Alandalus,
Salahm261@gmail.com

Fekri M. Abduljalil
Department of Computer Science,
Faculty of Educ., Arts, & Science,
University of Sana'a,
fmabduljalil@gmail.com

Abstract

Recently many Tracking system proposed in VANET as an applications to do specific function such as anti theft car systems ,vehicles monitoring systems,fleet management ,this type of applications depend on real-time vehicles' location ,we call it Location-Based Applications(LBA) or Location-Based Services(LBS) .

From the VCC top view point, this paper propose platform location tracking system for support built many LBAs in VCC environment. Authors propose optimized VCC architecture for improve bandwidth utilizationand decrease number of location update message.via handling impact increase number of LBAs in a single vehicle

The performance of the proposed system is evaluated on the basis of percentage of bandwidth consume and number of packets send to server vs number of LBAs.

Keywords: Vehicular Network; Vehicular Cloud Computing, Location Based Application.

1. Introduction

Recently Vehicular networking has become a unusual research area due to : (a) its specific applications such as efficient traffic management, road safety and infotainment[1].(b)Vehicles are considered as computers on wheels due to they are carrying more resources On Board Unite(OBU) such as communication systems ,storage and sensors[2].in Fig.1, OBU has many components consider as a resources on a vehicle,many researches proposed to develop ITS solutions.

Concept of Vehicular Cloud Computing (VCC) is solution shifted from Mobile Cloud Computing (MCC) for best utilities of a vehicular

network resources.The VCC is a new technology that has an unusual impact on traffic management, road safety and other application by on-demand using vehicular resources, such as computing, storage and internet for decision making and support customers by new services and applications [3][4].

Location-based applications (LBS/LBA) have seen fast growth in the last decade due to pervasive adoption of GPS enabled mobile devices, Location-Based-Application (LBA) /Location-Based-Service (LBS) is type of vehicular cloud computing services and applications depends on real-time vehicles location, LBA dose tracking vehicles location and maintaining an accurate up-to-date view of the entire network. Each a vehicle update its real-time location Abscissas "GPS" periodically to server, but this scheme have some challenges, storage, bandwidth overhead and server overhead. This called Location Update Problem (LUP) .

Many researches proposed Location update problem solutions, main goal was minimizing number location update message required between vehicles and servers if you want know more about location update problem and methods to solve it read [5], real-time vehicle's location is a cornerstone in several proposed LBS in vehicular cloud computing such as Pic on wheel as a service, Vehicle Witnesses as a Service [6], Video capture, Tracking systems, and traffic analysis.

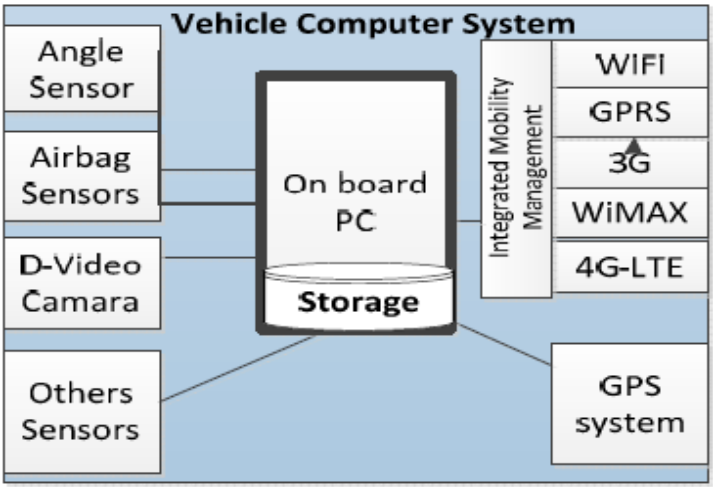


Fig. 1 On-Board Unit (OBU)

Note that in VCC environment what if single vehicle has more than one LBA? normally the LUP will be doubled by number of LBAs

installed on a single vehicle. Vehicular cloud computing LBAs needs a scalable and efficient tracking real time vehicles' location not just as an application also as a platform make others build LBS. In this paper, we propose a Tracking Service in vehicular Networks based on Cloud Computing(TSVCC).TSVCC platform is tracking real-time vehicles' location as a service in VCC environment it supports development many Location based applications, TSVCC improves bandwidth consume and number of location update message.via handling impact increase number of LBAs in a single vehicle

The remainder of this paper is organized as follows. In Section 2, related works are surveyed and we have discussed the problem formulation. The proposed solutionis introduced in Section 3. Section 4 presents the results and evaluation of the proposed schemes. Section 5 presents the conclusions and future work. Section 6 is reference.

2. Related work

In [3], authors presented state-of- art survey of vehicular cloud computing, they classified of a VCC's applications and services, they designed VCC architecture which we can use it as a VCC reference model, they showed open issues and challenges in VCC.

In [5] it is proposed a query-aware location update framework for Mobile Clients. For decrease number location update message send to location server. All the mobile clients must aware of query. In this scheme assume that mobile clients and the location server have a local copy of the same road network database .the focused was how mobile clients send little number of the location update message to location server to reduce bandwidth consume and server overhead, it take in account each a single vehicle has a single LBA, In fact, the VCC maybe a single vehicle has more LBA.

In [7] a Service-Oriented Security Framework for Vehicular Cloud Computing is proposed to support applications security in VCC environment, the author's idea based on this work from a different issue we deal with real-time vehicle locations

[11] Proposed a new vehicular cloud service. It is on demand utility of the available resource of vehicles to capture image and send it to

customer, after several year in future will see this service consider as a software application, pics-on-wheel service has some challenges how system can know where and when vehicles shots image, so that author proposed solution, each vehicle sending location update message periodically to location server. This scheme causesextra bandwidth consume and cause network congestion [5].

Problem formulation

Recently VCC open area to produce new kind of services that help human to get better and easy life on this world location based applications in VCC is faced with location update problem , many researchers provide solution to location update problem which take in account each a single vehicle contain a single LBA but this not enough in real-life. Note that, in current smart phone system is single phone has many LBAs .VCC has been shifted from MCC, So In real life a single vehicle may be contain more than one of LBA on the same vehicle . So the traffic in vehicular networksoverhead can be unmanageable when thousands of vehicles must update their Locations within short time. It is proposed to develop efficient Tracking Service in Vehicular networks based on Cloud Computing Platform as a Service for support every LBAs in vehicular networks and handle impact increase number of LBAs on a single vehicle.It should has the following features:

- Less Overhead Bandwidth
- Less number of location update massages
- Efficient Support LBAs

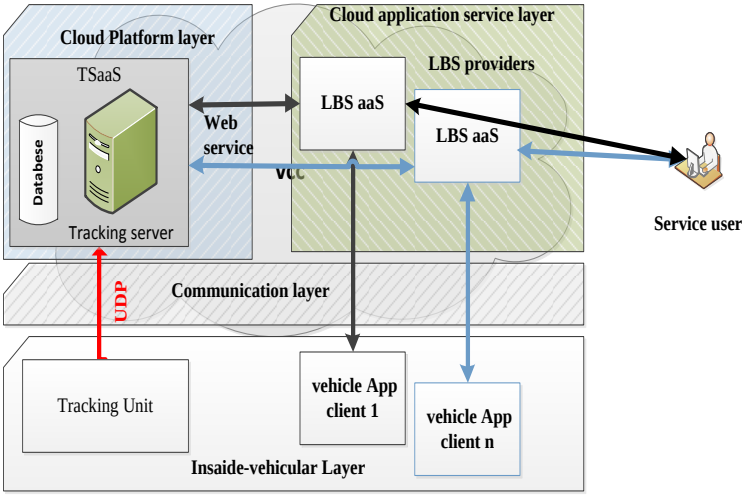


Fig. 2.TSVCC System Architecture

3. Proposed Tracking Service in Vehicular Cloud Computing (Platform as a Service)

In this section the proposed system is presented in four subsections. The first subsection presentssystem architecture of the proposed system. The second subsection describes real-time vehicle location process. The third subsection describes details of the location update .The last subsection describes Location server sharing (platform tracking system)

3.1. Proposed System Architecture

Proposed system architecture, as illustrated in Fig. 2, consists of six of components, namely, the Tracking unit, the tracking servers, the database server, the Vehicle app client, the LBSaaS server and the service user.

Tracking Unit: it is the software components that run on the OBU of the vehicle. It computes the current location of the vehicle (GPS Coordination) [8]. Then, it periodically sends the location to the tracking Server and another information such as vehicle status, speed, and temperature

Tracking Server: It is component that do two main functions. First, it receives the locationupdated messages fromvehicle's tracking unit and then it save this information in the data base server. Second, it provides a set of web services to be used and consumed by LBSaaS providers.

Database Server: This server is used to store the vehicle location and information in order to be easy to store large information and retrieve it easily.

Vehicle app client: It the application installed in vehicle. For each LBSaaS application server has application client .this client runs aninstructions that is sent by server application such as images capture, video capture, or any other request then it send response to application server. The single vehicle can have many app clients. In the previous system, the application client also periodically send the vehicle location to its app server location. In Proposed system we remove this function from app-client's responsibilities and assign it to tracking unit. Then, we proposed one location server for all applications server. This share the location server with all application servers that works on the same platform using service oriented architecture.

LBSaaS: This is an application server. It is a service that are produced to the user base on real-time vehicle location in VCC environment such pic on wheel, traffic analysis, tracking systems, and witness as a service.

Service user: It is the user. It is consumed utility of LBS- aaS provider services on demand model, this user may be inside vehicle or anywhere connected to internet.

3.2 Real-time Vehicle Location.

The location calculation in the modern vehicle is one of the basic function due to the availably of many location support techniques such as GPS. In Proposed research we used GPS because the availability of a satellite based global position system (GPS) enhances the accuracy of position of vehicle inhigh dynamic environment (PMB-648 GPS receiver accuracy is 2 meters) [8].The vehicle identification number (VIN) is used as a unique Id for the vehicle. The VIN is used for

distinguished each a vehicle from other in the database system. The tracking unit should be installed in the Vehicle. It can be implemented as a part of the vehicle operating system. It has the IP address and port number of the location server.The tracking unit computes the current location using GPS. Then, it add the VIN to location and it send the message to location server using UDP protocol. The UDP protocol is used because it is connectionless and need less time to send message to server.

The tracking unit can support more than one location server through multicast address

Algorithm Real-time Vehicle Location

Begin

While(tracking unit switch On)

initialization// check software update , get parametrats

Var VIN=Get Vehicle serial number

While(true)

Var POS=Get Current location // by GPS

Send current location to servers(VIN,POS) // by udp protocol

End while

End while

End

3.3 Location update

Location server receives location updatemessage from tracking unit of vehicle. The message contains VIN and location. Every vehicle should send location update message periodically to location server.The location server should check the message validity using DTD.In figure **3** we present an example of a DTD for the location update message validation[9], it add this message elements to the database server. The follow algorithm describe the location update steps.

Algorithm location update

Begin

While (true)

VarInComeL =Receive Location update message

IF InComeL is valide=true then

Add new record in database server

End if

End while

End

```
<!—Location Update Message  DTD file-- >
<!ELEMENTLocation_Message (timestamp,
Vehicle_Info+,Location_Info+, Velocity)>
<!ELEMENTUTC_DateTime (#PCDATA)>
<!ELEMENTVehicle_Info (VIN)>
<!ELEMENTVIN (#PCDATA)>
<!ELEMENTLocation_Info (Lati,Long)>
<!ELEMENTLatitude  (#PCDATA)>
<!ATTLISTLatitude  Indicator (S | N ) >
<!ELEMENTLongitude (#PCDATA)>
<!ATTLISTLongitude Indicator (E | W ) >
<!ELEMENTVelocity (#PCDATA)>
```

Fig. 3 DTD schema to validate Location Update

3.4 Location server sharing (platform tracking system)

The system use web server to handle web services requests from LBSaaSproviders.There are many web services can be provided to be used by application developers to develop applications that depend on real time vehicle location.Another part of the system is an application works as UDP server handles message update from tracking unit of vehicle.The web services provided by tracking server convert between location information and VIN and vice versa. The following algorithm show web service request handling.

Algorithm Tracking web service

Begin

Var request =Receive lbs provider request

Convert lbs provider request to suitable SQL query

Run SQL query on database

Convert result to lbs provider response

Send Response to lbs provider

End

4. Results and Evaluation

In this section, the proposed system is evaluated through studying the relationship between the number of app client and bandwidth or in terms of communication overhead (bandwidth consume) [10] and number of location update message should be send to the location server of Proposed optimized VCC which contain TSVCC platform as a service. The followed figers show results of quantativly evaluation to impact increase number LBAs vs single vehicle bandwidth consume and Location update message

The Size_LUMis minimum necessary size of location update message for carry Vehicle's information such aslocation(GPS data), VIN and speed.one UDP packet can carry one location update message.[11]

The NLBA is number of LBA's installed on single vehicle

Then, one position update operation for one vehicle costsUDP Packet's =NLBA* Size_LUM.So, one position update operation for one vehicle costs bandwidth =NLBA * Size_LUM /available bandwidth of vehicle.

Figure 4 show impact increase number LBAs in single vehicle on bandwidth consume ,note that in previous system when number LBA increase ,the bandwidth consume increase but in Proposed system stop impact increase number of LBA on a single vehicle.

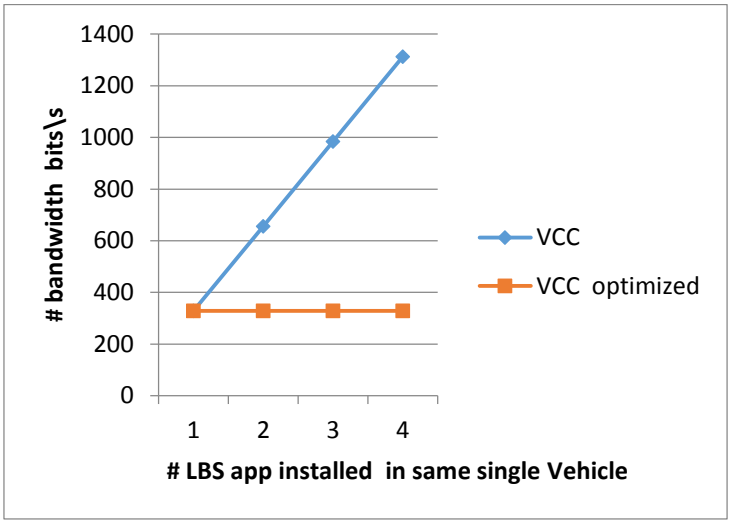


Fig. 4 app client vs bandwidth consume

Figure 5 show the impact increase number LBAs in single vehicle on number of necessary location update massage ,note that in previous system when number LBA increase ,the of necessary location update massage increase but in Proposed system stop impact increase number LBA on single vehicle.

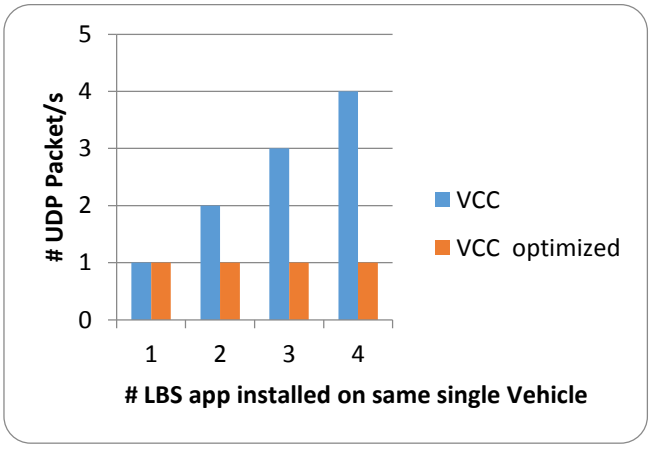


Fig.5 app client vs location update message

5. Conclusions and Future Work

In this paper, a novel enhancement to VCC architecture is proposed. The proposed scheme supports development real-time vehicle location applications in the intelligent transport system. The contributions of this paper are that it to solve the problem ofhuge bandwidth consume required by previous model. It addefficiently SOA architecture to previous VCC at platform layer .It open future work in many issues such as add Identification service to platform ,Optimize security and privacy Open area to researchers to add new function for VCC platform and can create new applications in future

7. References

[1] Rong Yu, Yan Zhang, Stein Gjessing, Wenlong Xia, Kun Yang, Toward Cloud-based Vehicular Networks with Efficient Resource Management, *IEEE Network*, (Volume:27 ,Issue: 5), Oct 2013, pp. 48 - 55.

[2] Fekri M.Abduljalil, "Video Capture Service in the Intelligent Transportation System based on Cloud Computing", International *journal of Computer Applications(0975-8887),*Volume 97-No.5,July 2014.

[3] Md Whaiduzzaman, Mehdi Sookhak, Abdullah Gani, Rajkumar Buyya, "A survey on vehicular cloud computing",*Journal of Network and Computer Applications*(2013), http://dx.doi.org/10.1016/j.jnca2013.08.004

[4] Mario Gerla," Vehicular Cloud Computing", *Vehicular Communications and Applications Workshop* ,2012 IEEE

[5] P Pesti, L Liu, B Bamba, A Iyengar, M Weber "ROADTRACK: Scaling Location Updates for Mobile Clients on Road Networks with Query Awareness*", Proceeding of the VLDB Endowment*, Volume 3, Issue 1-2,September 2010.

[6] Rasheed Hussain, Fizza Abbas, Junggab Son, Donghyun Kim, Sangjin Kim, Heekuck Oh, "Vehicle Witnesses as a Service: Leveraging Vehicles as Witnesses on the Road in VANET Clouds", in *Cloud Computing Technology and Science (CloudCom), 2013 5th International Conference on*, vol. 1, pp. 439-444. IEEE, 2013.

[7] Won Min Kang , Jae Dong Lee , Young-SikJeong and Jong Hyuk Park,"VCC-SSF: Service-Oriented Security Framework for Vehicular Cloud Computing", *Sustainability* 2015, *7*, 2028-2044; doi:10.3390/su7022028

[8] Nagarjuna Karyemsetty, Badugu Samatha, K Himantha Rao, " Design and Deployment of Vehicle Tracking System in VANETs using Xbee Pro: Prototype Model", *International Conference on Communication Networks (ICCN)*,2015

[9] J. G. Silveira, R. Balbinot, J. A. Oliveira Neto, F. C. Castello, A. R. Vieira, A. Quadra," Telemetry Extensions for the Mobile Location Protocol",*IEEE*,2003

[10] Sulata Mitra, "Dynamic Resource Reservation for Authentic Vehicles in VANET", *International Conference on Innovations in Engineering and Technology (ICIET'2013)* Dec. 25-26, 2013 Bangkok (Thailand)

[11] Mario Gerla, Jui-Ting Weng, Giovanni Pau,"Pics-On-Wheels:Photo Surveillance in the Vehicular Cloud" ,*IEEEInternational Conference on Computing*, 2013,pp.1123-1127.

[12] Victor W. Chu∗, Raymond K. Wong∗, Wei Liu†, Fang Chen†, Chang-ShingPerng," Traffic Analysis as a Service via a Unified Model", *2014 IEEE International Conference on Services Computing.*

[13] Abdessamed Derder, Samira Moussaoui, "Target Tracking in VANETs Using V2I and V2V Communication",*International Conference on Advanced Networking Distributed Systems and Applications*,2014

[14] S. Boopathi , K .Govindaraju ,M . Sangeetha ,M. Jagadeeshraja ,M. Dhanasu, "REAL TIME BASED SMART VEHICLE MONITORING AND ALERT USING GSM", *International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 11, November 2014*

[15] Jetendra Joshi, Kritika Jain and YashAgarwal," CVMS: Cloud Based Vehicle Monitoring System in VANETs ", *International Conference on Connected Vehicles and Expo (ICCVE),*2015

Using Remote sensing and GIS Techniques in mapping of groundwater potential zones in the Amran basin.

Mohammed M. GahGah
Geography department, Education Faculty,
Hajjah University, Hajjah, Yemen.
E-mail: dr.gahgah@yahoo.com
Phone: 00967- 773457172

Abstract

The objective of this study is to explore the groundwater availability in the Amran basin. Remote sensing data and geographic information system were used to locate potential zones for groundwater in the Amran basin. Various maps (i.e., base, geological, structural, drainage, slope, land use/land cover hydrological, and groundwater level) were prepared using the remote sensing data along with the existing maps. The groundwater availability of the basin is qualitatively classified into different classes (i.e., good, good to moderate, moderate, moderate to poor and poor), based on its contribution of the groundwater conditions, on 1: 100,000 scale map. The alluvial plain in main valleys, highly fractural areas and sandy formation areas like Tawilah formation, Nayfa and Amran Group were successfully delineated and shown as the prospective zones of groundwater.

Keywords: *GIS; remote sensing; groundwater, Amran, Yemen*.

1. Introduction

Groundwater resources are vital for Yemen's agriculture. For their recharge they depend mainly on spate running water and rainfall. Runoffs and springs in catchment's areas are the main sources of groundwater recharges. In Yemen, the estimated groundwater is around 1000MCM, which makes the total renewable water resource sum 2.5 MCM, while the total demand is estimated to be 3,400MCM with 900MCM deficit, which is covered from deep aquifers [1].

Many hydrogeological applications of remote sensing rely on the user to link image or data interpretation to groundwater processes [2]. Remote-sensing techniques using Landsat (TM and ETM+) have

also been used to identify groundwater discharge areas by identifying changes in lakes temperature [3] and [4], and by mapping lineaments associated with groundwater springs [5].

A Geographical Information System (GIS) allows one to manage, analyze, and display geographic information which is represented as layers of information (ESRI-1, 2004). This information set includes – maps, globes, geographic data sets, processing and work flow models, data models, and metadata. The broad area of GIS can be viewed in different ways. GIS is one of the most important tools for integrating and analyzing spatial information from different sources or disciplines. It helps to integrate, analyze and represent spatial information and database of any resource, which could be easily used for planning of resource development, environmental protection and scientific researches and investigations. A GIS cross-overlay procedure can be used to merge different component maps and weight information relating to controls on recharge processes [6]. This approach requires integration of the satellite data with geomorphological data within a GIS framework [2].

In this study, a Geographical Information System (GIS) is included as an interactive tool to estimate and interpret groundwater related factors. The present research shows a combination of the Remote Sensing technique with a dedicated GIS toolbox to evaluate groundwater in Amran basin.

2. Data and methodology

2.1. Description of study area

The Amran basin is one the most important agricultural production center in Yemen with fast growing population and increasing urbanization. In contrast to land resources, there is a high interdependency among water users due to the movement of water in the hydrologic cycle. As a result, there is a growing awareness of the need for adequate water resources to maintain environmental requirements.

The study area is located at Amran Valley, about 50 km north west of the capital city of Yemen (Sana'a), The total geographical area covers 1580 km^2 within the latitudes 15º 30` and 15 º 55` North and

longitude 43º 45` and 44º 15` East (Fig.1). From a physiographical view, the area is located within the Yemen Plateau (the western Yemen's mountains composed of volcanic trap basement, with the average altitude of 2500 m above the sea level. Aman plain is surrounded by mountains from all sides. Amran basin is located within Central Yemeni Highland which called the western mountainous range. This range is built-up mainly by volcanic rocks [7], whereas the lowest part located in the northeast with elevation of 1950 m above the sea level and the highest part in the southwest with elevation of 3200 m above the sea level. Amran valley can by divided to three main parts: Wadi Akhar located in west of Amran city and extend to the northwest. Wadi Alboun express the main part of Amran basin and extending from southwest to northeast. Qa Hamidah located in the northwest of the study area getting the direction to the northeast.

The district has a semi-arid climate, characterized by cold winters during November till February and pleasant summers from April and October with temperatures varying 0º C in the winters from December till 32º C on summer day. The study area receives most rainfall from the northwest monsoon (90%). The maximum precipitation about 402 mm/year. The main direction flow in the basin is from south-west to north-east and has drainage density of 0.4–2.75 km/km^2. It is one of the major agricultural plain in western intermountain plains. In the Amran basin, irrigation is mainly supplied from groundwater causing the depletion of water in the area. The main crop in the basin is potato, vegetables, Qat, maize, wheat.

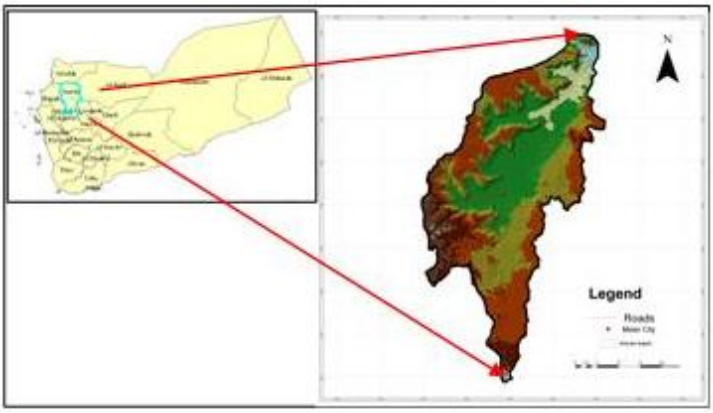


Fig. 1. Location of the study area.

Geological, the area consists of two main types of rocks, which are Cenozoic volcanic and Mesozoic sedimentary rocks (Fig. 2), as it will be described in the following (from the Youngest to the oldest):

1- *Quaternary Volcanic*

these type of volcanic activity appear in the south part of the study area consist a basaltic lava flows (alkaline basalt), which occurred between 100000 and 1500 years ago. The thickness of volcanic is more than 600 m. This volcanic contain a volcanic cones and thick volcanic ash. The volcanic cones could be seen in the satellite images in the southeastern part of the area. Meanwhile, the lava flows is shown in the north part. The Quaternary deposit is located in the Amran valley between Amran limestone and the Quaternary volcanic with varies thickness of few meters and 1500 meters [8].

2- Yemen Group (Tertiary) Volcanic

These group is presented in the area as dikes mostly in the north and west parts of the Amran basin.

3- Tawilah Group (Tertiary – Cretaceous)

Tawilah group is ranges in age from late Cretaceous to Eocene [9]. The rocks of the Tawilah group have a wide geographical distribution of the almost outcrop basement in the east part of the country and limited presence in the west part. The mainly composed of this unit is non-fossiliferous porous and fissured sandstones, gravel and conglomerate. The thickness of the Tawilah formation in the study area is more than 300m, and well exposed in the Thula area. The extrusion of lavas protected the Cretaceous sandstone from further erosion.

4- Amran Group (Jurassic)

The Amran Group is thick series of dominantly calcareous sediments, but locally including significant sequences of shales and evaporates. The carbonates are commonly massive, well cemented limestone and dolomites, fissured rocks.

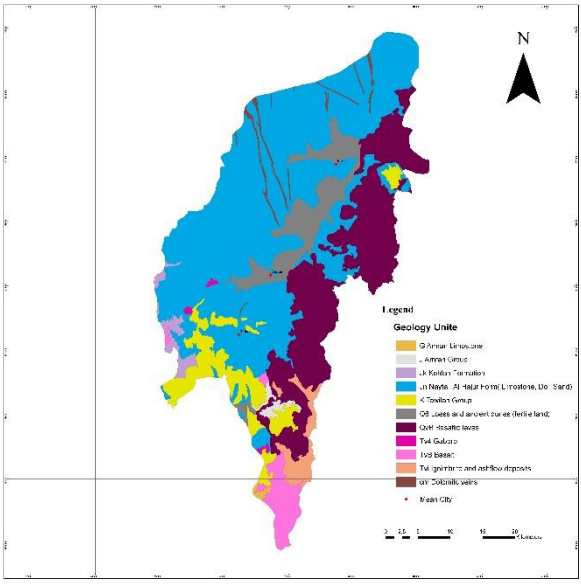


Fig. 2. Geological map of the study area.

Maximum thickness of the Amran group exceed 800m [10] and [11]. it is a complex group, with two main depositional environments: a neritic environment which resulted in a limestone and marly facies, and a shallow-water environment resulting in an evaporitic sequence. The limestone and dolomites contain more than 50% carbonates and the remaining is clay, silt and iron [12].

5- Kohlan Group (Triasic)

The Lower-Jurassic Kohlan Group is a minor unit, consisting mainly of sandstones with considerable vertical and lateral variation. Its base is always an unconformity with the underlying older rocks. [13] described Kohlan Group as a continental sediments of fluvio- lacustrine origin. The thickness of the unit is variable, with an average of some 60 meters.

Tectonic activity was accompanied by intensive volcanicity, including the intrusion of dikes, which started locally at the end of the Cretaceous and became widespread during the Tertiary period [13]. The major structure in Yemen is N40E extension [14] and [15]. The major faults in Yemen are associated with the opening of the red sea. The structure in the area was affected when the period of deformations occurred during Proterozoic and reactivated in straight phase.

2.2. Methodology

The satellite sub-images (ETM) were extracted and subjected to radiometric and geometric corrections prior to unsupervised classification. Unsupervised classification was used to get the classification based on natural clustering. The Iterative Self-Organizing Data Analysis (ISODATA) clustering algorithm available in ERDAS/IMAGINE image processing software was used for the purpose. The image has also been used to study the lineaments in relation to the hydrogeology of the basin, as it is play main role in the recharge of the basin.

The spatial probability can be obtained either through dynamic modeling or through analyzing the relation between the locations of a set of environmental factors, in order to produce areas of groundwater potential that have similar combinations of factors, using heuristic or statistical methods. The spatial information on groundwater potential locations is used in combination with the environmental factors such as lithology, slope and drainage.

With the fast development in geo-information science and earth observation, there are more and more tools available for carrying out a more dependable groundwater evaluation. As topography is one of the major factors in ground water analysis (as it is also for other types such as flooding, forest fires, volcanic eruptions [16] the generation of a digital representation of the surface elevation, called the Digital Elevation Model (DEM), plays a major role. Photogrammetrical methods using aerial photos, the use of (high-precision) GPS and the digitizing and interpolation of contour maps, X-band DEM data of a horizontal resolution of 30 m and relative and absolute height accuracies of 6 m and 16 m, respectively. These data are unclassified and available for public use from the German Aerospace Centre. Both C and X-band datasets are geometrically corrected and projected to the WGS72 datum. The rising availability of DEM data from radar interferometer (i.e., SRTM DEM data) give new possibilities to earth scientists and was successfully compared with other DEM extracted from aerial photography, ASTER, RADAR-SAT and SPOT-PAN [17] and [18]. With the initiation of geographical information systems (GIS) DEM can be used together with other spatial datasets such as geological, airborne magnetic, and hyper spectral datasets. For more advanced spatial analyses the DEM provides a basic spatial reference system and images, where vector data can automatically be draped over

the DEM. The ability of using spatial tools for generating geomorphometric models such as elevation, slope and aspect gives wide range of new analysis methods.

3. Analysis approaches and results

In order to demarcate the groundwater potential zones (availability) in the study area different thematic maps have to be prepared from remote sensing data, topographic maps and geological maps in conjunction with the existing maps. In this study, the data required for groundwater prospect have been determined as groundwater related factors in the area which are lithology, slope, digital elevation model (DEM), structures density, drainage density, rainfall and landuse and landcovers as follows..

3.1. Base map

The base map has to be prepared in ArcMap application of the ArcGIS software using geocoded toposheet. The base map contains the following details with the appropriate symbols: (i) cities/villages, (ii) main valleys, (iii) roads, (iv) district boundary (Fig.3).

3.2. Topography

Topographic information has to be collected from SRTM DEM data (Fig. 4). Drainage map is extracted from the SRTM DEM data by using the flow direction grid map. Digital Elevation Modeling techniques are adopted to derive parameters on geomorphology-based ground water influence factors like slope gradient and elevation. Slope Map (Fig.5) in degree has been prepared from SRTM DEM range from 0_ to 58º.

3.3. Lithology map

Lithology map was prepared by digitizing each lithologic unit/rock type in ArcGIS software package from geology maps obtained from geological survey of Yemen, Sana'a sheet (1990). Furthermore each lithologic unit or rock type is classified based on the legend available on the geology map (Fig. 2) and the rock unites in the area has been described in section 2.1.

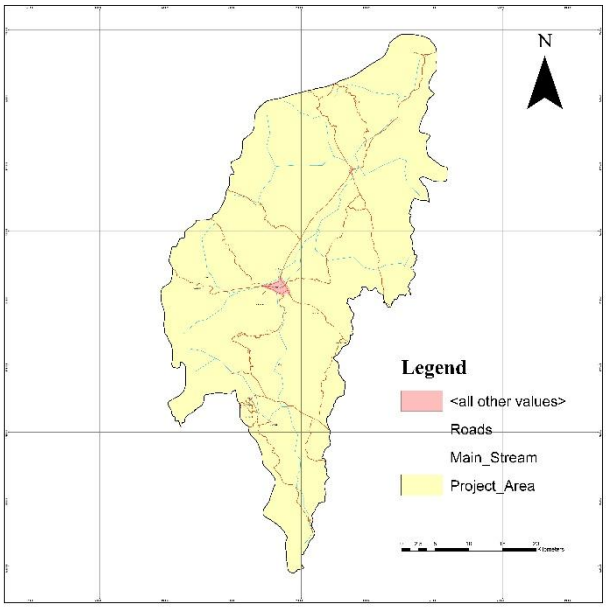


Fig.3. base map of the study area.

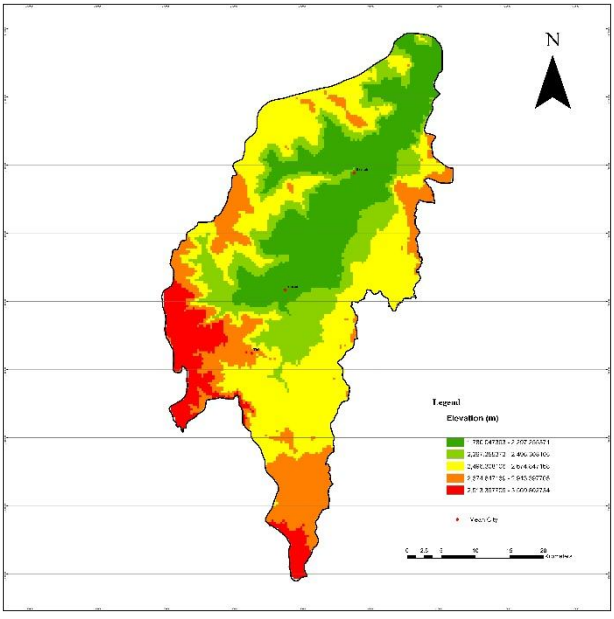


Fig. 4. DEM map of Amran basin.

3.4. Geological structural map

Geological structural map was prepared by digitizing lineaments, faults in ERDAS 9.1 software from satellite Landsat Image (ETM). Different structures that should be mapped have been represented on the map with appropriate line symbols. Lineament density which reflects the geological structure in the area and its orientation has been classified in zones to show its relation with the groundwater.

Actually, Amran basin is graben basin [19], as a result of the tectonical activities that the area faced. The majer trend of the lineaments in the area is NW-SE and NE-SW , these lineaments are contained the faults and dikes. The north and northwest parts of the basin showed high density (>1.3 km/km^2) (Fig. 6). These types of geological structures are acted as barriers and conduits for groundwater movement.

3.5. Hydrological map

A raster of flow direction shown in (Fig. 7) gives us the flow across a surface which will always be in the steepest down-slope direction and is used to determine the stream network. Flow direction map created from a raster shows the direction of flow of the stream is from south-southwest to north-northeast. A raster of flow accumulation is created. The drainage pattern observed in the Amran basin is dendritic in nature and the basin has a drainage density of 0.4–2.75 km/km^2. Stream ordering is given by using DEM and Drainage network in Arc GIS by using Hydrology module. Drainage network helps in the delineation of watersheds and for suggesting various water harvesting structures(Fig. 7, 8).

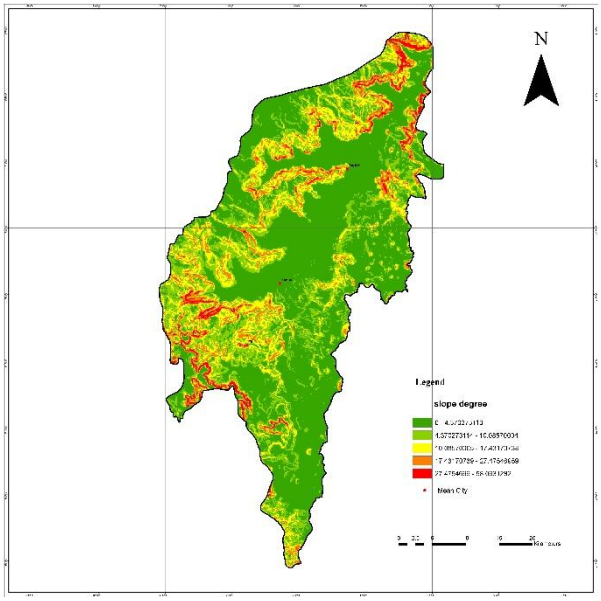


Fig.5. Slope map of Amran basin.

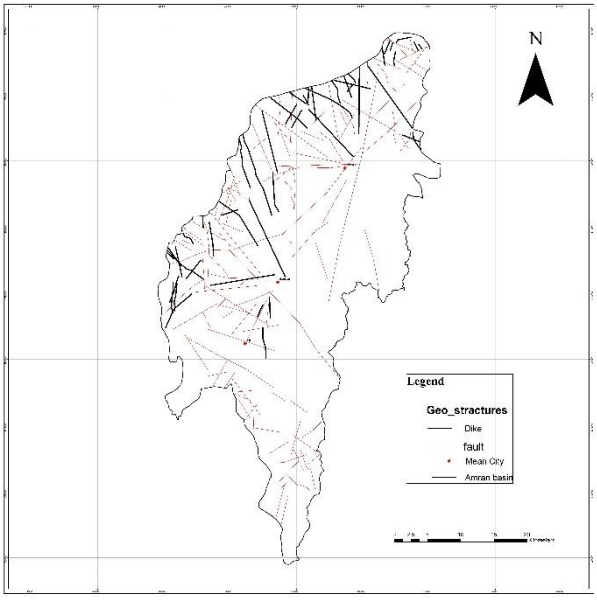


Fig. 6. Geological structure map of Amran basin.

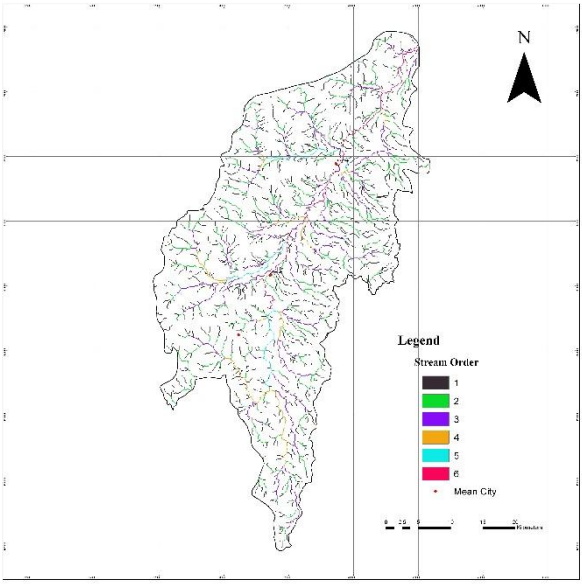


Fig. 7. Drainage map of Amran basin.

DEM provides excellent information on hydrologic aspects like steam courses, major reservoirs, canal commands, etc. with help of landuse landcover map produce from the satellite image all the information can be derived and integrated to the hydrological map.

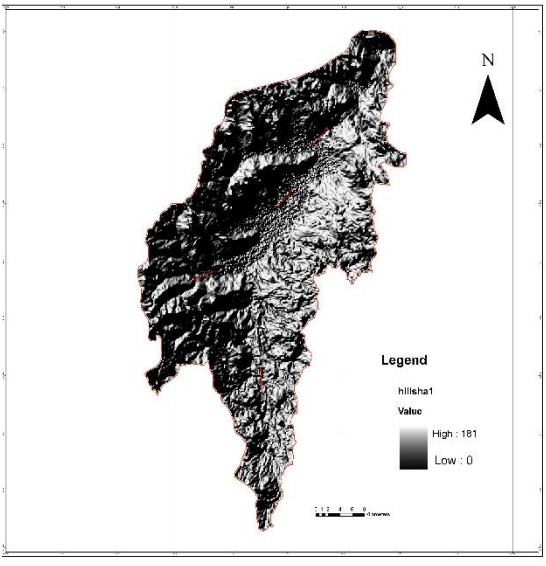


Fig.8. Earth relief and the main direction of flow, Amran basin.

The hydrological information derived from DEM in combination with collateral data has shown on a separate map overlay in a classified with appropriate symbols. For the preparation of hydrological map overlay, the following sources of information are required: (i) ground water levels, (ii) wells distributions and density, and (iii) meteorological data. Moreover the following details have to be shown in hydrological map overlay: (i) streams (ii) depth to Groundwater Level, (iii) existing wells distributions and (iv) rainfall data, which indicating average annual rainfall (average annual rainfall in the unit of mm). This source of rainfall data was obtained from Yemen Meteorological Department. All these data have been collected and incorporated for preparing the hydrological map overlay. Create a new layer in ArcGIS software and transfer by superimposing the integrated streams net, elevation, ground water levels and wells density map overlays. These integrated maps result in the hydrologic map (Fig.9).

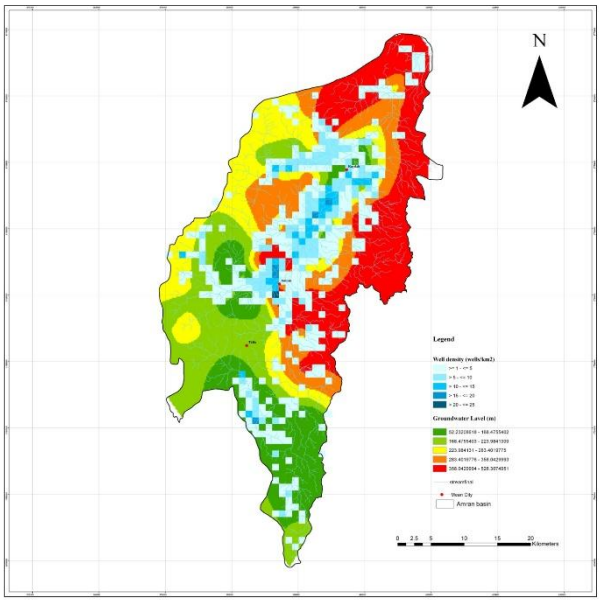


Fig.9. hydrological map of Amran basin.

3.6. Land use/land cover classification

Land use/land cover map was prepared for Amran basin using Landsat Thematic Mapper Remote Sensing Data with special emphasis on irrigated areas in the Amran basin. Geocoded Landsat TM data was digitally classified in ERDAS IMAGINE 9.1 Software package using supervised classification technique along with limited ground truth for mapping the broad categories. The parallelepiped supervised classification technique was applied as it has been widely used decision rule which is based on simple boolean. Training data in spectral bands are used in performing the classification. Ground truth was carry out using the previous knowledge and previous studies. Classes have the flexibility to merge to a higher class or break into a distinct class based on the land cover percentages. A total of thirteen landuse and landcover (LULC) classes were mapped for the study are (Fig. 10).

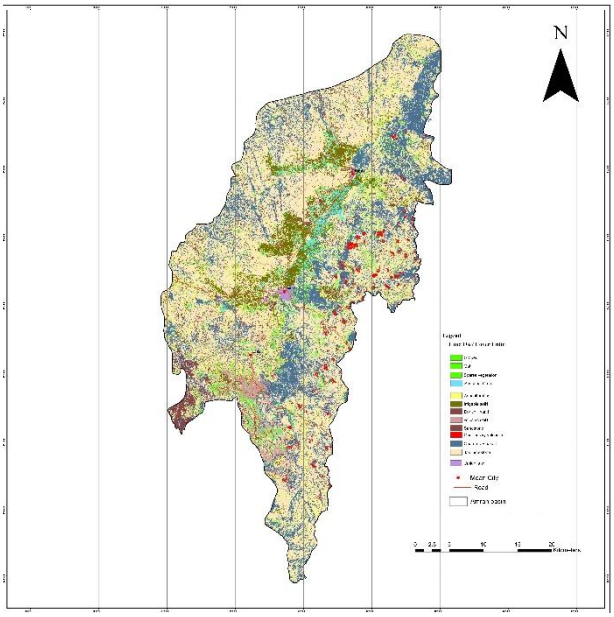


Fig.10. Landuse and landcover map of Amran basin.

3.7. Groundwater prospects map

For preparing the groundwater prospects map the following procedure has to be followed. First a groundwater prospects map has to be prepared by integrating the information from lithological, structural, geomorphological and hydrological maps in ArcGIS 9.1 as indicated in the following steps:

Step 1: integrate the lithology, slope, elevation, structures density, rainfall and drainage density to thematic maps. Based on the contribution of those factors to ground water control, the weightage values were given.

Step 2: The groundwater prospects of each unit in the thematic maps have to be evaluated by applying the overlay method considering all ground water related factors unites information. overlaying factor maps which is the usual practice in GIS analysis of groundwater zonation, in the present case effective factor ratings for each theme are calculated on each facet.

The groundwater potential map was produced on 1: 100,000 scale of the entire basin. Various hydrogeomorphic units are grouped into potential zones whereas, the basin has been classified on to five zones as following: Good potential zone, moderate to good potential zones, moderate to poor potential zone and poor zone (Fig.11). In general, alluvial in the main valleys, highly fractural areas and sandy formation areas like Tawilah formation, Nayfa and Amran Group are characterized by more porosity and high fractures. therefore, these areas are classified as good to moderate prospect zones for groundwater exploration. The basalt areas with not lineaments are grouped as poor zones. High density of the groundwater discharge (pumping) is concentrated in those areas, as consequent, the groundwater laver is low even the recharge is high.

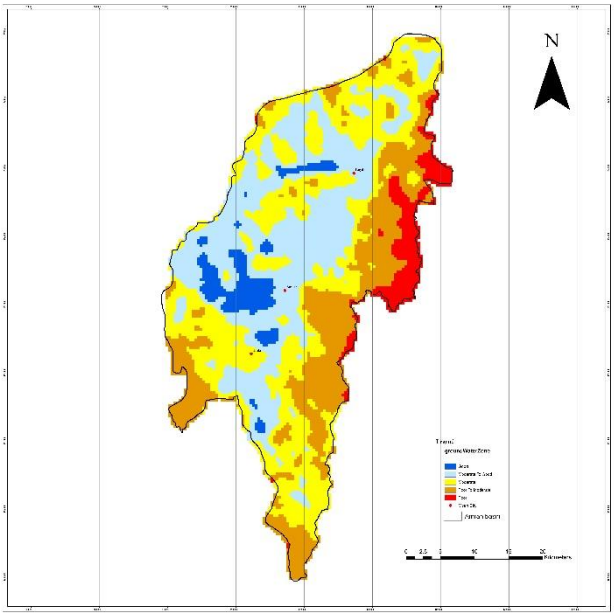


Fig. 11. Groundwater potential zones map of Amran basin.

4. Conclusions

The produced groundwater related database could help as information source to Institutions, researchers, groundwater practitioners, drilling companies and decision makers etc. Mapping of groundwater resources have been increasingly implemented in recent years because of increased demand for water. The data most commonly available for groundwater study are geological, geomorphological and hydrological information. In this study we attempted to identify groundwater potential zones using remote sensing and geographic information system techniques in the Amran basin. To determine the groundwater availability of the Amran basin, various thematic maps such as, base map, lithological map, structural map, elevation map and hydrological map were prepared from remote sensing data, topographic maps, geology maps and rainfall maps using ERDAS 9.1 software and Arc GIS and integrated these maps to produce groundwater potential zonation map. It is advisable to use new area in the plateau and south parts of the basin for future drilling.

5. References

[1] Q.Y. Al-Asbahi, Water Resources Information in Yemen. IWG-Env, International Work Session on Water Statistics, Vienna, June 2005, pp. 20-22.

[2] A.M. Meijerink,. Remote sensing applications to hydrology: groundwater. *Hydrol Sci* 41(4), 1996. Pp.549–5.

[3] A. K. Singh and S. R. Prakash, An integrated approach of remote sensing, geophysics and GIS to evaluation of groundwater potentiality of Ojhala subwatershed, Mirzapur district, UP, India. Map India conference, India, 2003.

[4] E. N. Tcherepanov, V. A. Zlotnik, and G. M.. Henebry, (). Using Landsat thermal imagery and GIS for identification of groundwater discharge into shallow groundwater-dominated lakes. *Int J Remote Sens* 26(17), 2005. Pp.3649–3661.

[5] E. Sener. A. Davraz and M. Ozcelik, (2005). An integration of GIS and remote sensing in groundwater investigations: a case study in Burdur. Turkey. *Hydrogeol J* : 13. 2005. Pp.826–834.

[6] M. W. Lubczynski and J. Gurwin. 2005. Integration of various data sources for transient groundwater modelling with spatio-temporally variable fluxes: Sardon study case, Spain. *J Hydrol*, 306. 2005. Pp.71–96,61.

[7] Al-Thour. K. A. Stratigraphy sedimentology and diagnosis of the Amran group (Jurassic) to the west and north of Sana'a Yemen. Ph.D thesis, birmengham Uni. UK. 1997. (unpublished).

[8] A. Abou Khadrah.. A review of the sedimentological evolution of Yemen Arab Republic. Bulletin Faculty Science, San'a University, **2**. 1982. Pp. 39-55.

[9] K. Al-Subbary, G. M. Nicols, and D. Basence. Cretaceous to Tertiary pre-Rift fluvial to marina sedimentation in Yemen. *Geological Society of Egypt*, Cairo.No.1. 1993. pp. 383-407.

[10] K. A. Al-Thour. Facies sequences of the middle-upper jurassic carbonate platform (Amran Group) in the Sana'a region, Republic of Yemen. *Marine and Petroleum Geology*, **14.6. 1997**. pp. 643-660.

[11] K. Weiss, (2008). Basin Platform Transitions in Upper Jurassic Carbonates of the Amran Group, Yemen. Middle East Conference and Exhibition, 2-5 March, Manama, Bahrain. 2008.

[12] Buol, S. W., Hole, F. D., and McCracken, R. J. (). Soil genesis and classification. 2nd et. Iowa State Univ. Press, Ames. 1980.

[13] A, N. Charalambous, Problems of groundwater development in the Sana'a basin, Yemen Arab Republic. Improvements of Methods of

long Term Prediction of Variations in Groundwater Resources and Regimes Due to Human Activity (Proceedings of the Exeter Symposium, July 1982). IAHS Publ. no. 136. 1982.

[14] Huchon, P., Jestin, F., Cantagrel, J. M., Gaulier, J. M., Al Khirbash, S., & Gafaneh, A. (1991). Extensional deformations in Yemen since Oligocene and the Africa-Arabia- Somalia triple junction: Annales Tectonicae, v. 5, p. 141-163.

[15] Davison, I., Al-Kadasi, M., Al-Khirbash, S., Al-Subbary, A. K., Baker, J., Blakey, S., Bosence, D., Dart, C., Heaton, R., McClay, K., Menzies, M., Nichols, G., Owen, G., & Yelland, A. (1994). Geological evolution of the southeastern Red Sea rift margin, Republic of Yemen: *Geological Society of America Bulletin*, **106**, p. 1474-1493.

[16] van Westen, C., van Asch, T. W. J., & Soeters, R. (2006). Landslide hazard and risk zonation-why is it so difficult? *Bulletin of engineering geology and the environment*. **65**, p. 167-184.

[17] Siva Subramanian, K. S., Singh, A., & Sudhakar, M. (2003). Evaluation of Digital Elevation Models Created from Different SatelliteImages. www.gisdevelopment.net/technology/rs/mi03079.htm.

[18] Reinartz, P., Lehner, M., Müller, R., Rentsch, M. & Schroeder, M. (2003). First Results on Accuracy Analysis for DEM and Orthoimages Derived from SPOT HRS Stereo Data over Bavaria. www.ipi.uni-hannover.de/html/publikationen/2003/workshop/baudoin.pdf

[19] Beydoun, Z. R., As-Saruri, M., & Baraba, R. S. (1996). Sedimentary basins of the Republic of Yemen: their structural evolution and geological characteristics: Revue Institute Francais Petrole, **51**, p. 763-775.

Estmating the Teaching Contact Hours based on Fuzzy Maps: New Approach

Mohammed Al-Sarem
Department of Information Science
Taibah University
Al Madinah Al Monawarah, KSA

Hasan Al-Shaikh
Department of Computer Science
AL Andalus University
Sana'a, Yemen

***Abstract*— Current paper presents an innovative approach for developing a course specification taking in consideration the impact of the covered topics in a course on the whole program learning outcomes. It presents a practical methodology allowing instructors to design an efficient course specification based on the national academic accreditation and assessment standards.**

Keywords—Course Specification; NCAAA; Fuzzy Set; Estimation.

I. INTRODUCTION

On the core of the National Academic Accreditation and Assessment Standard stated by NCAAA lies process of matching the learning Outcomes (LOs) of a course with the whole program learning outcomes (PLOs) [1]. Since the standard does not show how to cover and match the topics that will be taught with the program learning Outcomes, the purpose of this work is to present an approach that aims to facilitate, on one hand, the matching process of LOs to PLOs. On the other hand, to show how should instructor to distribute the course topics over the semester. The proposed approach suggests to use the fuzzy set theory [2] as a way to describe imprecision that is characteristic of much of human reasoning. It suggests also to use Two-Phases process in which experts' opinions are elicited and matched with PLOs and LOs, then transform them into so-called topic-program learning outcomes (T-PLOs) Table.

II. THE PROPOSED APPROACH

A. Phase I: Elicitation process

Step1: In order to get coherent matrix, a set of domain experts are invited to set the relationship, from her/his own point of view, among the LOs and PLOs. Since it is difficult to avoid imprecision in the matching process, the experts are invited to set their opinion using a list of linguistics variables (very strong, strong, week, very week).

Step 2: Aggregating all the (linguistic) weight of the LOs-PLOs relation using the well-known fuzzy logic method of SUM. The result of this step is an aggregated (linguistic) weight. In case of linguistic weight, the defuzzification method of center of gravity (COG) is applied and a numerical weight for the relationship is calculated. The output of this phase is two matrices: One for presenting the LOs-PLOs relation and the other is for presenting the topics-LOs table.

The before mentioned methodology is applied using the following algorithm:

Step 1: For all the *M* experts, set credibility weight $b_k = 1$

Step 2: For all the ordered pair((LO_i and PLO_i) / (T_i and LO_i)) each k[th] of the M experts is asked to set the relation using the linguistic variables.

Step 3:

IF for one relation more than $2M/3$ different linguistic weights are suggested **THEN**

- ask experts to reassign weights for this particular relation and go to step 2

ELSE

IF the k[th] expert has proposed for a relation a linguistic weight that does not belong to the neighborhood of weights[1] **THEN**

- disregard this particular linguistic weight and penalize the expert who chose the "distant" weight and set him a new credibility weight $b_k = r.b_k$

Step 4:

- Aggregate all the linguistic weights proposed for each relation using the *SUM* method where the membership function μ suggested by k[th] expert is multiplied by the corresponding credibility weight b_k.

- Use the *COG* defuzzification method to calculate the numerical weight w_{ij} for each relation.

Step 5: IF there is an ordered (LOs-PLOs/ Topic-LOs) pair not examined go to step 2

[1] A linguistic weight does not belong to a neighborhood when it is not partially overlapping with at least another linguistic weight proposed by another expert.

ELSE

- Construct the LOs-PLOs/ Topic-LOs table whose are the defuzzified weights w_{ij}. w_{ij}

END.

Phase II: Transformation process

After generating the LOs-PLOs and Topics-LOs tables, a Max-Min composition method is used to transfer them into the T-PLOs [3]. To illustrate the Max-Min composition, let:

$$R1 = \{(x,y) \mid (x,y) \in X \times Y\}$$
$$R2 = \{(y,z) \mid (y,z) \in Y \times Z\} \quad (1)$$

and the Max-Min composition will be:

$$R1 \circ R2 = \{(x,z) \mid (x,z) = Max\{Min\{\mu R1(x,y), \mu R2(y,z)\}\}\}$$

for $x \in X, y \in Y$ and $z \in Z$ (2)

III. BENEFITS OF THE PROPOSED APPROACH

As we mentioned before, the main goal of the proposed approach is to illustrate how to match the topics that will taught in a particular course with the whole program leaning Outcomes. However, we can extend this approach to estimate the required contact hours and topics distribution per weeks.

Let the generated T-PLOs Table, from the previous step, is presented as follows:

$$W = \begin{array}{c} \\ t_1 \\ t_2 \\ \vdots \\ t_{\mathrm{m}} \end{array} \begin{array}{c} \begin{array}{cccc} \mathrm{PLO}_1 & \mathrm{PLO}_2 & \cdots & \mathrm{PLO}_{\mathrm{n}} \end{array} \\ \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{m1} & w_{m1} & \cdots & w_{mn} \end{bmatrix} \end{array}$$

where,

$w_{ij} \in [0,1]$, w_{ij} = 1 denotes the program learning outcome PLO_j is covered in the topic t_i, whilst w_{ij} = 0 is vice versa, $0 \le i \le m$ and $0 \le j \le n$, m is number of topics in a course and n is number of program learning outcomes.

Let also, R_{ij} denotes impact of a topic t_i on the program learning outcome PLO_j and is calculated as follows:

$$R_i = \frac{\sum_{j=1}^{n} w_{ij}}{Max(\sum_{j=1}^{n} w_{ij})} \quad (3)$$

Then, topics distribution per weeks is calculated as follows:

$$D(t_i) = \frac{h(C_i).R_i}{\sum_{i=1}^{n} R_i} \quad (4)$$

where,

$h(C_i)$ denotes the total credit hours for the course C_i.

IV. ILLUSTRATIVE EXAMPLE

At IS department, five domain experts are invite to fill the course specification of the " IS- 448 Semi-structured Data" course that is taught at the second semester of the fourth year of the curriculum plan [4,5]. According to the PLOs of the curriculum (Table 1), the experts should to match them with the course LOs (Table 2) which also should to be covered by the course topics (Table 3).

Table 1: Program Learning Outcomes

#PLOs	Description
PLO_1	Attain an ability to apply knowledge of computing and mathematics appropriate to the program's student outcomes and to the discipline. **[ABET-Criterion 3a]**
PLO_2	Attain an ability to analyze a problem, and identify and define the computing requirements appropriate to its solution. **[ABET-Criterion 3b]**
PLO_3	Attain an understanding of professional, ethical, legal, security and social issues and responsibilities. **[ABET-Criterion 3e]**
PLO_4	Attain an ability to communicate effectively with a range of audiences. **[ABET-Criterion 3f]**

PLO_5	Attain an ability to analyze the local and global impact of computing on individuals, organizations, and society.**[ABET--Criterion 3g]**
PLO_6	Attain an understanding of and an ability to support the use, delivery, and management of information systems within an Information Systems environment. **[ABET- IS Criterion 3j]**
PLO_7	Attain an ability to analyze the local and global impact of computing on individuals, organizations, and society. **[ABET-Criterion 3g]**
PLO_8	Attain recognition of the need for and an ability to engage in continuing professional development. **[ABET-Criterion 3h]**
PLO_9	Attain an ability to use current techniques, skills, and tools necessary for computing practice. **[ABET-Criterion 3i]**
PLO_{10}	Attain an understanding of and an ability to support the use, delivery, and management of information systems within an Information Systems environment. **[ABET-IS Criterion 3j]**
PLO_{11}	Attain and demonstrate programming skills in at least one high-level programming language.
PLO_{12}	Attain an ability to effectively utilize relational databases to store, retrieve, and manipulate data.
PLO_{13}	Attain an ability to use web languages and web services to create and interact with web pages.

Table 2: Course Learning Outcomes

LOs#	**Description**
LO_1	Explore semi-structured data settings.
LO_2	Design query mechanisms for semi-structured data
LO_3	Apply the skills of design and modeling data structure
LO_4	Examine the basic concepts of semi-structured data
LO_5	Design semi-structured data solutions

Table 3: Course Topics

Topic#	Description
t_1	Introduction: Semi-structured data, XML core concepts and namespace
t_2	DTDs, a simple schema language for XML documents
t_3	XPath, a navigation language for XML documents, XML Schema, a more expressive schema language for XML documents
t_4	XQuery, a query language for XML documents
t_5	Validation of HTML 5, RDF and Linked Data
t_6	CSS and the DOM
t_7	Web Data vs. Web Documents vs. Web Applications

At the end of discussion session, the experts fill the LOs-PLOs as presented in Table 4.

Table 4: Map course LOs with the program LOs

Course LOs #	Program Learning Outcomes			
	PLO_3	PLO_8	PLO_9	PLO_{12}
LO_1		0.7	0.3	
LO_2	0.2			0.8
LO_3			1	
LO_4		1		
LO_5	0.5		0.5	

Table 5: Map course topics with the program LOs

Course Topic#	Course Learning Outcomes

	LO_1	LO_2	LO_3	LO_4	LO_5
t_1	0.6			0.4	
t_2	0.2			0.5	0.3
t_3		0.3	0.3		0.4
t_4		0.5			0.5
t_5			0.5		0.5
t_6			0.5		0.5
t_7		0.2	0.3		0.5

Based on formula in Equation (2), the result of transferring the LOs-PLOs and Topics-LOs tables into the T-PLOs.

Table 6: Transorming Result

Course Topic#	Course Learning Outcomes				
	PLO_3	PLO_8	PLO_9	PLO_{12}	R_{ij}
t_1	0	0.6	0.3	0	0.6
t_2	0.3	0.5	0.3	0	0.73
t_3	0.4	0	0.4	0.3	0.73
t_4	0.5	0	0.5	0.5	1
t_5	0.5	0	0.5	0	0.67
t_6	0.5	0	0.5	0	0.67
t_7	0.4	0	0.5	0.2	0.73

Let we are planning to arrange 45 meetings per a semester with 3 hours per week, then to estimate distribution per weeks we use the formula in Equation (4).

Table 7: Distribution Topics per weeks

Course Topic#	**No. of Weeks**	**Contact hours**

t_1	≈2	6.652174
t_2	≈1.5	4.304348
t_3	≈1.5	5.478261
t_4	≈2	5.869565
t_5	≈3.5	10.56522
t_6	≈3.5	7.826087
t_7	≈1.5	4.304348
Total	≈15	45h

V. CONCLUSIONS

This study proposed an innovative approach based on the Academic Accreditation and Assessment Standard stated by NCAAA, on one hand, to match the program learning outcomes with the topics that will be covered on a course. On the other hand, to estimate topics distribution per weeks. The proposed approach presents a practical methodology allowing instructors to design an efficient course specification and keep the required quality with acceptable level since it shows how the instructor can cover topics over the week.

References

[1] National Commission for Academic Accreditation & Assessment (2009) National Qualifications Framework for Higher Education in the Kingdom of Saudi Arabia, Available on:
[2] http://www.psu.edu.sa/AAPC/Files/Handbook2_QAASA.pdf
[3] A. L. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning-I." Information sciences 8.3 (1975): 199-249.
[4] Zadeh LA. Fuzzy sets. Inform Control 1965, 8:338–353.
[5] Visscher-Voerman, I., & Gustafson, K. (2004). Paradigms in the theory and practice of education and training design. Educational Technology Research & Development, 52(2), 69–89.
[6] Albashiry, Nabeel M., Joke M. Voogt, and Jules M. Pieters. "Maintaining Curriculum Consistency of Technical and Vocational Educational Programs through Teacher Design Teams." Man in India 96.1 (2016): 599-608.

ASSOCIATION RULES AS WAY FOR BUILDING OPTIMAL LEARNING PATH

Hasan Al-Shaikh
Department of Computer Science
AL Andalus University
Sana'a, Yemen

Mohammed Al-Sarem
Department of Information Science
Taibah University
Al Madinah Al Monawarah, KSA

Abstract

With huge number of learning materials distributed in the Internet, students are suffering finding the right materials. They spend a lot of time, firstly, in surveying and collecting such materials, and then, going through the documents to find the useful information. Furthermore, these materials may store in such way that make its reading is difficult. Thus, it will be great help if we guide students during their searching process. This paper presents a method that is based on well-known data mining technique for constructing browsing path. The proposed method combines two main approach, the learning path construction approach and the learning object (LO) recommendation approach. Firstly, we find a set of candidate LOs based on the well-known Apriori algorithm. Then, the learning path construction approach uses the extracted rules from the collected learning object to build a reading map.

Keywords: Adaptive and Intelligent Web-based Educational Systems (AIWBES); Association rules; Concept map; VSM; Prerequisite; Epistemological order

1. INTRODUCTION

As it is known that, most of the self-directed learners spend most of time in surveying and choosing the right learning materials collected from the Internet and most of those materials are imperfect and have no particular order in the content.

Adaptive and Intelligent Web-based Educational Systems (AIWBES) provide an alternative to the traditional "just-put-it-on-the-Web" approach in the development of Web-based educational courseware (Brusilovsky and

Miller, 2001). This can be explained by the multitude of advantages that it offers. In particular, these systems accumulate a great deal of information which is very valuable in analyzing students' behavior and assisting them in the detection of possible errors, shortcomings and improvements of their learning process by building optimal pedagogical path (Samia Azough et al., 2010). According to (Romero and Ventura, 2006, 2007) the use of data mining is a promising area in the achievement of these objectives. In recent years, researchers have begun to investigate various data mining methods in order to help learners improve their programming learning process. Some of the most useful data mining tasks and methods are clustering, classification and association rule mining for example, in the National University of Singapore the data mining application have been used for classifying and selecting those students who need extra classes in a given subject. With the help of data mining they are able to select the targeted students much more precisely than by traditional methods (Ma et al., 2000). (Myller al., 2002) applied EM (Expectation-Maximization) – algorithm for clustering the students to construct homogeneous groups in terms of programming skills according to the students' skills, to predate exam results according to the skills shown in exercises. (Mostafavi and Barnes, 2010) suggested to use educational data mining methods as a way for creation a system that can judge a student's performance by the way he/she responds to questions (they gathered data from an introductory programming course teaching C++) for determining where the student needs to help. (Dominguez et al., 2010) integrated data mining into an e-learning system to generate dynamically tailored hints for students who are completing programming exercises during a national programming online tutorial and competition.

These methods uncover new, interesting and useful knowledge based on students' usage data. Some of the mains e-learning problems or subjects to which data mining techniques have been applied (Castro et al., in press) are provide course adaptation and learning recommendations based on the students' learning behaviour, dealing with the evaluation of learning material and educational web-based courses, provide feedback to both teachers and students of e-learning courses, and detection of atypical student's learning behaviour. This knowledge, however, can be useful not only to the providers (educators) but also to the users themselves (students), as it can be oriented towards different ends for different partakers in the process (Zorrilla et al., 2005). It could be oriented towards students in order to recommend learners'activities, resources, suggest path or simply links that would favor and improve their learning. In this paper, we propose an approach to build the suitable pedagogical path based on the association rules. The remainder of the

paper is organized as follows. In section 2, we discuss the related work. Proposed approach is given in Section 3.In Section 4 we use an example to illustrate the process of constructing concept maps based on the proposed method. Section 5 concludes the whole paper and discusses the future work.

2. RELATED WORK.

The e-learning systems act as an adaptive system if they select the path of learning that meet the learner's requirements and needs and discard those paths, which are not in accordance with these needs (Alian and Jabri, 2009), to achieve this goal, learner model and domain model are two of the key problems. A human behavior based learner model can be learned by observing the learner's actions therefore learner model can be built based on learner's behavior during solving task. (Tsai et al., 2001) used two-phase fuzzy mining and learning algorithm. They integrated an association rule mining algorithm, called Apriori, with fuzzy set theory to find embedded information that could be fed back to teachers for refining or reorganizing the teaching materials and tests. In a second phase, they used an inductive learning algorithm of the AQ family: AQR, to find the concept descriptions indicating the missing concepts during students' learning. The results of this phase could also be fed back to teachers for refining or reorganizing the learning path. (Chen et al., 2005) used association rule learning to discover common learning misconception of learning. According to that, the obtained rules can be applied to tune courseware structure through modifying the difficulty parameters of courseware in the courseware database. (Carchiolo et al., 2002) proposed an adaptive system for e-learning, which provides students with all paths from an initial knowledge to a desired one. The paths are retrieved and optimized based on student profile and teacher profile. Thus discarding those paths, which are not in accordance with the student's needs and the remaining paths are presented to the student to select one path and learn its course units. (Chen et al., 2006) presented a Personalized Web-based Instruction System (PWIS) to construct suitable learning pathway based on a modified item response theory for helping learning. (Colace et al., 2005) also presented an approach, which can obtain the learning style and capabilities of each learner to arrange the learning path adaptively with most suitable teaching contents. (Zhang et al., 2007) proposed an assessment model based on Bayesian Networks, which assesses learning status by knowledge map after absorbing and analyzing test results and creates learning guidance. (Hsieh and Wang, 2010) presented a web-based learning support system that uses both the preference-based and the correlation-based algorithms for recommending the most suitable learning objects or documents for each unit of the courses in order to facilitate more

efficient learning for the learner. (Lee et al., 2009) proposed to apply the algorithm of Apriori for Concept Map to develop an intelligent concept diagnostic system (ICDS). This system provides teachers with constructed concept maps of learners rapidly, and enables them to diagnose the learning barriers and misconception of learners instantly. (Azough et al., 2010) proposed an adaptive system in order to generate pedagogical paths which are adapted to the learner profile and to the current formation pedagogical objective. They have studied the problem as an "Optimization Problem". Using Genetic Algorithms, the system seeks an optimal path starting from the learner profile to the pedagogic objective passing by intermediate courses. (Huang et al., 2007) also used genetic algorithm and case-based reasoning to construct a near-optimal learning path. In this paper we present an algorithm for building adaptive learning system that finds the suitable learning path based on the Concept Map and Good Learners Average Rating (GLAR).

3. RESEARCH APPROACH

In a learning activity, each step has a learning focus, which was called ‘‘concept”. The learning of these concepts should be done in a proper sequence. This kind of learning sequence called epistemological order (Chen et al., 1999). ‘‘Epistemological order” is used to standardize the learning order of different concepts. Let's present the epistemological order in the Fig. 1, C_i and C_{i+1} represent two concepts. The connecting line between them represents that there is a correlation of a certain epistemological order in between.

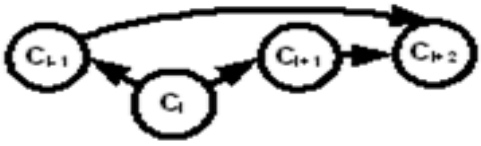


Fig.1. Epistemological order of concept

The arrow of the connected line represents the learning order. Therefore, Fig. 1 indicates that concept C_i precedes concept C_{i+1} in terms of the epistemological order. In addition, we can presume that the learner will have difficulty if he/she decided to read concept C_{i+1} before concept C_i. In general, the learner’s learning difficulty of concept C_{i+1} is caused by incomplete learning of concept C_i. Therefore, If C_i is a prerequisite to concept C_{i+1}, then teachers could identify the learning problems of students by tracing the relationship between concepts (Cheng et al., 2005; Hwang, 2003; Hwang et al., 2008), furthermore, it can be used to demonstrate how the learning status of a concept can possibly be influenced by learning status of other concepts

and give learners adaptive learning guidance, therefore to provide learning suggestions to individual students, we firstly construct the concept map, then we suggest a personalized material according to social learning theory (Bandura, 1977) that strongly supports idea that people can learn by observing the behavior of others and outcome of behavior of good learner can increase their performance.

Therefore, the proposed approach uses the algorithm of Apriori for providing the learning guidance and Good Learners Average Rating to guide learners in selecting good learning resources in order to improve their learning process.

Phase 1: *The construction of concept map*

Phase 2: *Generate the suitable learning path*

Phase 3: *Suggestion of adaptive materials*

Phase 1: *The construction of concept map*

Assuming that there are ***n*** learning objects $LO_1, LO_2, \dots, LO_n$ in learning objects repository and let each learning object ***LO*** contains a ***m*** set of concepts: ***LO***=$\langle C_1, C_2, C_3, \dots, C_m \rangle$.

The construction of concept map procedure includes five steps.

Step 1: Determining the learner's goal

Let's after analysis the learner's testing, found that the problem lies in the lack of understanding of a concrete concept and suppose that learner intends to learn more about this concept.

Step 2: This step consists of 2 sub steps, shown as follows:

Step 2a: Identify the candidate Learning Objects

Based on the learner's goal, the learning object repository (LOR) will search to find ***n*** learning objects that contained ***m*** related concepts. For simplicity, assume that the LOR found five learning object that are associated with the learner's goal and contain contents for among C_1, C_2, C_3, C_4 and C_5 concepts respectively where C_2 is a learner's goal as shown in Table 1.

Table 1. Learning objects that associated with the learner's goal

Learning Objects	Related Concepts
LO_1	C_1, C_2, C_3, C_4 and C_5
LO_2	C_2, C_3 and C_5
LO_3	C_1, C_2, C_3 and C_5
LO_4	C_2 and C_5
LO_5	C_2 and C_4

Applying the Apriori algorithm (Agrawal and Srikant, 1994), we identify the largest item set by extract association rules among all collected LO, where an itemset is called a large-itemset if its support value is greater or equal to the user-specified support threshold (called minSupport) and an association rule is an expression $X \rightarrow Y$ where X and Y are disjoint itemsets, which represents possibility when X appears that Y will also appear. The support of an association rule is the support of X∪Y, and the confidence of such a rule is the fraction of all transactions containing X that also contain Y (Hidber, 1999). The following is the detail of the Apriori algorithm.

Apriori algorithm

Input: Learning objects (**LO**), Threshold of minimum support value (**minSupport**).

Output: Large itemsets in learning objects (**LI**).

Procedure:

1: LI_1 = find large 1-itemsets in **LO**.

2: For (k=2; $LI_{K-1} \neq \phi$;k++){

C_k = apriori-gen(LI_{K-1}); // New candidates

3: for all of records r $\in$ **LO** {

C_t = subset(C_k, r) // Candidates contained in r

for all of candidates c $\in C_t$

c.count++;}

LI_k ={c ∈C_K | c.count ⩾minSupport}}

4: **Return** LI= $\cup_k LI_k$;

After the operation of the Apriori algorithm, the large itemset is{ C_2, C_3, C_5} and the confidence of an association rule "$C_i \rightarrow C_j$" is calculated as follows:

$conf(c_i \rightarrow c_j)=\frac{Sup(c_i,c_j)}{Sup(c_i)}$, where C_i is a concept in the large 1-itemset, C_j is a concept in the learning object, "$C_i \rightarrow C_j$" denotes the association rule from C_i to C_j , "$conf(C_i \rightarrow C_j)$" denotes the confidence of the association rule "$C_i \rightarrow C_j$", "$Sup(C_i, C_j)$" denotes the support of the 2-itemset "(C_i, C_j)" , "$Sup(C_i)$" denotes the support of the large 1-itemset C_i ; $i \neq j$; $1 \leq i \leq n$ and $1 \leq j \leq n$.

In Table 2 shown the confidence of an association rule "$C_i \rightarrow C_j$"

Step 2b: Find out all of the learning objects which are designated for the candidate concepts

Table2.The confidence of an association rule"$C_i \rightarrow C_j$"

Rules	Confidence
$C_1 \rightarrow C_2$	1
$C_1 \rightarrow C_3$	1
$C_1 \rightarrow C_4$	0.5
$C_1 \rightarrow C_5$	1
$C_2 \rightarrow C_1$	0.4
$C_2 \rightarrow C_3$	0.6
$C_2 \rightarrow C_4$	0.4
$C_2 \rightarrow C_5$	0.8
$C_3 \rightarrow C_1$	0.667
$C_3 \rightarrow C_2$	1

$C_3 \rightarrow C_4$	0.33
$C_3 \rightarrow C_5$	1
$C_4 \rightarrow C_1$	0.5
$C_4 \rightarrow C_2$	1
$C_4 \rightarrow C_3$	0.5
$C_4 \rightarrow C_5$	0.5
$C_5 \rightarrow C_1$	0.5
$C_5 \rightarrow C_2$	1
$C_5 \rightarrow C_3$	0.75
$C_5 \rightarrow C_4$	0.25

Step 3: Presetting the relevance between concepts and learning objects by a teacher

Assume that $\boldsymbol{r_{ij}}$ denotes the ratio of concept $\boldsymbol{C_i}$ contained in learning object $\boldsymbol{LO_j}$ that indicates the weight of the concept contained in the learning object. Then, we can get the Concept-Learning object matrix $\boldsymbol{R}$, shown as follows:

$$R = \begin{array}{c} \\ C_1 \\ C_2 \\ \vdots \\ C_m \end{array} \begin{array}{c} \begin{array}{cccc} LO_1 & LO_2 & \cdots & LO_n \end{array} \\ \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m1} & \cdots & r_{mn} \end{bmatrix} \end{array},$$

where $r_{ij} \in [0,1]$, $r_{ij} = 0$ denotes that the concept C_i doesn't appear in learning object LO_j, $r_{ij} = 1$ otherwise, $0 \leq i \leq m$ and $0 \leq j \leq n$. As shown in Table 3, five LOs contain these related concepts C_1, C_2, C_3, C_4 and C_5.Each $r_{C_i LO_j}$ is set by teacher.

Table 3. Concept-Learning Object relevance mapping table

Concept	Learning Objects						
	LO_1	LO_2	LO_3	LO_4	LO_5	LO_6	LO_7

C_1	0.2	0.0	0.3	0.0	0.0	0.0	0.0
C_2	0.3	0.3	0.3	0.8	0.7	0.0	0.0
C_3	0.1	0.2	0.3	0.0	0.0	0.1	0.7
C_4	0.2	0.0	0.0	0.0	0.3	0.0	0.0
C_5	0.2	0.5	0.1	0.2	0.0	0.4	0.0
C_6	0.0	0.0	0.0	0.0	0.0	0.0	0.3
C_7	0.0	0.0	0.0	0.0	0.0	0.5	0.0

Step 4: Calculate the relevance degree between LOs.

For all association rules type"$C_i \rightarrow C_j$" obtained in Step 2, we calculate the relevant degree $rev(LO_i \rightarrow LO_j)_{C_xC_y}$between learning object LO_i and LO_j from the relationship"$C_x \rightarrow C_y$", shown as follows:

$$rev(LO_i \rightarrow LO_j)_{C_xC_y} = Min(W_{C_xLO_i}, W_{C_yLO_j}) * conf(C_x \rightarrow C_y) \quad (1)$$

where" $rev(LO_i \rightarrow LO_j)_{C_xC_y}$"denotes the relevance degree of the relationship "$LO_i \rightarrow LO_j$"converted from the relationship"$C_x \rightarrow C_y$", $rev(LO_i \rightarrow LO_j)_{C_xC_y} \in [0,1]$, $W_{C_xLO_i}$denotes the weight of the concept $C_x$in the learning object LO_i, $W_{C_yLO_j}$denotes the weight of the concept $C_j$in the Learning object LO_y, "$conf(C_x \rightarrow C_y)$"denotes the confidence of the relationship $C_x \rightarrow C_y$", $x \neq y$; $1 \leq x \leq m$, $1 \leq y \leq m$, $1 \leq i \leq n$ and $1 \leq j \leq n$. Furthermore, let $conf(LO_x \rightarrow LO_y)$ be the confidence of the relationship $LO_i \rightarrow LO_j$. If there is more than one relationship between any two constructed LOs, the relationship between them chooses as follows:

$rev(LO_i \rightarrow LO_j) = Max(rev(LO_i \rightarrow LO_j)_{C_xC_y})$ (2) Based on Eq. (1) mentioned above the relevance degree for the relationship $LO_1 \rightarrow LO_2$, calculated as follows:

$C_1 \rightarrow C_2$: Min (0.2, 0.3) ×1=0.2 (confidence=1)

$C_1 \rightarrow C_3$: Min (0.2, 0.2) ×1= 0.2(confidence=1)

$C_1 \rightarrow C_5$: Min (0.2, 0.5) ×1=0.2 (confidence=1)

$C_2 \rightarrow C_3$: Min (0.3, 0.2) ×0.6=0.12 (confidence=0.6)

$C_2 \rightarrow C_5$: Min (0.3, 0.5) ×0.8=0.24 (confidence=0.8)

$C_3 \rightarrow C_2$: Min (0.1, 0.3) ×1=0.1 (confidence=1)

$C_3 \rightarrow C_5$: Min (0.1, 0.5) ×1=0.1 (confidence=1)

$C_4 \rightarrow C_2$: Min (0.2, 0.3) ×1=0.2 (confidence=1)

$C_4 \rightarrow C_3$: Min (0.2, 0.2) ×0.5= 0.1 (confidence=0.5)

$C_4 \rightarrow C_5$: Min (0.2, 0.5) ×0.5=0.1 (confidence=0.5)

$C_5 \rightarrow C_2$: Min (0.2, 0.3) ×1=0.2 (confidence=1)

$C_5 \rightarrow C_3$: Min (0.2, 0.2) ×0.75= 0.15 (confidence=0.75)

And based on Eq. (2) the maximum value among these relevance degrees is:

$$\text{rev}(LO_1 \rightarrow LO_2) = 0.24$$

$$\text{conf}(LO_1 \rightarrow LO_2) = \text{conf}(C_1 \rightarrow C_2) = 0.8$$

By analogy, we calculate the relevance degree for all relationship $LO_i \rightarrow LO_j$ according to that we can get the learning objects-relationship table, as shown in Table 4.

Learning Object-relationship	***Relevance degree***	***Confidence***
$LO_1 \rightarrow LO_2$	0.24	0.8
$LO_1 \rightarrow LO_3$	0.2	1
$LO_1 \rightarrow LO_4$	0.2	1

$LO_1 \rightarrow LO_5$	0.2	1
$LO_2 \rightarrow LO_1$	0.3	1
$LO_2 \rightarrow LO_3$	0.3	1
$LO_2 \rightarrow LO_4$	0.5	1
$LO_2 \rightarrow LO_5$	0.5	1
$LO_3 \rightarrow LO_1$	0.3	1
$LO_3 \rightarrow LO_2$	0.3	1
$LO_3 \rightarrow LO_4$	0.3	1
$LO_3 \rightarrow LO_5$	0.3	1
$LO_4 \rightarrow LO_1$	0.2	1
$LO_4 \rightarrow LO_2$	0.4	0.8
$LO_4 \rightarrow LO_3$	0.2	1
$LO_4 \rightarrow LO_5$	0.24	0.8
$LO_5 \rightarrow LO_1$	0.16	0.8
$LO_5 \rightarrow LO_2$	0.4	0.8
$LO_5 \rightarrow LO_3$	0.3	1
$LO_5 \rightarrow LO_4$	0.3	1

Phase 2: *Generate the suitable learning path*

To fulfill a goal issued by a learner, learning objects must be arranged subsequently according to their weight and taught in the front side of the learning path, therefore for each relationship "$LO_i \rightarrow LO_j$" add an edge

from LO_i to LO_j. The following algorithm briefs the learning path building procedure:

Learning Path Generation Algorithm

Input: Threshold value (θ), relevance degree matrix(REV), learning objects (LO) and total number of LOs (n).

Output: A suitable learning path (LP).

Procedure:

1: ***LP*** = {0}

2:While (length of LP < n){

if ($REV_{ij} < \theta$)

Set $rev_{ij} \leftarrow 0$

else

Arrange each LO_i according to their weight.

Add the $LO_i$to ***LP***

3: Return ***LP***.

[Step 3b]: Generate the optimal learning path

Assume that, the threshold value given by learner is 0.75 then the previous figure will appear as follows:

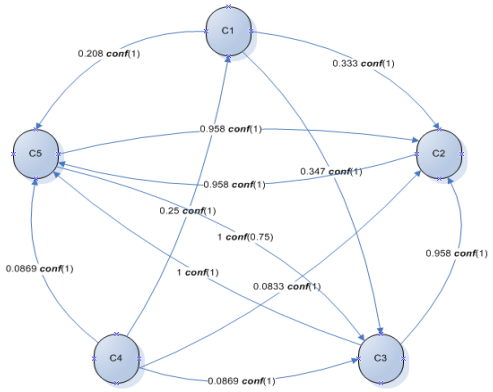


Fig3. Concepts-relationship map according step3b

[Step 3c]: Generate the optimal learning path

If a learner obtained to learn concept C_1 then the recommender concepts are C_3, C_2 and C_5, respectively, where $C_3 \in LO_3$, $C_2 \in LO_2$ and $C_5 \in LO_5$(for all other see Table 4.)

Table4.Recommender learning path

Learning Object	*Prerequisites LOs*	*Recommending LO*
C_1	C_4	C_3, C_2 and C_5
C_2	$(C_3, C_5), C_1$and C_4	C_5
C_3	$C_5, C_1 and\ C_4$	C_2
C_4	_	$C_1, (C_5, C_3) and\ C_2$
C_5	$C_1, C_2, C_1 and\ C_4$	$C_3$and C_2

4. Future work and conclusion

In this paper, we have presented a new method for constructing learning path based on data mining techniques for adaptive learning systems. The proposed method combines tow main approach, the learning path construction approach and the learning object recommendation approach. Firstly, we find a set of candidate LOs based on the well-known Apriority algorithm. Then, the learning path construction approach use the rules extracted from the collected learning object to build a Concept map, moreover, in order to assist a learner

to study efficiently, the recommendation approach arranged the collected learning objects in the correct order to understand them efficiently.

In the future, we will try to integrate the proposed method with other efficient method. We also hope to study the user feedback and the impact on the profile of the learner.

References:

[1] Brusilovsky, P. and Miller, P. (2001). Course Delivery Systems for the Virtual University. In: Tschang, T. and Della Senta, T. (eds.): Access to Knowledge: New Information Technologies and the Emergence of the Virtual University. Elsevier Science, Amsterdam, 2001, 167-206.

[2] Samia Azough, Mostafa Bellafkih and El Houssine Bouyakhf, "Adaptive E-learning using Genetic Algorithms",International Journal of Computer Science and Network Security, Vol. 10, No. 7, July 2010.

[3] Romero, C., & Ventura, S. (2006). Data mining in e-learning. Southampton, UK: Wit Press.

[4] Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. Expert Systems with Applications, 33(1),135–146.

[5] Andrade MA, et al. (2000) Homology-based method for identification of protein repeats using statistical significance estimates. J Mol Biol 298(3):521-37.

[6] Myller N., Suhonen J, Sutinen E. Using data mining for improving web-based course design,Proceedings ICCE'02 of the International Conference on Computers in Education,Auckland, New Zealand vol.2; December, 2002. pp.959 – 963.

[7] Behrooz Mostafavi, Tiffany Barnes:Towards the Creation of a Data-Driven Programming Tutor. Intelligent Tutoring Systems (2) 2010: 239-241.

[8] Dominguez, M.D., Saka, M.C., Lieb, R., Wittchen, H.U., van Os, J., 2010. Early expressionof negative/disorganized symptoms predicting psychotic experiences and subsequent clinical psychosis: a 10-year study. Am. J. Psychiatry 167, 1075–1082.

[9] A. Castro et al.(1992) "productive systems and international integration"international integration labour market organisation ,London:Academic Press.

[10] Zorrilla, M. E., Menasalvas, E., Marin, D., Mora, E., & Segovia, J. (2005).Web usage mining project for improving web-based learning sites. InWeb mining workshop, Cataluna.

[11] Marwah Alian and Riad Jabri, (2009), A Shortest Adaptive Learning Path in eLearning Systems: Mathematical View,Journal of American Science, 2009.

[12] Tsai Jeanne L,Yu-Wen Ying, Peter Allen Lee,Cultural predictors of self-esteem: A study of Chinese American female and male young adults.Cultural Diversity and Ethnic Minority Psychology 7 (3), 284.

[13] Chen L, F Yang, J Yang, X Zhang, Y Jiang, Y Yan, X Tang, J Wang., 2005,Genome dynamics and diversity of Shigella species, the etiologic agents of bacillary dysentery.Nucleic acids research 33 (19), 6445-6458.
[14] Carchiolo, V., Longheu, A., & Malgeri, M. (2002). Adaptive formative paths in a Web-based learning environment.Educational Technology & Society, 5 (4), 64-75.
[15] Chen, B.E., Kondo, M., Garnier, A., Watson, F.L., Puettmann-Holgado, R., Lamar, D.R., Schmucker, D. (2006). The molecular diversity of Dscam is functionally required for neuronal wiring specificity in Drosophila. Cell 125(3): 607--620.
[16] Colace, F., M. Desanto and M. Vento, 2005. E-learning platform developing an evaluation strategy in real case. Proceedings of the 35th Annual Conference on Frontiers in Education, Oct. 19-22, Indianopolis, IN., pp: F1C-20-F1C-20.
[17] Zhang Z, et al. (2007) Genome-wide identification of spliced introns using a tiling microarray. Genome Res 17(4):503-9.
[18] Hsieh,T.,C.,& Wang, T.l.(2010).a mining-based approach ondiscovering courses pattern for constructing suitable learning path.Expert systems with applications,37(6),4156-4167.doi:10.1016/j.cswa.2009.11.007.
[19] Samia Azough, Mostafa Bellafkih and El Houssine Bouyakhf,"Adaptive E-learning using Genetic Algorithms", International Journal of Computer Science and Network Security, Vol. 10,No. 7, July 2010.
[20] Huang, X., Hood, E., and Cai, L. (2007) Protein tertiary structure prediction by protein threading: a survey, Plant Biotechnology Journal, to appear.
[21] Chen, C.& Kelly, G. J. (1999). The sound of music: Constructing science as sociocultural practices through oral and written discourse. Journal of Research in Science Teaching, 36, 883–915.
[22] Cheng,Jill et al.5-Nucleotide Resolution Transcriptional Maps of 10 Human Chromosomes at;DOI: 10.1126/science.1108625 Science 308, 1149 (2005).
[23] Hwang E, et al. (2003) Spindle orientation in Saccharomyces cerevisiae depends on the transport of microtubule ends along polarized actin cables. J Cell Biol 161(3):483-8.
[24] Hwang Y.,M.Y. Yi,/ Int. J. Human-Computer Studies 59 (2003) 431–449.
[25] Hwang, G.,J.,Tsai, C.C.,and Tseng,J.C.R (2008) A novel approach for assisting teachers in analyzing student web-searching behaviors. computer and Education,52(1),926-938.
[26] Bandura, A. (1977). Social learning theory. Englewood Cliffs, NJ: Prentice Hall.
[27] R. Agrawal and R. Srikant. Fast algorithms for mining association rules in large databases. Research Report RJ 9839, IBM Almaden Research Center, San Jose, California, June 1994.
[28] C.Hidber, Online association rule mining. In:Proceedings of the ACM SIGMOD Conference on Management of data 1999:145-156.

Improved Model For Semantic Information Retrieval

Mossa ghurab
Department of Computer Science
Sana'a University
Ghurab@gmail.com

Hiba Mohammed Al-Marwi
Department of Computer Science
Sana'a University
Hiba_almarwea@yahoo.com

Abstract:- *with a rapid expansion of new available information presented to us on the internet ,information retrieval gaining importance. Several approach introduce in this filed which record success ,but it take long time. in this paper we discuses a model based on ontology which we built using naives bayes algorithm, and apply k-means clustering algorithm to find semantically similarly terms on the Ontology ,Also we present document using improved concept vector model ICVS which is improved model for traditional CVS also we apply fuzzy classification to limited number of retrieval document to particularly threshold.*

***Index Terms*—concept Relevance, Concept vector space, Query Expansion ,K-means clustering, Naivse Bayes .**

I. INTRODUCTION

As number of document on the internet increase in everyday life. Traditional approach on information retrieval proven to be less efficient in providing relevant information to user query .Several methodology appears to generate more accuracy result. To achieve this goals new upcoming semantic approach try to establish a semantic relationship among the document . In semantic web information is stored in conceptual hierarchy referred as ontology to build some domain specified ontology. we use Naïve Bayes algorithm . Many researches are developed to enhance Naïve Bayes [1] [2] [3].We use naives bayes using map reduce to enhance its performance[5].

Text document are represent using vector space model[4]. Vector space model consist of concepts extracted from ontology and concept relevance which is calculate using frequency of concept occurrence . We use improved vector space model ICVS [4]. which take into account both concept frequency and concept important which computed using page rank algorithm as one of the most well known of markov based algorithms[4].

Fuzzy logic very useful in information retrieval .IR system find difficulty to make decision in providing accurate information .Each element of fuzzy set has a membership value between 0 to 1.This membership function play important role in defining degree of membership of element in fuzzy set.

Rest of the paper is organized as section 2 provide related work; section 3 presented discuses and section 4 conclude paper and finally section 5 presented my opinion.

II. RELATED WORKS

There is a lot of research in IR field such as[1][2] [5] which extract concept from internet using Naïve bayes algorithm which used as preprocessing step to build specific domain ontology.

Another research focus on enhance models which used to represented document as helper technique to enhance information retrieval[4].

In [6] research developed approach to get more relative document to user query by using k-means clustering algorithm fuzzy classification ,query expansion technique .

III. PROPOSAL MODEL

We use architecture (as shown in fig. 1) for semantic information retrieval based on document [6]

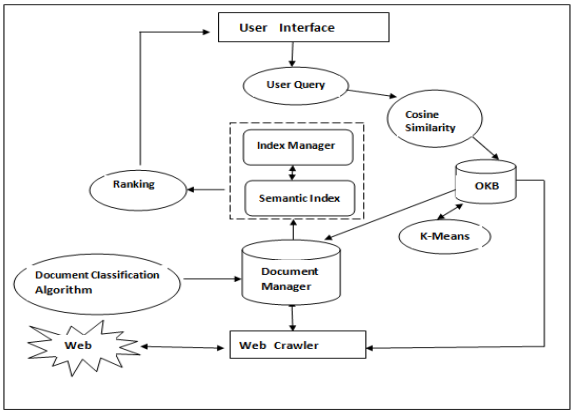


Fig 1:Overall Architecture for proposed system

A. Naivse Bayes

In the beginning we build our ontology using Naïve Bayes algorithm. Naïve Bayes algorithm has two phases .First phase is learning process to build some knowledge .And second phase is document classification. Features extraction is needed to help classifier to learn or classify document .We use standard bag-of words representation as feature in our classification algorithm . Naivse Bayes classification use probability to find output target class with maximum posterior

$$\hat{y} = \underset{c_i \in C}{argmax}\ P(c_i|D) \qquad (1)$$

we can define Naïve Bayes classifier model in equation 1.

Where :

- D : document

- c_i :target class c_i in C

- C : all target class in dataset

- P(ci|D) : posterior probability class ci if given document D

We implement Naïve Bayes using map reduce programming model on both phases. Algorithm1 and 2 show using map and reduce in classifying phase

Algorithm 1 Map Function for Naïve Bayes Classification Algorithm

```
Input : document,learning_knwlege,list_of_class
Output: Intermediate Key, Intermediate Value

Nb_model  build Classifier(Learning_Knowlege)

For each c in list_of_class do
      Prior ← nb_model.getPriorProbability(c)
       For each token in document do
           Temp ← nb_model.getPosterior(c,token)
          Posterior ← Posterior X temp
          End for
         Probability ← prior X Posterior
        emit Intermediate(document,(class, Probability))
end for
```

Algorithm 2 function for Naïve Bayes Classification Algorithm

```
Input :key, list(Value)
Output: Key, Value
MaxValue ← -1
For each value in list(value) do
  If (value. Probability ≥ maxValue then
    maxValue ← value. Probability
    tclass ← value. Class
  end if
end for
emit(key, tclass)
```

B. IMPROVED CONCEPT VECTOR MODEL

Firstly k-means clustering algorithm is applied to the ontology to find semantically related terms between ontologes using mapping technique which maybe Syntactic, Semantic or comparing weight .Secondly if user enter a query through user interface the cosine similarity algorithm apply between user query and clusters which yield from apply k-means algorithm in ontology as show in Algorithm 3 Then extracted concept submit to web crewel to get most related document to user query. Return document must process by removing stop words and convert every word to it's root .To find number of semantic term in each document .Document must represent using improved vector space model ICVS which is enhance form of traditional CVS . ICVS take into account concept relevance and concept important to calculate concept weight The Proposed Model to computing concept weight as shown in fig . 2 [4]

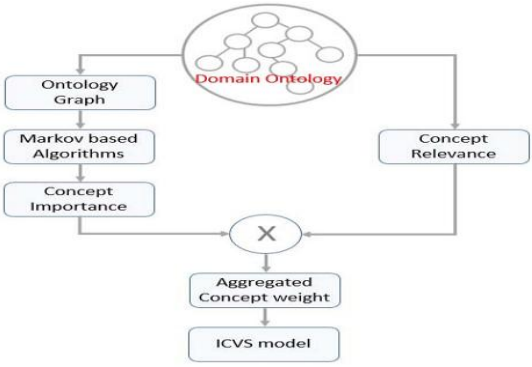


Fig 2 :The proposed model for an improved concept vector space .

ConceptWeight(w_i)=Concept Important(Imp) +Relevance(Rel)

Where :

Concept Important: which computed by the number of relations a concept has to other concepts

Concept Relevance : which is computed using the frequency of concept occurrences in the dataset

Concept important computed using page rank algorithm as one of the most well known example of markov based algorithm

Page rank algorithm .It assigns an initial importance to a vertex i as show in equation 2.

$$PR^{(0)}(i) = \frac{1}{N} \quad (2)$$

Page rank iterates as per equation 3 and continues to iterate until a convergence criterion is satisfied .

$$PR^{(k+1)}(i) = \sum_{j \in V_i} \frac{PR^{(k)}(j)}{Outdegree(j)} \quad (3)$$

The final step of proposed model is building a concept vector space model .This model consists of two components: concept and their weights. Concepts are taken from our domain ontology using matching method. The idea of matching method is to search for concept in the ontology which matching either partially or exactly/fully with a dataset term. The mapping of term $t_i \in d$ into concept $c_j \in O$ is defended as exact match $EM(t_i,c_j)$

$$EM(ti,cj) = \begin{cases} 1 & \text{if label}(c_j) = t_i \\ 0 & \text{if label}(c_j) \neq t_i \end{cases} \quad (4)$$

Or partial match where $PM(t_i,C_j)$

$$PM(ti,cj) = \begin{cases} 1 & \text{if label contain } t_i \end{cases} \quad (5)$$

If EM(ti,cj) =1 or PM(ti,cj)=1 then term ti is replaced with concept cj Finally , a document is represented using concept space representation model by the following tuple:

$$d_i = \{ (c_1,w_1),(c_2,w_2),(c_3,w_3),\ldots\ldots\ldots(c_i,w_i)\} \quad (6)$$

After represent document in ICVS .Then membership function pi(x) calculate by finding the number of semantic terms presented in each document based on the pi(x) value the document classified as document close to user query and document far from user query. Document which stratify specific threshold value store with their membership value in semantic index. Then document are ordered in decreasing order

of their value and presented to user as final response. Algorithm 4 show all above steps.

Algorithm 3 extract most relevant term

```
Let Q User Query
Let C Set Of Cluster
Let Co Set Of Concepts In Cluster

For each wi in Q          do
    For each ci in C          do
        For each coi in Co  do
          Apply k-means between coi and wi
           List1← List1.AddSimilrityValue()
        End For
      Maxsim ← List1.FindMaxSim()
      List2 ←— List2.Add(Maxsim)
   End For
  MaximumSimall cluster← List2. FindMaxSim()
  List3 ←List3.add(MaximumSimall cluster )
 End For
MaximumSimcluster← List3. FindMaxSim()
Submit all concept in cluster in
MaximumSimcluster
To web crawler to get web document contain only
this concept.
```

Algorithm 4 Document Classification

let S(D) is the set of document i.e D(Web Document). SK,Semantic Keywords. $SK_j(D_i)$ is the $_j$th term in I^{th} document.

Input: set of document fetched by crawler

Output: Classified document within a boundary

Step 1: A set of Document $\sum_{i=1}^{n} s(d_i)$ is fetched by the web crewel for concept fetched through ontology

Step 2: Find the number of semantic terms presented in each document $\sum Sk(D_i)$,where SK_j is the set of semantic terms found in Di

Step 3: Find the percentage of keyword match i.e P(x) of each document Di $Pi(x)=(Sk_j/SK)*100)$

Step 4: Compute membership function (μd(x)) for each document as $\mu d_i(x)$ Pi(x)

Step 5: Classify the document through creating a fuzzy Set of document by setting a boundary of fuzzy value Through membership function as:

$$K[i]=\begin{cases} 0 & \text{If } Mdi(x)=0 \\ 1 & \text{if } \mu di(x)=1 \\ 1 < \mu di(x) < 0 & \text{Otherwise} \end{cases}$$

Step 7:Associate SKj to Di along with its url and value of K index and create semantic index

IV. CONCLUSION AND FUTURE WORK

In this paper we View a model to semantic information retrieval using specific domain ontology which build using Naïve Bayes algorithm ,and use fuzzy classification document which represented using ICVS. A Threshold value used to limited number of return document. In future the approach may be generalized for various domain

References

[1] He Youquan, and Xie jianfang and Xu cheng, “An improved Naive Bayesian algorithm for web page text classification ”. chines, , Apr. 2011.

[2] Yu Wang Weikang and Qian, Shuchang and Zhang, Xiaoyao ," A Learning Algorithm for Bayesian Networks and Its Efficient Implementation on GPUs ,2016 .

[3] T. L. Gilbert, Formulation,"Chinese web text classification system model based on Naive Bayes, July.2015

[4] Zenun Kastrati , Ali Shariq Imran and Sule Yildirim Yayilgan,"An Improved Comcept Vector Space Model for Ontology Based Classification", 2015

[5] Joan Santoso, Eko Mulyanto Yuniarno, and Mochamad HariadiC. " Large Scale Text Classification using Map Reduce and Naive Bayes Algorithm for Domain Specified Ontology Building",2015.

[6] Kajal Joshi, Ashish Verma, Ankita Kandpal, Shalini Garg, Rashmi Chauhan and R.H Goudar," Ontology based Fuzzy Classification of Web Documents for Semantic Information Retrieval,2014.

Policy Makers' Preceptions twords Success and Failure Factors of e-Government project implementation in Yemen:

An explore study

Mahdi Abdullah Alsebaei,
Ph.D,Technology Management & Policy Program
College of Engineering, Seoul National University
Seoul, South Korea.
Assistant Professor at School of Computer and Information Technology,
AL-Razi University, Sana'a, Republic of Yemen

E-mail: mahdiyemen80@gmail.com

Abstract:

The aim of this paper is to explore the current situation of e-government project's implementation in Yemen and understand the success and failure factors from the perception of the Yemeni e-government project team. The methodology is exploratory method which is to investigate the academic literature about e-government implementation in developing countries focusing on the success and failure factors of implementation. Also the structured telephone interviews were done with the top decision makers of the e-government in Yemen. The situation of Yemen in the area of communication and technologies will be presented. The findings show that Yemen started to establish a good plan for e-government projects and the implementation concept started to be well organized by the e-government project's team. The researcher as a member of this team has identified many activities in order to determine the most significant factors that help the government for a proper implementation for the system of e-government. However, Yemen's government still need to work hard regarding the Internet quality and the level of education for citizens in terms of computer technologies and IT skills. Yemen also needs to overcome the issue of the gap between genders as well as between urban and rural.

The study will address important and significant factors preventing the implementation of e-government and establishing of e-services in Yemen. Finally; the requirements for successes E-government implementation will be elaborated. And the study ends up with Recommendations as solutions for better E-government implementation in the future for the Republic of Yemen.

Keywords: E-government, Success factors, failure factors, E-government in Yemen, Yemen.

I. INTRODUCTION

Electronic government (e-government) suggests the use of information technology to enable and improve the efficiency with which government services are provided to citizens, employees, businesses and agencies; it increases the convenience and accessibility of government services and information to citizens, (Carter and Belanger, 2005).

E-government comprises the use of ICT in order to deliver public services to citizens and businesses. It entails the transformation of public services available to citizens using new organizational processes as well as new technological trends (Gunter, 2006).

E-government is designed to facilitate a more integrated mode of governance. It encapsulates the relationships between governments, their citizens as well as their suppliers by the use of electronic means (Means and Schneider, 2000). In brief, the goal of e-government is to achieve in making government services more accessible, more citizen focused, more relevant to citizens as well as more responsive to their needs and expectations.

In this perspective, e-government is the use of technology, especially web-based Internet applications in order to enhance access to and efficiently deliver government information and services (Brown and Brudney, 2001). Heeks (2003) asserts that the use of information and communication technologies as a means to improve the activities of public sector organizations is the definition of e-government. According to the World Bank (2004), e-government refers to the use of information technologies by government agencies that have the ability to transform relations with citizens, businesses, and other arms of government. Last but not least, according to the Centre for Democracy and Technology (2002), "e-government is the use of information and communications technologies in order to transform government by making it more accessible, effective and accountable".

The researcher conducted this study to review the situation of e-government in Yemen and analysis its related concepts which is among the early stage of E-government project. The study will analyze several factors that contribute to the success implementation of E-government project in local government of Sana'a city- the capital of Yemen.

II. E-GOVERNMENT CONCEPTS

A. Main objectives of e-government

- Make it easier to find information about government services.
- Improving access to government services by making them online 24/7.
- Helping to improve the "QoS" quality of services.
- Helping to interact more effectively with government organizations and related agencies so that more services can be provided.

- Helping to build trust between government
- Challenging the existing ways of working.

B. E-government Applications

■ Government-to-Government (G2G):

G2G is the interconnection between the government agencies. This function or application serves both internal processes and activities (between public organizations themselves) and external ones also (between government organizations, citizens and businesses). The ultimate aim of the (G2G) function is to enhance inter-government organizations' processes by streamlining collaboration and coordination.

In order to realize a single access point, collaboration and cooperation among different governmental departments and agencies is compulsory. Online communication and cooperation allows government agencies and departments to share databases, resources, pool skills and capabilities, enhancing the efficiency and effectivity of processes (Valantina, 2004).

■ Government-to-Citizens (G2C):

G2C is the interconnection between the government agencies and the public citizens. With this function or application, the public organizations publish information and contact details, and offer regular services online. The ultimate aim of this application is to give users different options and communication channels for government transactions.

G2C allows customers to access government information and services instantly, conveniently, from everywhere, by use of multiple channels (PC, Web TV, mobile phone or wireless device). It also enables and reinforces their participation in local community life “send an email or contribute to an online discussion forum” (Valantina, 2004).

■ Government-to-Businesses (G2B):

G2C is the government communication with the private business sector. G2B allows businesses as well as individuals to have transactions with the government, examples being: renewing registrations, lodging taxes, downloading tenders’ information, and many others. The government-to-business (G2B) application is as useful as the G2C system, enhancing the efficiency and quality of communication and transactions with business. Therefore, this should be considered by the governments.

Companies everywhere are conducting business-to-business e-commerce in order to lower their costs and improve inventory control. The opportunity to conduct online transactions with government reduces red tape and simplifies

regulatory processes, therefore helping businesses to become more competitive. The delivery of integrated, single-source public services creates opportunities for businesses and government to partner together for establishing a web presence faster and cheaper (Valantina, 2004).

- Government-to-Employees (G2E):

G2E is the interconnection between the government agencies and the employees from different sectors (public and private G2E is perhaps the least adopted application of e-government. Scholars and countries around the world usually focus on the first three applications only.

G2E is an effective way to provide e-learning, bring employees together and to promote knowledge sharing among them. It gives employees the possibility of accessing relevant information regarding: compensation and benefit policies, training and learning opportunities, civil rights laws, etc (Valantina, 2004).

However, Yemen is still in the first function G2G which is the interconnection between the government agencies. There is no full interaction with citizens because they can only brows the information about government services from the different government agencies through the government official portal which is announced in 2009. Moreover, there is no electronic interaction with the Yemeni government with the private business sector yet as well as the employees from both public and private sectors, these maybe targeted to be established after completing all the stages required to be implemented for the e-government system in Yemen.

C. Benefits of E-government

OECD (2003) thoroughly examined e-government initiatives in its members' countries (Australia, Austria, Belgium, Canada, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Japan, Korea, Luxemburg, Mexico, Netherlands, New Zealand, Norway, Poland, Portugal, Slovak Republic, Spain, Sweden, Switzerland, Turkey, UK, USA), and listed the advantages of e-government as follows:

- Improves efficiency in processing large quantities of data

- Improves services through better understanding of users' requirements, thus aiming for seamless online services.

- Helps achieve specific policy outcomes by enabling stakeholders to share information and ideas.

- Assists a government's economic policy objectives by promoting productivity gains inherent in ICT and e-commerce.

- Contributes to governments' reform by improving transparency, facilitating information sharing and highlighting internal inconsistencies.

- Helps in building trust between governments and their citizens, an essential factor in good governance by using internet-based strategies to involve citizens in the policy process, illustrating government transparency and accountability.

III. PURPOSE OF THE STUDY

The main purposes of this study are as follows:

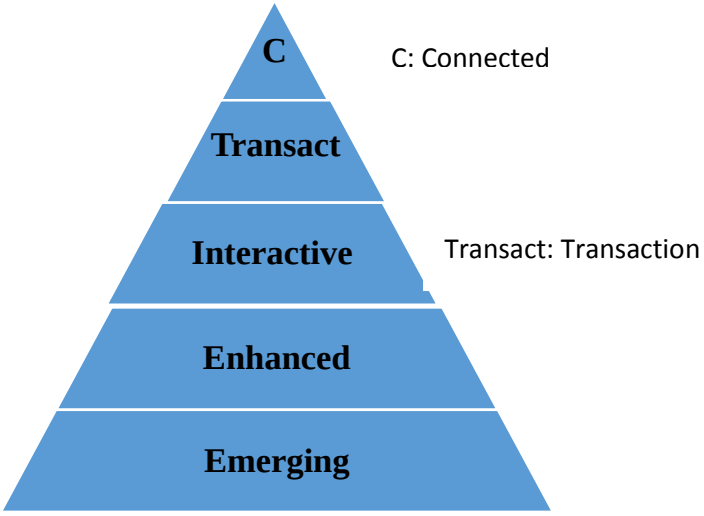
1. To understand the current level of e-government project implementation in Yemen.
2. To determine the critical success and failure factors of e-government project in Yemen as being perceived by top decision makers and project's team.
3. To come up with some recommendations that should bring into attention for the e-government project implementation in Sana'a city the capital of Yemen.

IV. E-GOVERNMENT MODELS

Several e-government models have been proposed by researchers and International organizations such as (UN 2008, Layne & Lee 2001) and some others. We briefly discuss these models in the following subsections:

A. UN e-government maturity model 2008: A Five stages model:

United Nations has identified stages of e-government from different perspective. UN developed a five stage e-government model which builds upon the sophistication of online presence. As a country migrates upwards through the various stages as shown in the figure, it is ranked higher in the web measure index (UN, 2008).



- UN E-government Maturity Model

Source: UN E-government Index 2008.

Stage I - Emerging: A government's online presence is mainly comprised of a web page and/or an official website; links to ministries or departments of education, health, social welfare, labor and finance may/may not exist.

Stage II - Enhanced: Governments provide more information on public policy and governance. They have created links to archived information that is easily accessible to citizens, as for instance, documents, forms, reports, laws and regulations, and newsletters.

Stage III - Interactive: Governments deliver online services such as downloadable forms for tax payments and applications for license renewals.

Stage IV - Transactional: Governments begin to transform themselves by introducing two-way interactions between 'citizen and government'. It includes options for paying taxes, applying for ID cards, birth certificates, passports and license renewals, as well as other similar G to C interactions, and allows the citizen to access these services online24/7. All transactions are conducted online.

Stage V - Connected: Governments transform themselves into a connected entity that responds to the needs of its citizens by developing an integrated back office infrastructure. This is the most sophisticated level of online e-government initiatives and is characterized by:

1. Horizontal connections (among government agencies)

2. Vertical connections (central and local government agencies)

3. Infrastructure connections (interoperability issues)

4. Connections between governments and citizens

5. Connections among stakeholders (government, private sector, academic institutions, NGOs and civil society)

B. Developing fully functionally E-government: A four stage model By Layne & Lee 2001:

Based on technical and managerial feasibility, Layne and Lee (2001) proposed a four-stage e-government model. These four stages are:

- *Stage 1- Cataloguing:* government online presence
- *Stage 2- Transaction:* citizens transact with government electronically.
- *Stage 3- Vertical integration:* Local government connected to different functions or service of the government.
- *Stage 4- Horizontal Integration:* Integration across different functions and services of government.

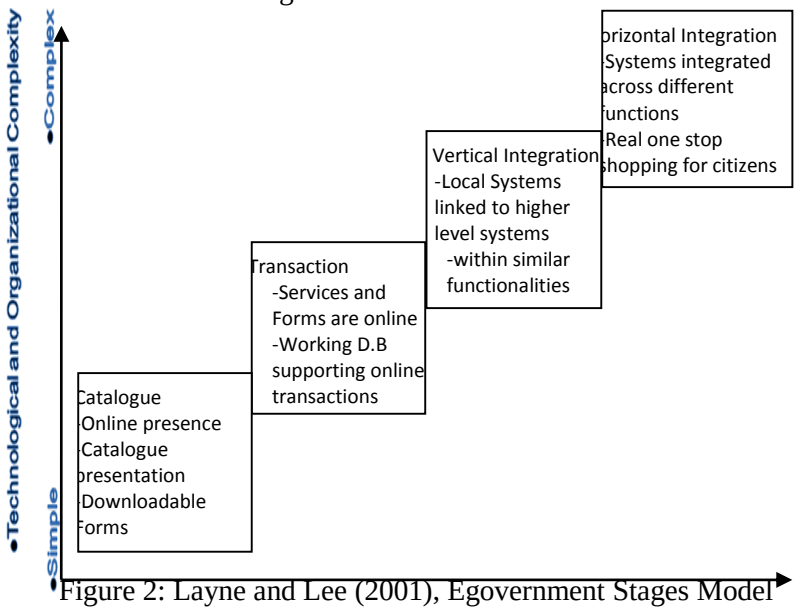


Figure 2: Layne and Lee (2001), Egovernment Stages Model

Source: Developing fully functionally E-government: A four stage model by Layne & Lee 2001.

According to the model, the first stage is the cataloguing of information on a Web site. This involves the initial efforts of government to be present online. In this stage, the information is not a transactional operation. Online presence

of the government saves its time and money because that will reduce the workload on frontline offices. The second stage is, Transaction, aims to extend the capability of catalogue stage and enables to perform a simple online transaction. The Vertical Integration stage transforms government services and integrating them at different levels. The horizontal Integration stage is the maturity stage with unified and seamless services.

The two models differ because they are developed from different perspectives. UN (2008) five-stage model focus on a web-based public service. Layne & Lee's (2001) four-stage model is developed based on a general or an integrated perspective combining technical, organizational, and managerial feasibility. However, the model proposed by Layne & Lee (2001), four-stage model misses out on the political participation component and does not address the possible changes in the way decisions are made in government. Yemen government is most likely deal with the different stages of e-government by following both of these models. (E-government project team Documentation, 2008).

V. ICT AND E-GOVERNMENT IN YEMEN

In this part the researcher is trying to present the current situation of Yemen's ICT and e-government. Therefore this part will be explained in five subsections as follows:

A. Yemen overview:

Yemen is one of the developing countries which located in the southern part of the Arabian Peninsula. It is bounded on the north by Saudi Arabia and by the Arab Sea. Oman lies in the west of the Republic of Yemen and the Red Sea lies in the west of the Republic of Yemen. The total area of the Republic of Yemen is about 555,000 square Kilometer, and the population of Yemen is 23 million (NIC, 2010). The official language of the country is Arabic and Islam is the official religion of the country. The Yemeni Rial (YR) is the official currency unit. The Republic of Yemen has three national independence days: September 26, 1962 when the king of North the Republic of Yemen, at that time, was overthrown and making the country a republic instead of a kingdom, November 30, 1967 when South the Republic of Yemen, at that time, had become independent from United Kingdom. The Unification Day on May 22, 1990, when the Republic of Yemen was established by the merger of South and North of Yemen.

Source: Yemen e-government portal at: www.yemen.gov.ye

B. ICT Sector in Yemen

Information and Communication Technology (ICT) is considered nowadays an efficient tool serving the progress and development of countries (ESCWA, 2007). Like other less developed countries, the republic of Yemen is seeking to develop this vital sector to carry on its development process (ESCWA, 2007). Yemen is in lowest level of ICT infrastructure in world, this level is characterized by the following: (a) low penetration rates of fixed and mobile telephone lines; (b) lack of an environment conducive to widespread use of telecommunication services by businesses and individuals; and (c) insufficient national bandwidth, inadequate backbone for voice and data telecommunication and insufficient number of Internet players in the market (ESCWA, 2009).

The following table 1- shows that Yemen country index in term of using computers and Internet, telephone usage.

Table 1: Telecommunication Infrastructure index and its components

Country	Index value	Estimated Internet users per 100 inhabitants	Main fixed telephone lines per 100 inhabitants	Mobile subscribers per 100 inhabitants	Personal computers per 100 inhabitants	Total fixed broadband per 100 inhabitants
Yemen	0.0298	1.44	4.48	13.76	2.77	0.00

Source: United Nations E-government Survey 2010.

According to (ESCWA, 2007), Yemen's effort towards building an Information Society can be summarized as follows:

- The National Information Center (NIC) was established during the second half of 1995.
- The ministry of Telecommunication and Information Technology (MTIT) in 2003, it was entrusted with IT management in addition to development and organization of communication policies. It has many campuses deal with the IT such as:
  1. Telecommunication and Information Technology city
  2. Electronic Library
  3. General Telecom Council
  4. Data communication
  5. Yemen Electronic Government Portal.
- The central statistics Department, affiliated to the Ministry of Planning and International Cooperation, adopted a national strategy on statistical work.

C. TelephoneCapacity

According to (ESCWA, 2007), Yemen is still among the countries that benefit little from communication services. The main challenges and obstacles are the country's difficult topography, scattered population groups and low density, weak infrastructure services and low level of distribution networks in the cities as well as in rural areas. The following table shows the majority indicators for Yemen telephone communications.

TABLE2: TELEPHONE COMMUNICATION INDICATORS (URBAN AND RURAL) DURING 2003-2005

INDICATOR/YEAR	2003	2004	2005	2006	2007

URBAN COMMUNICATION					
TOTAL CAPACITY OF COMMUNICATION NETWORK	1,505,192	1,564,349	1575,169	1,580,966	1,573,476
CAPACITY FITTED(LINE)	1,161,041	1,243,838	1,278,315	1,300,042	1,326,125
Population (People)	19,131,977	19,721,631	20,305,269	20,900,532	21,538,995
Telephone density (telephone/ per 100 inhabitances)	3.5	4.1	4.4	4.63	.74
Lines Vacancies (line)	476,157	445,702	376,930	331,714	304,137
Number of working telephone lines (line)	684,884	798,136	901,385	968,328	1,021,988
The increase in the working telephone lines	142,680	113,252	103,249	66,943	53,660
The number of main exchanges (number)	47	52	54	49	49
The number of sub exchanges (number)	208	222	223	229	229
Rural Communication**					
Rural Communication Capacities fitted (Line)	101,344	128,385	240,245	190,575	193,578
Lines operating (line)	62,593	96,027	124,921	156,190	168,718
Percentage of lines operating to equipped capacities (%)	62	75	52	82	87

Source: NATIONAL PROFILE OF THE INFORMATION SOCIETY IN YEMEN, ESCWA 2009.

D. D-Internet and mobile services in Yemen

According to (Al-Eryani, 2009), Internet services in Yemen started in 1996 with one single provider. Second provider has been started working in April

2002. Wireless access to network was lunched at the end of December 2006. Between2004-2006, Yemen improved its mobile communication services as the number of mobile telephone subscribers jumped from 1.483 million in 2004 to 3.201 million 2006 (ESCWA, 2007). The following table shows the majority indicators for Yemen internet and mobile services.

E. E-government development in Yemen

The Yemen government, like most other developing countries, is still trying to implement IT in government organization, and try establishing some projects as initial step to make e-government such as the National Program for Information Technology, that known as Yemen e-government project 2003. The following Table-4 shows that Yemen is still one of the lowest countries in terms of E-government development between 2008-and 2010.

Table 4: E-government developments in western Asia (including Yemen) between 2008 and 2010.

Country	**E-government development index value**		**World e-government development ranking**	
	2010	**2008**	**2010**	**2008**
Bahrain	0.7363	0.5723	13	42
Cyprus	0.5705	0.6019	42	35
United Arab Emirates	0.5349	0.6301	49	32
Kuwait	0.5290	0.5202	50	57
Jordan	0.5278	0.5480	51	20
Saudi Arabia	0.5142	0.4935	58	70
Qatar	0.4928	0.5314	62	53
Turkey	0.4780	0.4834	69	76

Oman	0.4576	0.4691	82	84
Azerbaijan	0.4571	0.4609	83	89
Lebanon	0.4388	0.4840	93	74
Georgia	0.4248	0.4598	100	90
Armenia	0.4025	0.4182	110	103
Syrian Arab Republic	0.3103	0.3614	133	119
Iraq	0.2996	0.2690	136	151
Yemen	0.2154	0.2142	164	**164**

E-government development for some Countries including Yemen between 2008 and 2010.

Source: United Nations E-Government Survey 2010.

VI. E-GOVERNMENT PROJECTS IN YEMEN

Information Communication Technology (ICT) is considered nowadays an efficient tool serving the progress and development of countries (ESCWA, 2007). The Yemen government, like most other developing countries, is trying to implement ICT in government organization, and try establishing some projects as initial step to make e-government system such as the National Program for Information Technology, that known as Yemen e-government project (MCIT, 2003), and the official government portal which has been announced in 2009. As the previous Table 4 compare Yemen to other countries around the world and show that Yemen is still one of the lowest countries in term of E-government development between "2008-2010".

*A. Old E-government project in Yemen (2003**):***

The first attempt to build e-government in Yemen was announced in 2003 and its websites was launched at the same year but with failure. However, the project was derailed in the absence of change-inducing environment (ESCWA, 2007). Moreover, most of the available information systems lack of technical compatibility and updated information (NIC, 2005). This first e-government project faced some problems which lead to the failure of the project; these problems can be summarized as follows:

1. Absence of clear vision and strategy for the purpose of the e-government project.
2. Absence of coordination and integration between the e-government project team and different stockholders from the government ministries and related agencies.
3. Absence of political leadership.
4. Conflicting priorities about the responsibilities among involved organizations toward the e-government project.
5. Carefulness from some organization to diffusion their information.
6. Lack of technical knowledge and common technical culture.
7. Lack of sense and social awareness of the importance of e-government system.
8. Lack of clear methodologies and models.
9. Lack of e-readiness among the government organizations employees including the top management and leaders.
10. The gap between the project vision and the real situation.

Because of the above factors, the project was failed. The failure of the first e-government project was seen from other top management point of view as adjournment not failure due to the government prior projects such as improving education sectors and health sectors.

*B. New E-government project in Yemen (2008)**:***

This project was established in September 2008.The main aim was to implement and completely design the Yemen official electronic portal. The first stage was to design the electronic information content of all government organizations, taking into account the daily updating. In this stage the citizen will benefit from e-government website in two different ways. First way is finding information from the portal via the internet. Second way, is inquiring about the government services via the emails and contacts available on the government portal.

From the new e-government plan, there will be projects working in parallel with the main project aim to develop the technical culture. Other projects will aim to diffusion the technical skills by doing many technical skills by doing many technical courses in many locations in Yemen such as summer camps (Al-eryani2009).

The second stage in the new e-government plan will be considering the concept of interaction between e-government and citizens in one direction (semi-interaction). The third stage of this plan will be focusing on e-payment. This stage will aim for full interaction in two directions.

However, the Yemen e-government portal at www.yemen.gov.ye is the first step toward the full implementation of the e-government project. The main aim of this government portal is to make a unified container for all government ministries and agencies in one site. This electronic portal must be the official

source of all the government information. This portal can be considered as the first track to the e-services and the first step toward the full e-government.
According to Al-Eryani (2009) this portal has many benefits such as: reduce cost of searching information, reduce time and effort of searching information, and prepare the society for the next step which toward the semi interaction with the government websites.
One of the main reasons behind the success of designing and launching the government portal with all ministries of Yemen government is the combination of selecting the project team which consists of ten members five of them are from the ministry of telecommunication and information technology for the technical and hosting issues and the other five are from the Information center of the Secretary of the cabinet at the Prime minister office for the administration and policy issues. This leads to get a good response from the different ministries and related agencies to upload their information as required from the project team to the official portal in a particular time and each of them take the responsibility of upgrading their information. This could not be happen without the interfere of the secretary of the cabinet in the Prime minister office which gives directions and instruction to the different ministries and agencies to be on time to do their responsibilities with the respect to the scheduling time specified by the project team.

VII. *CONCLUSION AND RECOMMENDATIONS*

The OECD (2003) defines e-government as "the use of information and communication technologies, and particularly the internet, as a tool to achieve better government". In this description, the internet is defined as a requirement and a possible medium for e-government. Thus, the first step, countries need to define their clear vision and strategy for e-government system. This broad vision of e-government should be shared with the all different stakeholders (government agencies, Citizens, Businesses, and others). From the e-government project's team information and phone call interviews and also researcher point of view as one member of the team, it has been clear that there are failure and success factors of implementing e-government system in the metropolitan city "Sana'a"-the capital city of the republic of Yemen. These factors can be summarized as follows:

Success factors:
1. Make the government employees e-ready. That means train them for the required ICT skills and knowledge as well as make them awareness of the benefits and importance of the e-government system.
2. Make the citizens e-ready. That means to reduce the gap between the citizens in the technology culture and skills.

Failure factors:
1. The ambiguous vision of the conceptual of the e-government.

2. The absence of the readiness and awareness among the employees and the citizens.
3. The digital divide between the communities.

Recommendations:

From the researcher point of view, some recommendation that Yemeni government should strongly consider towards e-government implementation which could be no technological factors as well. These recommendations can be summarized in two stages:

Stage1: E-readiness Infrastructure: This has to include the following:

1. Education: improve the quality of education and offer education for all.
2. Gender: reduce the gap between males and females and offer them equal education and chances.
3. Income: since the GDP per capita is around $1000 USD, the government should reduce the cost of computers, training, and internet access and make them available for the all layers of citizens with cheaper prices.
4. Rural and Urban: reduce the gap between the citizens in the rural and urban areas by offering them.

S*tage2: E-readiness Diffusion:* This has to include the following:

1. Diffusion of computers and internet culture
2. Clarify the thought that "Computer is only for specialist".
3. Clarify the misunderstanding of dangers of the internet websites.
4. Clarify the conceptual of e-government benefits and importance for citizens and employees.

Finally, it can be said that Yemen published its official portal in 2009 towards preparing for the whole e-government project implementation. The vision is now clear and the processes are somehow identifies and established. One of the main reasons behind the success of designing and launching the government portal with all ministries of Yemen government is the combination of selecting the project team which consists of ten members, five of them are from the ministry of telecommunication and information technology for the technical and hosting issues, and the other five are from the Information center of the Secretary of the cabinet at the Prime minister office for the administration and policy issues. This leads to get a good response from the different ministries and related agencies to upload their information as required from the project team to the official portal in a particular time, therefore each of them take the responsibility of upgrading their information. This could not be happen without the interfere of the secretary of the cabinet in the Prime minister office which gives directions and instruction to the different ministries and agencies to be on time to do their responsibilities with the respect to the scheduling time specified by the project team.

REFERENCES

[1] United Nations (2008). UN E-Government Survey 2008: From e-government to Connected Governance, Department of Economic and Social Affairs, Division for Public Administration and Development Management, available at: http://unpan1.un.org/intradoc/groups/public/documents/un/unpan028607.pdf

[1] Layne, K. and Lee, J.W. (2001).Developing fully functional E-government: A four stage Model, Government Information Quarterly 18(2), 122-136.

[2] Carter, L. and Belanger, F. (2005). The utilization of e-government services: citizen trust, innovation and acceptance factors, Information system J 15, 5-25.

[3] Valentina (2004). E – government for Developing Countrie; Opportinuties and Challenges. *The Electronic Journal on Information Systems in Developing Countries*, EJISDC (2004) 18, 1, pp. 1-24.

[4] Heeks, R. (2003) E-government for Development Projects Fail: How can Risks be Reduce? iGovernment working paper series, Paper no. 14. Available at http:// http://unpan1.un.org/intradoc/groups/public/documents/NISPAcee/UNPAN015488.pdf

[5] OECD,2003 "OECD e-Government Studies: The e-Government Imperative," OECD Publications, France, 2003

[6] ESCWA (Economic and Social Commission for western Asia), (2007) National Profile of the information society in Yemen, United Nation available at: http://www.escwa.un.org/wsis/reports/docs/Yemen_2005-E.pdf

[7] The Yemeni e-government portal project documentation, (2008, 2009).

[8] Dada, D. (2006). "The Failure of E-government in Developing Couturiers: A Literature Review". The Electronic Journal on Information Systems in Developing Countries, EJISDC (2006) 26, 7, 1-10.

Message authentication on vehicular cloud computing

Mousa Mohammed Gharsan
Department of Computer Network,
Faculty of Engineering and IT,
University of Alandalus
alqadaci2013@gmail.com

Fekri M. Abduljalil
Department of Computer Science,
Faculty of Educ., Arts, & Science,
University of Sana'a,
fmabduljalil@gmail.com

***Abstract*—** *Today vehicular cloud computing technology has been realized to provide better services on demand. A smart vehicle provides powerful resources for computing, storage, sensing, and data delivery. In addition, with the combination of smart vehicle with cloud computing allowing the user to access the hardware, data, and software in the vehicle, for example the user can use the vehicle's on board camera to deliver image on demand. In addition, because open commination is used in Vehicular cloud computing (VCC), there is a risk for important message exchange in VCC environment leads to misguiding users, message forgery, modification wrong information sharing. In this paper VC message authentication service framework (VC-MASF) is proposed to address the limitation of message integrity and message's source authentication in VCC applications.*

Keywords: Vehicular Cloud Computing, PKI, Key manager, Symmetric key, Message signature, Message authentication Code.

1. INTRODUCTION:

For the last few years, smarter vehicles, safer, and less stressful driving experiences have been realized. Currently, ordinary vehicles have devices such as GPS, radio transceiver, small-scale collision radars, cameras, on board computers and different types of sensing devices to alert the driver to all types of road safety conditions and mechanical malfunctions. Vehicles are becoming more sophisticated with on-board storage, powerful on-board computing capabilities, significant communication capabilities and less power limitations which are supported by hosts of sensors, actuators, on board radar and GPS [12].

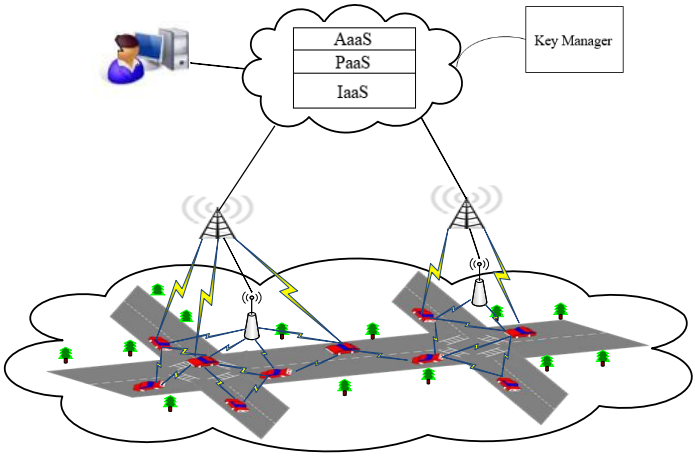


Figure 1: vehicular cloud architecture

The intelligent transportation system plays an important role in road safety, better utilization of traffic signals, traffic management etc…, which aim to improve driving conditions on road and road safety, this known as Vehicular ad hoc network(VANET) [2]**.** VANETs are considered important due to their huge potential and numerous applications. VANETs not only offer immense safety enhancements but also many commercial opportunities. Vehicular Internet: Security & Privacy Challenges and Opportunities. the Vehicular Environment, it is not possible for any user to get the dynamic information from anywhere, anytime. In VANET, the Messages shares between vehicles and with Road Side Units (RSU's) only when the vehicle is within the range of RSU,The vehicles communicate through the supported wireless medium i.e. Dedicated Short Range Communications (DSRC), Wireless Access in Vehicular Environments (WAVE) IEEE 802.11p standard [13].
The vehicles equipped with not only sensors to measure temperature, humidity, wind, etc., but also with cameras and other powerful resources. The large amount of unused storage space and computing power capabilities coupled with on-board provide a unique opportunity to utilize vehicles to be a part of cloud computing. This together with high speed (4G/LTE) connectivity can be a great advantage for a variety of applications[13].

The increasing vehicular cloud computing usage, the significance of security in this area is also increasing, providing security in a vehicular network is more difficult than in other network due to open communication, high mobility and wide range of vehicles. Therefore, there are various security that threaten the security of vehicular cloud computing, these threats can be classified into five parts, namely, confidentiality, authentication, on-repudiation, localization and verification data [12]. In this paper message

authentication is addressed, which confirm that the message came from an authorized user and has not been changed in transit. There are various security techniques used for message authentication, but some of the them are not compatible in vehicular cloud network due to their large computation power and time According to researcher's knowledge all the previous researches address the message authentication from vehicular ad hoc communication view, in this message authentication from vehicular cloud computing view is addressed by proposed a vehicular cloud message authentication service framework (VCMASF).

The remaining of the paper is organized as follows. section 1. reviews the background and related works. section 2. contains the proposed work. section 3. HMAC background. section 4. EC background. Section 5. concludes the paper. section 6. References

1. RELATED WORK

Chenxi Zhang [1] Proposed a message authentication scheme, named RAISE, which makes RSUs responsible for verifying the authenticity of messages sent from vehicles and for notifying the results back to vehicles. The proposed scheme provides an efficient identity-based batch signature verification scheme for vehicular communications, which enable vehicles to verify a batch of signatures once instead of one after another and efficiently increase vehicle's message verification speed.

Al-Sultan,others [2] Described that VANET is a subclass of mobile ad hoc network MANET, in which each vehicle acts as a node creating a network in the road with either another node or with a road side unit(RSU)located along the road. Each vehicle is supported with wireless sensing devices, which helps to establish communications between vehicles and RSU.This technology has been used in a range of applications such as predicting the correct route, controlling accidents and avoiding traffic jams and congestion.

Sherali Zeadally,others [3] Presented a lot of attentions from researcher to take care in vehicular network that still need further investigation and innovative solutions due to its interesting and promising applications in vehicular safety services, location based services and traffic congestion avoidance. The main focus must be on safe driving application wherever each vehicle periodically broadcasts messages with its current position, road formation and direction as well as speed.

Khana,others [4] Described that vehicular networks are the suitable networks that enable a communication among vehicles and vehicles to road side unit which can be used for intelligent transportation systems to obtain render safety, comfort and convenience on the road. However, because of lack of infrastructure and centralized administration, it becomes vulnerable to misbehaviors, proposed algorithm DMN-Detection of Malicious Nodes in VANETs improves DMV Algorithm in terms of effective selection of verifiers for detection of malicious nodes and hence improves the network performance.

Papadimitratos,others [5] Described that vehicular networks emerge as one of the most convincing and yet most challenging instantiations of the mobile ad hoc networking technology, security and privacy are critical factors and significant challenges to be met, outline security requirements for vehicular communication systems, provide models for the system and the communication, propose a set of design principles for future security and privacy solutions for vehicular communication systems.

Supriya, others [6] Introduced an overview of vehicular clouds, discussed the security challenges of VC, provide a directional security scheme to illustration an appropriate security architecture that handles several challenges of security in VCs

Abhale, others [7] Described the architecture of vehicular ad-hoc network, introduced the vision and architecture of mobile-vehicular cloud computing, discussed the security challenges in vehicular cloud.

Saurabh.others [8] Proposed an authentication scheme “Message Digest and Location based Authentication (MDLA)” to validate the mobile client and the cloud server participating in the mobile cloud computing, MDLA consists of three key phases, which are registration, authentication, and update. The operations of the scheme MDLA begins if and only if a mobile client is registered with the cloud service provider, used a protocol analyzer to validate the registration and authentication phases.
Mamun ,others [9] Proposed a GS scheme, based on pairing-based construction of Groth with additional properties. He presented a reliable and standard CPA-secure GS solution to a vehicular network application, suggested using of LM (Link Manager) that provides restricted privacy appropriate for a real time VANET environment and protects against DoS (Deny of Service) and Sybil attacks as well. He used batch verification which can significantly improve the performance of signature verification that makes the solution applicable for real life vehicular communication.

Sharma,others [10] Proposed a key authentication scheme for vehicular cloud environment. The proposed scheme for authenticating both the sender and receiver based on the ECC. The scheme also used one-way hash function and concatenation operation for secure communication. The proposed system aims to detect malicious vehicles in the network and maintain overall trust between the vehicles. From the obtained results, it is proved that proposed scheme better results as compared to other

Papadimitratos,others [11] Described security architecture for VC systems aiming at a solution that is both comprehensive and practical, discussed problem of identifying threats and models of adversarial behavior as well as security and privacy requirements that are relevant to the VC, He introduced a range of mechanisms to handle identity and credential management, and to secure communication while enhancing privacy.

2. PROPOSED WORK

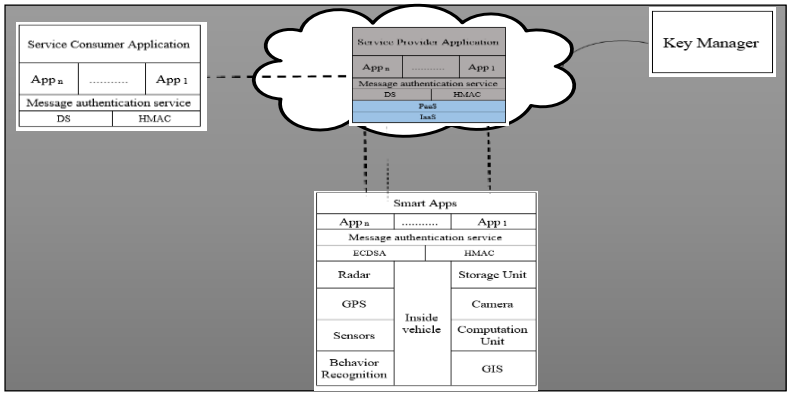


Figure 2: VC-MASF

In the proposed vehicular cloud message authentication service framework (VC_MASF), the message exchange between vehicular cloud applications is two type as the following:

1- Periodic message (Message sent periodically)

2- None periodic message (Message sent on demand)

In periodic message: the proposed message authentication scheme used is HMAC because signing/verifying every message every period of time (for example every 100ms) produce overhead and it is impossible.

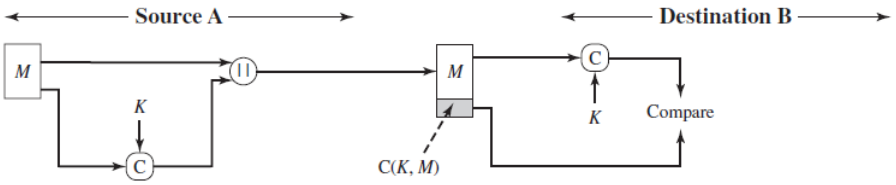


Figure 3: Message authentication Code,source=[16]

HMAC scheme process

1- Registration: In these phase, each vehicle needs to resister itself to a key manager, who is responsible for distributing a pair of private/public (SK/PK) key, and public key certificate to each vehicle. In addition, the key manager also plays a role as a security authority, who is capable of tracing their real identities of vehicles form their certificates. This role is important and necessary specially when criminal events happen. Thereby, during the registration, vehicles/drivers must provide the key manager with the real identity information. The registration process needs to be done before application communication. The Public/ private key generation can be achieved by adopting Elliptic curve private/public generation algorithm, because EC provide high level security with minimum key size comparable to other public key algorithm like RSA, gamal …etc[15]

Public/private generation and public key corresponding certificate distribution algorithm.

ACTIONS:

Step 1: Select an elliptic curve E defined over Zp. The number of points in E(Zp) should be divisible by a large prime n.

Step 2: Select a point P € E(Zp) of order n.

Step 3: Select a statistically unique and unpredictable integer d in the interval [1; n - 1].

Step 4: Compute Q = dP.

Step 5: public key is (E, P, n, Q), A's private key is d

Step 6: getting the public key certificate from certification authority C= $E(SK_{auth}, [T||ID||PK])$ where SK_{auth} is the private key used by

the authority and T is a timestamp

2- Manual authentication &Symmetric Key Establishment:
In these phase, vehicle (V) initiate a mutual authentication process with service provider (SP), SP verifying the vehicle signature and its corresponding public key certificate. A valid signature means that the vehicle is a legitimate user in VCC. The messages that have been signed by vehicle and service provider include secret credentials, which can be used to compute a shared key. Here, Diffie-Hellman key exchange secured with signature could be adopted to establish the shared symmetric key[14], the shard key used for a period of time (time stamp) and need to be reestablished.
$V \rightarrow SP : \{aP \mid Cert_V\}PKV.$

$SP \rightarrow V : ID \mid bP \mid \{ID \mid aP \mid bP\}SKSP.$

$V \rightarrow SP : \{ID \mid bP \mid aP\}SKV$

where aP and bP (a, b ∈ Z q, P is a generator of an addition group G) are random elements of the Diffie-Hellman key establishment protocol, and the shared session key between the Receiver and Sender is K ← abP . When receiving the first message from the Sender, the SP decrypts {aP | $Cert_V$}PK_{SP} (| as a concatenation operation) with its private key SK_{SP}, and then verifies the V's public key PK_V in the anonymous certificate $Cert_V$. Then, the SP sends ID | bP | {ID | aP | bP }SKsp to the Sender. The Sender verifies the signature {ID | aP | bP }

on ID | aP | bP . At last, the Sender sends back the signature {IDsp | bP | aP}SKv, and the receiver verify the signature.

3- HMAC computing and Verification: Having the shared key between sender and receiver. The sender can use it to compute a HMAC of a message, and then sends a tuple (ID, message, HMAC) to receiver (CS). Since the receiver has the shared key, it can verify the HMAC and then accept or reject the message.

In none periodic message: the proposed message authentication scheme used is DS, because signing/verifying only to message requesting on demand which doesn't produce overhead. The message digital signature can be achieved by adopting Elliptic curve digital signature algorithm (ECDSA), because ECDSA

provide high level security with minimum key size comparable to other public key algorithm like RSA, Elgamal …etc[15]

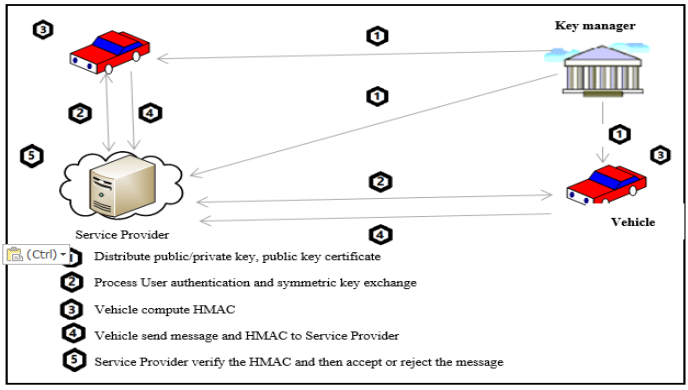


Figure 3: VC-MASF

DS scheme process [17]

1- Registration: In these phase, each vehicle needs to resister itself to a key manager, who is responsible for distributing a pair of private/public (SK/PK) key, and public key certificate to each vehicle. In addition, the key manager also plays a role as a security authority, who is capable of tracing their real identities of vehicles form their certificates. This role is important and necessary specially when criminal events happen. Thereby, during the registration, vehicles/drivers must provide the key manager with the real identity information. The registration process needs to be done before application communication. The Public/ private key generation can be achieved by adopting Elliptic curve private/public generation algorithm, because EC provide high level security with minimum key size comparable to other public key algorithm like RSA, Elgamal …etc[15]

Public / private key generation and public key corresponding certificate distribution.

ACTIONS:

Step 1: Select an elliptic curve E defined over Zp. The number of points in E(Zp) should be divisible by a large prime n.

Step 2: Select a point P € E(Zp) of order n.

Step 3: Select a statistically unique and unpredictable integer d in the interval [1; n - 1].

Step 4: Compute Q = dP.

Step 5: public key is (E, P, n, Q), A's private key is d

Step 6: getting the public key certificate from certification authority C= $E(SK_{auth}, [T||ID||PK])$ where SK_{auth} is the private key used by the authority and T is a timestamp

2- Signing message:

In these step, sender compute hash value of a message H(M), Signing the message using its private key: sign $sender_{PRV}$ H(M)||M, and then send M, signature, certificate to receiver.

Signing message algorithm

INPUT: Message m, private key d.

OUTPUT: Signature (r, s).

ACTIONS:

Step 1: Select a statistically unique and unpredictable integer k in the interval [1; n - 1].

Step 2: Compute kP = (x1, y1) and r = x1 mod n. (Here x1 is regarded as an integer, for example by conversion from its binary representation.) If r = 0, then go to step 1 (This is a security condition: if r = 0, then the signing equation s = k-1{h(m)+dr} mod n does not involve the private key d.)

Step 3: Compute k^{-1} mod n.

Step 4: Compute s = k^{-1}{h(m) + dr} mod n, where h is the Secure Hash Algorithm (SHA-1).

Step 5: If s = 0, then go to step 1. (If s = 0, then s^{-1} mod n does not exist; s^{-1} is required in step 3 of signature verification.)

Step 6: The signature for the message m is the pair of integers (r; s).

3- Signed message verification:

In these step, the receiver verifies the certificate of sender, if it is legitimate user in VCC, the receiver verifying the signed message using sender public key :very senderPK (H(M)), In addition the hash value compared with the hash, the hash value H is obtained after the manual verification to check whether message integrity is guaranteed.

Signing message verification

INPUT: Message m, signature (r, s);Public signing key Q(x,y) , the signature components r and s, and the base point G(x,y)

OUTPUT: Accept or reject signature.

ACTIONS:

Step 1: Verify that r and s are integers in the interval [1; n - 1].

Step 2:. Compute $w = s^{-1} \bmod n$ and h(m).

Step 3: Compute $u1 = h(m)w \bmod n$ and $u2 = rw \bmod n$.

Step 4: Compute $u1P + u2Q = (x0, y0)$ and $v = x0 \bmod n$.

Step 5: Accept the signature if and only if v = r

3. BACKGROUND IN HMAC

In this section a quick introduction to the theory of message authentication based on hash function (HMAC) is introduced . Chapter 12.2 of William stalling book [16] provides a background of HMAC. MAC (Message authentication code) is an alternative authentication technique involves the use of a secret key to generate a small fixed-size block of data, known as a cryptographic checksum or MAC, that is appended to the message. This technique assumes that two communicating parties, say A and B, share a common secret key. When A has a message to send to B, it calculates the MAC as a function of the message and the key: where
M= input message
C = MAC function
K= shared secret key
MAC = message authentication code
The message plus MAC are transmitted to the intended recipient. The recipient performs the same calculation on the received message, using the same secret key, to generate a new MAC. The received MAC is compared to the calculated MAC. If we assume that only the receiver and the sender know the identity of the secret key, and if the received MAC matches the calculated MAC, then
1. The receiver is assured that the message has not been altered. If an attacker alters the message but does not alter the MAC, then the receiver's calculation of the MAC will differ from the received MAC. Because the attacker is assumed not to know the secret key, the attacker cannot alter the MAC to correspond to the alterations in the message.

2. The receiver is assured that the message is from the alleged sender. Because no one else knows the secret key, no one else could prepare a message with a proper
MAC.
3. If the message includes a sequence number (such as is used with HDLC, X.25, and TCP), then the receiver can be assured of the proper sequence because an attacker cannot successfully alter the sequence number.
A MAC function is similar to encryption. One difference is that the MAC algorithm need not be reversible, as it must be for decryption.
A MAC, also known as a cryptographic checksum, is generated by a function C of the form T = MAC (K, M) Where M is a variable-length message, K is a secret key shared only by sender and receiver, and MAC (K, M) is the fixed-length authenticator, sometimes called a tag. The tag is appended to the message at the source at a time when the message is assumed or known to be correct. The receiver authenticates that message by precomputing the tag.
A MAC derived from a cryptographic hash function. The motivations for this interest are
1. Cryptographic hash functions such as MD5 and SHA generally execute faster in software than symmetric block ciphers such as DES.
2. Library code for cryptographic hash functions is widely available. In cryptography, a keyed-hash message authentication code (HMAC) is a specific type of message authentication code (MAC) involving a cryptographic hash function (hence the 'H') in combination with a secret cryptographic key. As with any MAC, it may be used to simultaneously verify both the data integrity and the authentication of a message. Any cryptographic hash function, such as MD5 or SHA-1, may be used in the calculation of an HMAC; the resulting MAC algorithm is termed HMAC-MD5 or HMAC-SHA1 accordingly. The cryptographic strength of the HMAC depends upon the cryptographic strength of the underlying hash function, the size of its hash output, and on the size and quality of the key.

HMAC Algorithm

Defining algorithm terms.

H = embedded hash function (e.g., MD5, SHA-1, RIPEMD-160)

IV = initial value input to hash function

M = message input to HMAC (including the padding specified in the embedded hash function)

Yi _ i th block of M, 0 > i < (L – 1)

L _ number of blocks in M

b _ number of bits in a block

n _ length of hash code produced by embedded hash function

K _ secret key; recommended length is n; if key length is greater than b,

the key is input to the hash function to produce an n-bit key

K+ _ K padded with zeros on the left so that the result is b bits in length

ipad _ 00110110 (36 in hexadecimal) repeated b/8 times

opad _ 01011100 (5C in hexadecimal) repeated b/8 times

Then HMAC can be expressed as
HMAC(K, M) = H[(K+ _ opad) || H[(K+ _ ipad) || M]]
Algorithm

1. Append zeros to the left end of K to create a –bit string K+ (e.g., if K is of length 160 bits and b = 512, K will be appended with 44 zeroes).

2. XOR (bitwise exclusive-OR) K+ with ipad to produce the -bit block Si .

3. Append to Si`.

4. Apply H to the stream generated in step 3.

5. XOR K+ with opad to produce the b-bit block So .

6. Append the hash result from step 4 to So.

7. Apply H to the stream generated in step 6 and output the result.

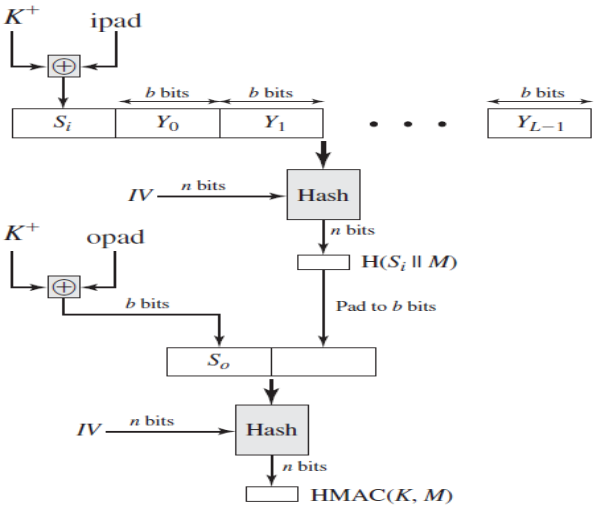


Figure 4: HMAC algorithm, source=[16]

4. BACKGROUND IN ELLIPTIC CURVES:

In this section a quick introduction to the theory of elliptic curves (EC) is introduced. Chapter 10.3 of William stalling book [16] provides a background of elliptic curves arithmetic and elliptic curve cryptography. For simplicity, I will restrict background to EC over Finite group. Most of the products and standards that use public-key cryptography for encryption and digital signatures use RSA. As we have seen, the key length for secure RSA use has increased over recent years, and this has put a heavier processing load on applications using RSA.This burden has ramifications, especially for electronic commerce sites that conduct large numbers of secure transactions. A competing system challenges RSA: elliptic curve cryptography (ECC). ECC is showing up in standardization efforts, including the IEEE P1363 Standard for Public-Key Cryptography. The principal attraction of ECC, compared to RSA, is that it appears to offer equal security for a far smaller key size, thereby reducing processing overhead. Elliptic curve cryptography makes use of curves in which the variables and coefficients are all restricted to elements of a finite field. Two families of elliptic curves are used applications: prime curves over Zp and binary curves over GF(2m) . For prime curve over Zp , we use a cubic equation in which the variables and coefficients all take on values in the set of integers from 0 through p - 1 and in which calculations are performed modulo p.

$y2 \bmod p = (x3 + ax + b) \bmod p$ (10.1)

For example, Equation (10.5) is satisfied for $a = 1$, $b = 1$, x = 9, y = 7, α= 1, p = 23

72 mod 23 = (93+ 9 + 1) mod 23

49 mod 23 = 739 mod 23

3 = 3

Now consider set consisting of all pairs of integers that satisfy Equation (10.5), together with a point at infinity *O*. The coefficients a and b and the variables *x* and *y* are all elements of Z*p*, example, let $p = 23$ and consider the elliptic curve $y2 = x3 + x + 1$. In this case, a = b = 1. For the set E23(1, 1), we are only interested in the integers in the quadrant from (0, 0) through (p - 1, p - 1) that satisfy the equation mod *p*. Figure 5 lists the points (other than *O*) that are part of E23(1, 1), Figure 6 plots the points of E23(1, 1); note that the points, with one exception, are symmetric about $y = 11.5$.

(0, 1)	(6, 4)	(12, 19)
(0, 22)	(6, 19)	(13, 7)
(1, 7)	(7, 11)	(13, 16)
(1, 16)	(7, 12)	(17, 3)
(3, 10)	(9, 7)	(17, 20)
(3, 13)	(9, 16)	(18, 3)
(4, 0)	(11, 3)	(18, 20)
(5, 4)	(11, 20)	(19, 5)
(5, 19)	(12, 4)	(19, 18)

Figure 5: Points on the Elliptic Curve E23 (1,1)

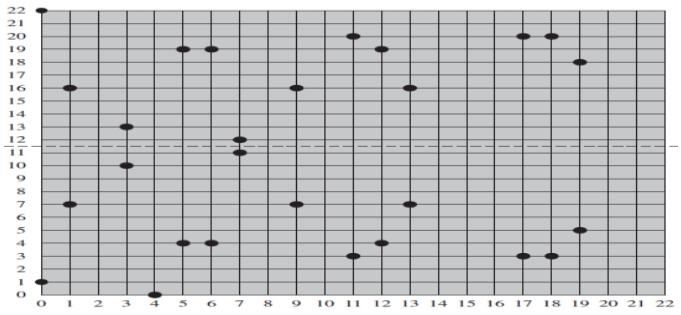


Figure 6:The Elliptic Curve E23(1, 1)

It can be shown that a finite abelian group can be defined based on the set Ep(a, b) provided that (x3 + ax + b) mod p has no repeated factors. This is equivalent to the condition

$(4a^3 + 27b^2) \bmod p \neq 0 \bmod p$ **(10.2).**

The rules for addition over E$p(a, b)$, correspond to the algebraic technique described for elliptic curves defined over real numbers. For all points P, Q _ E$p(a, b)$:

1- $P + O = P$
2- If $P = (xP, yP)$, then $P + (xP,-yP) = O$. The point $(xP,-yP)$ is the negative Of P , denoted as $-P$. For example $E23(1, 1)$, in , for $P = (13, 7)$, we have $-P = (13, -7)$.
   But -7 mod 23 = 16 . Therefore, $-P = (13,16)$, which is also In $E23(1, 1)$.
3- If $P = (xp, yp)$ and $Q = (xQ, yQ)$ with P Z $-Q$,then $R = P + Q = (xR, yR)$ is determined by the following rules:
   $xR = (l2 - xP - xQ) \bmod p$
   $yR = (l(xP - xR) - yP) \bmod p$
   where

$$\lambda = \begin{cases} \left(\frac{y_Q - y_P}{x_Q - x_P}\right) \bmod p & \text{if } P \neq Q \\ \left(\frac{3x_P^2 + a}{2y_P}\right) \bmod p & \text{if } P = Q \end{cases}$$

4- Multiplication is defined as repeated addition; for example, $4P = P + P + P + P$.
   For example, let $P = (3, 10)$ and $Q = (9, 7)$ in $E23(1, 1)$. Then

$$\lambda = \left(\frac{7 - 10}{9 - 3}\right) \bmod 23 = \left(\frac{-3}{6}\right) \bmod 23 = \left(\frac{-1}{2}\right) \bmod 23 = 11$$
$$x_R = (11^2 - 3 - 9) \bmod 23 = 109 \bmod 23 = 17$$
$$y_R = (11(3 - 17) - 10) \bmod 23 = -164 \bmod 23 = 20$$

So $P + Q = (17, 20)$. To find $2P$,

$$\lambda = \left(\frac{3(3^2) + 1}{2 \times 10}\right) \bmod 23 = \left(\frac{5}{20}\right) \bmod 23 = \left(\frac{1}{4}\right) \bmod 23 = 6$$

xR = $(6^2 - 3 - 3) \bmod 23 = 30 \bmod 23 = 7$

yR = (6(3 - 7) - 10) mod 23 = (-34) mod 23 = 12

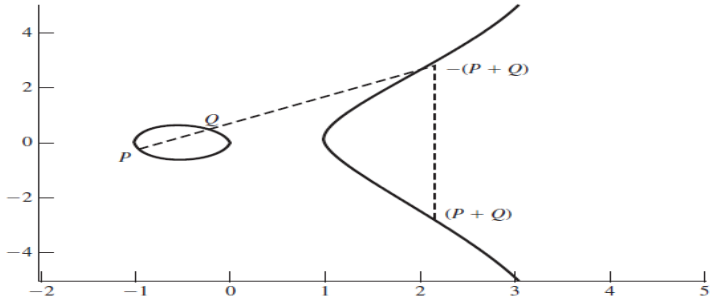


Figure 7: Geometric description of the addition of two points

Elliptic Curve Cryptography

The addition operation in ECC is the counterpart of modular multiplication in RSA, and multiple addition is the counterpart of modular exponentiation. To form a cryptographic system using elliptic curves, we need to find a "hard problem" corresponding to factoring the product of two primes or taking the discrete logarithm. Consider the equation $Q = kP$ where $Q\ P\ \_\ EP(a, b)$, and k 6 p . It is relatively easy to calculate Q given and P , but it is relatively hard to determine given Q and P .This is called the discrete logarithm problem for elliptic curves. We give an example taken from the Certicom Web site (www.certicom.com).Consider the group E23(9,17) . This is the group defined by the equation y^2 mod 23 = (x3 + 9x + 17) mod 23. What is the discrete logarithm k of Q = (4, 5) to the base P = (16, 5) ? The brute-force method is to compute multiples of P until is found *T*hus,

P = (16, 5); 2P = (20, 20); 3P = (14, 14); 4P = (19, 20); 5P = (13, 10);6P = 17, 32; 7P = 18, 72; 8P = (12, 17); 9P = (4, 5). Because 9P = (4, 5) = Q , the discrete logarithm to the base Q = (4, 5) is k = 9 . In a real application, k would be so large as to make the bruteforce approach infeasible. In the remainder of this section, we show two approaches to ECC that give the flavor of this technique.

Security of Elliptic Curve Cryptography

The security of ECC depends on how difficult it is to determine k given kP and P. This is referred to as the elliptic curve logarithm problem. The fastest known technique for taking the elliptic curve logarithm is known as the Pollard rho method. 10.3 compares various algorithms by showing comparable key sizes in terms of computational effort for cryptanalysis. As

can be seen, a considerably smaller key size can be used for ECC compared to RSA. Furthermore, for equal key lengths, the computational effort required for ECC and RSA is comparable [JURI97]. Thus, there is a computational advantage to using ECC with a shorter key length than a comparably secure RSA.

Symmetric Scheme (key size in bits)	ECC-Based Scheme (size of *n* in bits)	RSA/DSA (modulus size in bits)
56	112	512
80	160	1024
112	224	2048
128	256	3072
192	384	7680
256	512	15360

Figure 8:Comparable Key Sizes in Terms of Computational Effort for Cryptanalysis, Source =certicom

5. CONCLUSION

Increasing vehicular cloud computing usage, the significance of security in this area is also increasing, providing security in a vehicular network is more difficult than in other network due to open communication, high mobility and wide range of vehicles. Therefore, there are various security issues that threaten the security of vehicular cloud computing and the one of the challenges task is to provide message authentication in VCC. The network has always the possibility of attacks that leads to misguiding users, message forgery, modification, and extortion by man-in-the-middle attacks, wrong information sharing are possible due to the open communication nature. In the proposed work, the main aim is to provide a message authentication in vehicular cloud using the proposed VC-MASF.The proposed VC-MASF is also aim to detect malicious vehicles in the network and maintain trust between the vehicles and the cloud enviroment.

6. REFERENCES

[1] *Chenxi Zhang."On Achieving Secure Message Authentication for Vehicular Communications ,thesis requirement for the degree of Doctor of Philosophy in Electrical and Computer Engineering Waterloo, Ontario, Canada, 2010.*

[2] *Saif Al-Sultan, Moath M.Al-Doori,AliH.Al-Bayatti,HussienZedan." A comprehensive survey on vehicular Ad Hoc network, Journal of Network and Computer Applications, 2013*

[3] *Sherali Zeadally · Ray Hunt · Yuh-Shyan Chen ·Angela Irwin · Aamir Hassan."Vehicular ad hoc networks (VANETS): status, results, and challenges, Springer Science+Business Media, LLC 2010.*

[4] *Uzma Khana, Shikha Agrawal , Sanjay Silakari." Detection of Malicious Nodes (DMN) in Vehicular Ad-Hoc Networks, International Conference on Information and Communication Technologies (ICICT 2014).*

[5] *P. Papadimitratos, V. Gligor , J-P. Hubaux." Securing Vehicular Communications - Assumptions,Requirements, and Principles".*

[6]*Ms. RajbhojSupriya K, Dr..S.V.Gumaste."Security Challenges Addressed in Vehicular cloud computing",International Journal of Science, Engineering and Technology Research (IJSETR), Volume 4, Issue 6, June 2015.*

[7] *Ms. Ashwini Abhale, Mr. Sumit Khandelwal., Prof. Uma Nagaraj "Shifting VANET to Cloud - Survey.",International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 11, November 2013 ISSN: 2277 128X*

[8] *Saurabh Dey , Srinivas Sampalli , Qiang Ye "A Light-weight Authentication Scheme*
Based on Message Digest and Location for Cloud Computing".

[9] *Mohammad Saiful Islam Mamun , Atsuko Miyaji ,Hiroaki Takada "A multi-purpose Group Signature for Vehicular Network Security".,2014 International Conference on Network-Based Information Systems.*

[10] *Manish Kumar Sharma , Rasmeet S. Bali , Arvinder Kaur "Dyanimc Key based Authentication Scheme for Vehicular Cloud Computing".,2015 International Conference on Green Computing and Internet of Things (ICGCIoT).*

[11] *P. Papadimitratos, L. Buttyan, T. Holczer, E. Schoch, J. Freudiger, M. Raya*

Z. Ma, F. Kargl, A. Kung, J.-P. Hubaux ,"Secure Vehicular Communication Systems: Design and Architecture ",arXiv:0912.5391v1 [cs.CR] 30 Dec 2009.

[12] *Md Whaiduzzaman, MehdiSookhak, AbdullahGani, RajkumarBuyya," A survey on vehicular cloud computing ",* Journal ofNetworkandComputerApplications40(2014)325–344.

[13] *Kamran Zaidi, Muttukrishnan Rajarajan "* Vehicular Internet: Security & Privacy Challenges and Opportunities, Future Internet 2015, ISSN 1999-5903 ,www.mdpi.com/journal/futureinternet

[14] Douglas R. Stingson, "Cryptography: Theory and Practice, third edition," CRC Press, 2005.

[15] Sneha Charjan,D.H.Kulkarni "Quantum Key Distribution by Exploitation Public Key Cryptography (ECC) In Resource Constrained Devices, "International Journal of Emerging Engineering Research and Technology ,Volume 3,Issue 7,2015.

[16] William Stallinds "CRYPTOGRAPHY AND NETWORK SECURITY PRINCIPLES AND PRACTICE" ,FIFTH EDITION.

[17] N. Koblitz, A Course in Number Theory and Cryptography, 2nd edition, Springer-Verlag,1994.

[18] Don B. Johnson, Alfred J. Menezes, Elliptic Curve DSA (ECDSA): An Enhanced DSA

TIU: A Hybrid Model for the Adoption of Mobile Banking Application

Sultan Y. As-Sultan
Department of Computer Science and IT, Yemen Academy for Graduate Studies, Sana'a, Yemen
sultan2013sultan92@gmail.com

Ibrahim Ahmed Al-Baltah
Department of Information Technology Sana'a University, Sana'a, Yemen
albalta2020@gmail.com

Fua'ad Hassan Abdulrazzak
Department of Communication Technology and Networks Thamar University, Thamar, Yemen
Fuaad.abdulrazzak@gmail.com

Abstract

Many banks around the world are starting to offer banking services through mobile phones. Therefore, many studies have been proposed to help bankers to increase the number of customers. Despite the effort that has been founded in the state of the art, the adoption rate of mobile banking has not reached the expected amount yet. This study aims to fill this gap by proposing a hybrid model called TIU, which extends TAM/IDT/UTAUT models to analysis customers' attitudes toward using Mobile Banking Applications (MBA).The adapted models are considered the most accepted and well-known models in the literature. Consequently, the TIU constructs are adapted from those models to ensure its success.

Keywords: Mobile Banking Application ;Adoption; Intention to use; TAM;IDT;UTAUT ;Hybrid model TIU

1. Introduction

Over the past years, mobile banking has played an import role in our daily life [1]. Actually, the realization of mobile banking has gone through several stages to be able to deliver the targeted transactions. For example, in the past, customers had to go to the bank and deal with the bank employer perform their transaction, such as enquiry transfer money, pay bill etc [2]. This traditional process was tedious, which in turns lead to customer dissatisfaction. After a period of time, the development in information technology offered considerable effortless services, which is defined as web based application and this was perfect, because it allows the customers to connect and interact to their accounts by using web browsers of the desktop computer. However, these services were limited, because consumers had to present in their offices to perform the bank transactions. As a new and modern banking orientation, mobile banking has emerged with the aim of giving

customers many kinds of services in a very easy and fast manner [2-4]. Consequently, this could improve the ability of costumers to perform the required transactions and to manage their accounts at anytime and anywhere. Nowadays, mobile communications technologies offer vast additional value for consumers' banking transactions due to their always-on functionality and the option to bank virtually any time and anywhere[5] [2-4, 6-7]. For example, the Mobile Internet allows hundreds of millions of users to frequently use their mobile devices to access their business information and bank accounts and interact with friends, buy stuff online, search interesting places to visit on-the-go, ask for driving directions, and more. [8] .Therefore, it is worthy to point out that, the new information technology is becoming an important factor in the future development of financial services industry, and especially banking industry. Growing international trading and problems in transferring money have motivated researchers to propose a new structure E-banking and mobile banking. Most of banks are using the Internet as a new distribution channel [9]. Banking and financial systems constitute a critical service for our economy. However, the adaption of banking and financial systems involves many risks. Therefore, researchers have attempt to minimize the effect by proposing some strategies, models, indexes and frameworks [6-7, 10]. In general, most of the current studies have tried to solve and reduce system complexity in order to make systems as easy to use as possible , even with the presence of some challenges such as risks and uncertainty[7] .According to [11], there is still much work to be done in understanding and improving the trust that users place in a system's outputs, and the extent to which they understand the associated risks of decisions recommended by a system to help them to adopt to mobile banking transaction[11].

Accordingly, we can point out that, a greater understanding of the affection of the externals constructs and increased application of mobile banking would enhance and increase the intention of the customer toward this new kind of technology, which called Mobile Banking Application (MBA), and therefore, increase the adoption ratio to this technology [2-3, 6]. Due to the urgent need of MBA, many countries have tried to adapt and enhance MBAs, such as China, India, Finland ,Singapore , Thailand ,Malaysia , Saudi Arabia [6], [2-3], [12-13].

However, the aim of this study is to propose a hybrid model called TIU, which extends TAM/IDT/UTAUT models to analysis customers' attitudes toward using MBA. Furthermore, this study attempts to fill the gap in this area of research by understanding the necessary and important constructs that could influence user intention to use such new technique in general, and mobile banking application in more specific. In addition, it highlights some of the

inappropriate factors or constructs that could have no or little influence for the adoption of MBA.

The reminder or this paper is organized as follow. Section 2 presents the related works. Section 3 analyzes the research models. Section 4 introduces the proposed model. Section 5 presents the conclusion and highlights some future works. Section 6 presents study limitation.

2. Related works

In the literature, many studies have been conducted to examine the adoption of MBA, and to understand to what extent customers could adapt this technology.

A comprehensive review of the state of the art related to mobile banking adoption was conducted by [2], which covered the published studies in the period of time between year 2005 and 2014. One of the main finding of this review is that, many assumption of external constructs in previous studies have been refused or not supported or have a minimum affect toward mobile banking user intention. The reason for that is because the selection of constructs was randomly with poor understanding about.

From another perspective, [14] assumed that the lack of learning management systems (LMS) availability is negatively affects the perceived ease of use, but after performing the analysis based the respondents of the survey, this assumption was wrong and the result was not supported .

On the another hand, the main assumption of [7] was that tradition and usage barriers would have a strong negative effect or resistance to innovations toward mobile banking services, but the result showed that these two barriers have just a little effect for customer resistant to adapt mobile banking services/application. In addition, this study supposed that demographic factors like age and gender would play a main role for distinguish the dominate among users of mobile banking services by different rage of age and two types of gender male or female and after analyzed the respondents this suppose was also wrong [7] .

The conducted study in[12], firstly assumed that effort expectation is significantly affects individual intention to use mobile banking, but this assumption was not accurate because this factor have taken the minimum effect to individual intention to use mobile banking from the each factors that have been supposed in that study.

As it has been observed from the literature, three will-known and accepted models where proposed namely Technology Acceptance Model (TAM),

Innovation Diffusion Theory (IDT), and Unified Theory of Acceptance, and Use of Technology (UTAUT). These models share in common the aim of analyzing user acceptance, diffusion or intention a new emerged technology. As a matter of fact, most of the proposed studies that have been attempting to analyze user acceptance, diffusion or intention to use any technology (e.g. MBA) have adapted and extend one of the three mentioned model by adding another constructs. While some of the existing studies have tried to proposed hybrid models by adapting two of the above mentioned models.

For example, [15] proposed a model contains the most common construct which called trust and its independent factors, this construct trust have used to extent the two models IDT/UTAUT from different previous studies. Moreover, [15] adapted two fundamental constructs of TAM model which are perceived usefulness and perceived ease of use .These constructs and factors have been taken to increase user intention to use internet baking. The result of the study showed that these constructs have a significant influence toward internet banking in Gordian customer.

3. Models analyses

In this study we have summarized the fundamental constructs of the most three popular models TAM, IDT, and UTAUT which explicit in Table1.

Table 1. The fundamental constructs

N	TAM Fundamental Constructs	IDT Fundamental Constructs	UTAUT Fundamental Constructs
1	Perceived usefulness	Relative advantage	Performance expectancy
2	Perceived ease of use	Compatibility	Effort expectancy
3		Complexity	Social influence
4		Observability	Facilitating condition
5		Trialability	

In the literature, these three models have been extended to many versions by joining or adding some external construct or ,in some cases avoid some fundamental constructs depending to the field of the study to explain the intention to adopt mobile banking applications in different regions and countries [8, 16-18].As it has been mentioned in[2],and these three models and their extension were most used in south east Asia ,east Asia ,and Africa

for mobile banking adoption . Those studies have used them to explain and predict user behavior intention to use different kinds of payment technology services such as mobile payment, mobile banking ,and electronic payment[2] .

In our study we will do some analysis(comparing) and data mining for external constructs belonging to these three models TAM/IDT/UTAUT that have been used in previous studies[12, 19]to find and understand what is best and appropriate constructs that could improve user decision towards intention to use mobile banking ,and to build a perfect model which can help banking managers to understand factors that would increase their customers numbers and save customer loyalty to the banks by bringing appropriate services to the customer .After analyzing the three popular models TAM/IDT/UTAUT from the previous studies in mobile banking adoption we have found that each of these models was extent to its special and different external constructs with its times of frequent as shown in Table 2 [2].

Table 2. The extended constructs of each model

N	The Constructs Used to Extent TAM Model Only	The Construct Used to Extent IDT Model Only	Construct Used to Extent UTAUT Model Only
1	Customer awarenesss, 2 times	Banking needs , 1 time	Alertness , 1 time
2	Perceived elitism , 1 time	Benevolence , 1 time	mobile phone efficacy , 1 time
3	Individual mobility, 1 time	Fire reputation , 1 time	Personalization , 1 time
4	Innovation resistance , 1 time	Integrity , 1 time	service compatibility , 1 time
5	Situational normality , 1 time	perceived competence , 1 time	task characteristics , 1 time
6			Technology characteristics , 1 time

7			utility expectancy , 1 time
8			power distance , 1 time

After that ,We have also done a deep analysis and data mining to know what is the common external constructs that have been shared and added between each two three groups (TAM/IDT) , (TAM/UTAUT) and (IDT/UTAUT) of the three models (TAM/IDT/UTAUT) and their frequent times. The result is explicit in Table 3.

Table 3.The external constructs that found in to models

N	The Common Constructs in The Two models TAM/IDT	The Common Constructs in The Two Models TAM/UTAUT	The Common Constructs in The Two Models IDT/UTAUT
1	Ease of use ,27 times	Trust ,10 times	Trust ,10 times
2	perceived risk ,14 times	social infulance/subject norm ,15 times	Perceived risk ,14 times
3	self efficacy ,11 times	Perceived risk ,14 times	self efficacy ,11 times
4	compatibility(life style ,device), 10 times	self efficacy ,11 times	Credibility, 8 times
5	credibility ,8 times	facilitation condition ,9 times	demographic factors,8 times
6	demogrghic factors, 8 times	cost ,8 times	trialability 5 times

7	relative advantage, 8 times	Credibility, 8 times	Accessibility ,3 times
8	trialability ,5 times	demogrghic factors ,8 times	Convenience, 3 times
9	Perceived benefits , 3 times	structural assurance, 8 times	
10		Trialability, 5 times	

Then we have done additional analysis to know and understand the most common external constructs that have been shared between three models (TAM/IDT/UTAUT) and its frequents time .The result is demonstrated in Table 4 .

Table 4. The common constructs in the three models

N	The Common Constructs in the Three Models TAM/IDT/UTAUT
1	Trust ,10 times
2	Perceived risk ,14 times
3	self efficacy, 11 times
4	credibility ,8 times
5	demographic factors, 8 times
6	Trialability, 5 times

4. The proposed model TIU

In our study we propose to use more than one model to have more understanding and explanation for user intention to use mobile banking application(MBA) by combining some of fundamental constructs in these three popular models(TAM/IDT/UTAUT) .Moreover, we propose to add some effected external construct which have been tested and supported in the previous studies and which have been used and evaluated by more than 8

studies because these constructs could really play a main role for understanding customer decision and intention to use (MBA)[2].

Therefore, our proposed model adapts two fundamental factors or constructs from TAM namely perceived usefulness and perceived ease of use as in Table 1, and two external constructs perceived risk and self efficacy that were common in two models TAM/IDT as in Table 2. In addition, we propose to add perceived social influence construct that was common in two models TAM/ UTAUT as in Table 2 .Also, the demographic construct is adapted in the proposed model as a moderating construct to category the respondents to different groups. It worth nothing to point out that, the demographic construct has been considered and adapted in all groups of models (TAM/IDT), (TAM/UTAUT) ,(IDT/UTAUT) ,(TAM/IDT/UTAUT) as shown in Table 3 and Table 4. Our hybrid model TIU is illustrated in Figure.1.

5. Conclusion

This study has analyzed the common constructs that have been used to extend the most popular models namely TAM, IDT, and UTAUT. Then the study proposes a hybrid model called TIU which extends TAM/IDT/UTAUT models to analysis customers' attitudes toward using Mobile Banking Applications (MBA). Besides, the analysis that performed in this study will give researchers a deep understanding that certainly help them to select the most appropriate constructs required in their studies.

As a future work we are planning to implement our proposed model, and then design an empirical experiment to evaluate it.

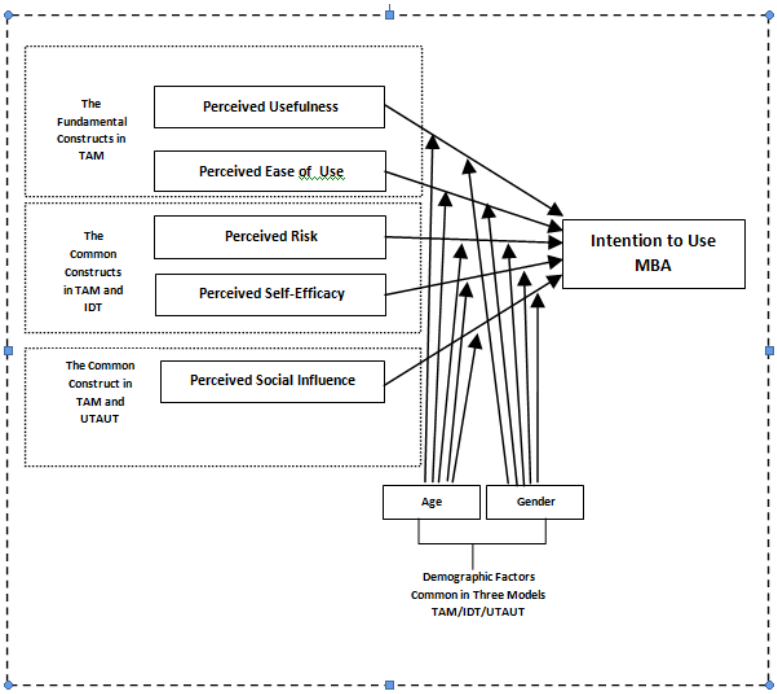


Figure. 1. The hybrid model TIU

6. Study limitation

This study have focus only on the most frequent constructs that have been used in previous studies, which are related to mobile banking adoption .We think rare external constructs which were used in previous works such as perceived enjoyment or Innovation resistance may have influence to customer intention to use(MBA).This study recommend next researchers to search for such constructs that could influence customer intention even if that constructs have less frequent or not used in such studies.

7. The References

[1] B. Cao and Q.-m. Fan, "The Infrastructure and Security Management of Mobile Banking System," in 2010 International Conference on E-Product E-Service and E-Entertainment, 2010.

[2] A. A. Shaikh and H. Karjaluoto, "Mobile banking adoption: A literature review," Telematics and Informatics, vol. 32, pp. 129-142, 2015.

[3] T. Wai-Ming and L. S. L. Lai, "Mobile Banking and Payment in China," IT Professional, vol. 16, pp. 22-27, 2014.

[4] X. Suping, "Adoption of mobile banking model based on perceived value and trust," in Information Management, Innovation Management and Industrial Engineering (ICIII), 2013 6th International Conference on, 2013, pp. 632-635.

[5] T. Laukkanen, et al., "Segmenting bank customers by resistance to mobile banking," International Journal of Mobile Communications, vol. 6, pp. 309-320, 2008.

[6] L. Mohan, et al., "Mobile App Usability Index (MAUI) for improving mobile banking adoption," in Evaluation of Novel Approaches to Software Engineering (ENASE), 2015 International Conference on, 2015, pp. 313-320.

[7] T. Laukkanen, "How Uncertainty Avoidance Affects Innovation Resistance in Mobile Banking: The Moderating Role of Age and Gender," in System Sciences (HICSS), 2015 48th Hawaii International Conference on, 2015, pp. 3601-3610.

[8] M. Sarwat and M. F. Mokbel, "Mobi Social (Mobile and Social) Data Management: A Tutorial," in Cloud Engineering (IC2E), 2015 IEEE International Conference on, 2015, pp. 4-4.

[9] M. R. Nami, "E-Banking: Issues and Challenges," in Software Engineering, Artificial Intelligences, Networking and Parallel/Distributed Computing, 2009. SNPD'09. 10th ACIS International Conference on, 2009, pp. 263-266.

[10] G. Guemkam, et al., "Financial Critical Infrastructure: A MAS Trusted Architecture for Alert Detection and Authenticated Transactions," in Network and Information Systems Security (SAR-SSI), 2011 Conference on, 2011, pp. 1-8.

[11] P. Townend, et al., "Risk assessment and trust in services computing: applications and experience," in Services Computing (SCC), 2013 IEEE International Conference on, 2013, pp. 392-399.

[12] C.-S. Yu, "Factors affecting individuals to adopt mobile banking: Empirical evidence from the UTAUT model," Journal of Electronic Commerce Research, vol. 13, p. 104, 2012.

[13] B. K. Jeong and T. E. Yoon, "An empirical investigation on consumer acceptance of mobile banking services," Business and Management Research, vol. 2, p. p31, 2013.

[14] S. Alharbi and S. Drew, "Using the technology acceptance model in understanding academics' behavioural intention to use learning management systems," International Journal of Advanced Computer Science and Applications (IJACSA), vol. 5, 2014.

[15] M. A. Al-Sharafi, et al., "The Effect Of Security And Privacy Perceptions On Customers' Trust To Accept Internet Banking Services: An Extension Of TAM," 2016.

[16] Z. Guo, et al., "Mechanism Design for Near Real-Time Retail Payment and Settlement Systems," in System Sciences (HICSS), 2015 48th Hawaii International Conference on, 2015, pp. 4824-4833.

[17] J. Hewitt and J. Pham, "Quantitative Risk Assessment enhances customer trust in aviation," in Reliability and Maintainability Symposium (RAMS), 2014 Annual, 2014, pp. 1-6.

[18] X. Xia, et al., "Web viral marketing based construction strategy of customer trust Network," in Computational and Information Sciences (ICCIS), 2012 Fourth International Conference on, 2012, pp. 425-429.

[19] I. Govender and W. Sihlali, "A study of mobile banking adoption among university students using an extended TAM," Mediterranean Journal of Social Sciences, vol. 5, p. 451, 2014.

Towards a Data Fusion Framework for Heterogeneous Sensors in Semantic Web of Things

Ghilan Mohammed Al-Gomai
Department of Computer Networks Yemen Academy for Graduate Studies, Sana'a, Yemen
ghilan20@gmail.com

Ibrahim Ahmed Al-Baltah
Department of Information Technology Sana'a University, Sana'a, Yemen
albalta2020@gmail.com

Fua'ad Hassan Abdulrazzak
Department of Communication Technology and Networks Thamar University, Thamar, Yemen
Fuaad.abdulrazzak@gmail.com

Abstract

Over the last few years, the number of embedded sensors have been increased in smart devices (e.g., mobile phones or smart watches) supported by several vendors. Furthermore, a variety of data will be generated by heterogeneous sensors. Traditional multi-sensor data fusion can deal with the same type of data effectively. However, heterogeneous sources with different types of data and measurement units must be compatible to provide a comprehensive, timely and reliable feedback related to a specific phenomenon. To address this issue, we propose a framework to detect and solve semantic conflict of data measure unites that might arise between the heterogeneous sensors.

Keywords: Web of things; Internet of things; Sementic web; Web server; Data integration, Data fusion Sensor data; Measurment units

1. Introduction

The tremendous and rapid advance of the internet has been shifted to the internet of things IoT. Consequently, the Web also has benefited from this advanced infrastructure and has moved forward to fully considerate the communication between heterogeneous devices at application-level[1], and this known as Web of things (WoT).

Over the last few years, the number of sensors are growing very fast, which are normally embedded in smart devices (e.g., mobile phones or smart watches). According to Cisco's predictions, there will be more than 50 billions of devices connected to the Internet by 2020 [2]. Typically, these sensors are provided from different vendors, which are in fact, designed and built according the vendors polices. As a natural results, these sensors will be

built based on different data models, which in turns, leads to the incompatibility between the massive quantity of the collected data from the heterogeneous sensors. Therefore, some researchers have attempted to mitigate this problem.

As a matter of fact, to facilitate meaningful data exchange between machines, the semantic web promotes common data formats through ontology utilization. Some of the current approaches have attempted to achieve data integration from heterogeneous sensors by employing the idea of semantic annotation to semantically annotate the sensor data according to ontology concepts that applying linked data principles [3]. Most of the existing studies [4-11] proposed architectures or frameworks to facilitate the integration and seamless access to heterogeneous sensors data with the semantic web technologies. They annotated the resources with semantic metadata to ensure the same interpretation of their capabilities and measurement units by different machines, and to support decision making and reasoning mechanisms by applications.

In order to analyses and understand a given phenomenon extensively such as (monitoring the environment, monitoring the ambient temperature in hall) collecting data from a single sensor may not create a significant value. Therefore, data fusion from set of heterogeneous sensors is critical to generate an accurate value. During the fusion process on data sensors , any possible conflict (e.g., data types: int, boolean / measurement units: Celsius, Fahrenheit) should be detected and resolved in order to ensure successful data fusion. According to the relations among sources, as discussed in[12], information fusion could be classified to complementary, redundant, and cooperative. Complementary means when information provided by sources represents different portions of a broader scene, in this case, information fusion would be applied to obtain a broad piece of information. While redundant means if two or more independent sources provide the same piece of information, these pieces can be fused to increase the reliability and confidence. Whereas cooperative refers to the situation where two independent sources are cooperative and the provided information would be integrated to produce new information (usually more complex than the original data), which, from an application perspective, better represents reality.

The aim of this paper, is to propose a framework to detect and solve semantic conflict of data measure unites that might arise between the heterogeneous data sensors. The process of this framework starts by matching the measurement units, and then data fusion process can take place. The remainder of this paper is organized as follows: The next section presents the motivation scenario. Section 3 reviews and summarizes related works of this study. Section 4 explains the proposed framework. Finally, Section 5 concludes this study.

2. Motivation scenario

Yemen Academy for Graduate Studies have smart hall (i.e., a large room for meetings, conferences, or other events) contains three distinct temperature sensors, which return the temperature values with different measurement units (e.g., Celsius, Fahrenheit). The user have interface that could, for instance, be a smartphone or tablet computer, or a Web application that allows the user to configure the desired temperature. In order to cover a large room and get the ambient temperature with increase the associated confidence and the reality, this involves fusing the collected temperature data from the sensors that allocated in the smart hall. Before the data fusion process take place, it is important to verify whether the sensors data measurement units are not different; otherwise a conversion process should be performed to ensure the compatibility of the measurement units of the collected data from the sensors.

3. Related work

Semantic Web of things is a new topic, which still in its infancy. Therefore, researchers have contributed with excellent efforts to SWoT and attempting to reach its maturity level. Thus, this section reviews the state of the art related to the topic of this study.

One of the studies proposed a platform called Sense2Web, in which the main attributes of sensors are captured and represented in RDF to create the necessary links to other resources [4]. In the same time, Sense2Web submits the sensor descriptions manually. These descriptions are stored in XML format, and then transferred into RDF. However, the focus of this platform was only on publishing the important sensor features, with no consideration to sensor readings [4].

SEMSENSE is an architecture that aims at data collection from single source and publishes the collected data to the Web according to Linked Open Data (LOD) principles. In SEMSENSE, mappings are done manually to enrich sensor data that stored in MySQL database with Semantic Sensor Network (SSN) ontology[13] concepts through D2RQ language [5]. From another perspective, Linked Sensor Middleware (LSM) is a cloud platform that provides real time data collection, annotation and publishing. It also support unified linked stream data query [6]. In [7], the authors proposed a system that helps users control their smart environment and produces a machine-readable description. They exploit reasoning mechanism that accepts different types of queries. The things require to be annotated with semantic descriptions and exposed as RESTful resources. Another architecture is proposed in [8], and

its aim was to merge heterogeneous sensor networks, convert measurements into semantic data and reason on them. However, this architecture mainly focuses on information coming from sensors and integration of different ontologies and protocols based on semantic reasoning. F. Wang, et al. [9], discussed multi-source heterogeneous information fusion in the IoT, they compared the features of data and information in the IoT to an existing wireless sensor network and the Internet. Furthermore, they proposed architecture for information fusion in the IoT that focus on computation of structural IoT data to improve data quality, which helps in making appropriate decisions. On the other hand, SIGHTED is a framework for collecting and publishing sensor data of multiple sources. It extends SEMSENSE architecture in order to deal with multiple heterogeneous sensor data sources and semantically enriches the sensor data that stored in MySQL database through the use of D2RQ mapping file to ontologies concepts [10]. Another IoT middleware called CASCOM [11], which is a configuration model acting as a mediator between the hardware layer and the application layer to simplify data retrieval process from sensors for non-technical user. In order to perform these bindings, middleware solutions need to be configured depending on the context information and user requirements. A sophisticate graphical user interface it will be used to get user requirements.

However, from examining the state of the art, the problem of ensuring data compatibility especially data measurement units that collected from heterogeneous sensors is still need further research. Therefore, this study proposes a framework that attempts to solve this problem and to fills the gap that is found in the literature.

4. The proposed framework

The proposed framework is amid to detect and solve semantic conflict of data measure unites that might arise between the heterogeneous sensors. To tackle this problem, this framework starts by detecting and solving any possible conflicts of measurement units, and then performs data fusion.

The framework consists of four main components as illustrated in Fig. 1, and these components are discussed in the following subsection.

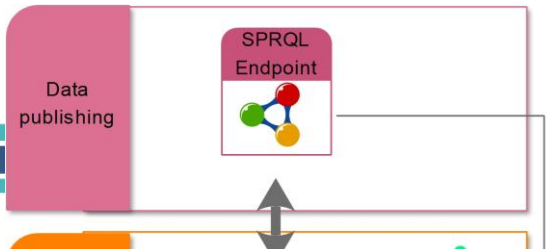


Fig. 1. The proposed framework components

4.1. Raw data collection

The predominant consideration of this component is to provide an interface to various IoT data sources. It is responsible for collecting data from the connected IoT devices. Data is collected by WoT Server that is based on Express framework[2] . Some devices can't directly offer a web API because the device might not support HTTP directly. The WoT gateway is working as a proxy for the Thing by offering a web API in the Thing's name.

4.2. Data storage

This component is responsible for storing the data. According to [14], the most suitable and common solution for data storage could be provided by

[2] http://expressjs.com/

database management systems DBMS. It provides an easy interaction between user and the stored data, also it allows user to query and analyze large amount of data with little effort.

Thus, in the proposed framework, a centralized My SQL database is used for storing the collected raw data.

4.3. Semantic annotation and mapping

This component performs the process of transforming the raw data that has been collected from heterogeneous sources into RDF format through the use of common ontologies. To perform this task, this study suggests reusing some specific existing ontologies, namely, Semantic Sensor Network (SSN) ontology for sensor domain, Quantity Kinds and Units (QU) ontology for quantities and units, and the DOLCE Ultra Lite (DUL) ontology[15] for location. D2RQ platform is adapted in this framework to construct the mapping between the relational database and the RDF triple store based on the unified data model.

4.4. Data publishing

The responsibility of this component is to publish the integrated data via standard interfaces and to make it accessible. The published data will be in RDF format, which requires using SPARQL as a query language. It is worthy to point out that, SPARQL is a W3C recommendation as a standard query language for RDF.

5. Conclusion

The Internet of Things connect objects or things to Internet, and integrate with semantic web technology through Semantic Web of Things (SWoT).This study proposes a data fusion framework to detect and solve semantic conflict of data measure unites that might arise between the heterogeneous data sensors.

The framework started with collecting data from the connected IoT devices and stores this data in a centralized My SQL database. Then, construct the mapping between the relational database and the RDF triple store through D2RQ platform. Finally, the framework publishes the data as RDF format using SPARQL query language.

In the future work, we plan to implement the proposed framework as a proof-of-concept to demonstrate its applicability. Furthermore, we plan to evaluate the proposed framework by collecting data sensors from multi heterogeneous sensors to ensure its ability for detecting and solving the conflicts.

6. References

[1] D. Guinard, V. Trifa, F. Mattern, and E. Wilde, "From the internet of things to the web of things: Resource-oriented architecture and best practices," *Architecting the Internet of Things*, pp. 97-129: Springer, 2011.

[2] D. Evans, "CISCO Internet Business Solutions Group (IBSG)," 2015.

[3] C. Bizer, T. Heath, and T. Berners-Lee, "Linked data-the story so far," *Semantic Services, Interoperability and Web Applications: Emerging Concepts*, pp. 205-227, 2009.

[4] S. De, T. Elsaleh, P. Barnaghi, and S. Meissner, "An internet of things platform for real-world and digital objects," *Scalable Computing: Practice and Experience*, vol. 13, no. 1, pp. 45-58, 2012.

[5] A. Moraru, D. Mladenic, M. Vucnik, M. Porcius, C. Fortuna, and M. Mohorcic, "Exposing real world information for the web of things." p. 6.

[6] D. Le-Phuoc, H. N. M. Quoc, J. X. Parreira, and M. Hauswirth, "The linked sensor middleware–connecting the real world and the semantic web," *Proceedings of the Semantic Web Challenge*, vol. 152, 2011.

[7] S. Mayer, and G. Basler, "Semantic metadata to support device interaction in smart environments." pp. 1505-1514.

[8] A. Gyrard, "An architecture to aggregate heterogeneous and semantic sensed data." pp. 697-701.

[9] F. Wang, L. Hu, J. Zhou, J. Hu, and K. Zhao, "A semantics-based approach to multi-source heterogeneous information fusion in the internet of things," *Soft Computing*, pp. 1-9, 2015.

[10] A. M. Nagib, and H. S. Hamza, "SIGHTED: A Framework for Semantic Integration of Heterogeneous Sensor Data on the Internet of Things," *Procedia Computer Science*, vol. 83, pp. 529-536, 2016.

[11] C. Perera, and A. V. Vasilakos, “A knowledge-based resource discovery for Internet of Things,” *Knowledge-Based Systems,* vol. 109, pp. 122-136, 2016.
[12] E. F. Nakamura, A. A. Loureiro, and A. C. Frery, “Information fusion for wireless sensor networks: Methods, models, and classifications,” *ACM Computing Surveys (CSUR),* vol. 39, no. 3, pp. 9, 2007.
[13] M. Compton, P. Barnaghi, L. Bermudez, R. GarcíA-Castro, O. Corcho, S. Cox, J. Graybeal, M. Hauswirth, C. Henson, and A. Herzog, “The SSN ontology of the W3C semantic sensor network incubator group,” *Web Semantics: Science, Services and Agents on the World Wide Web,* vol. 17, pp. 25-32, 2012.
[14] M. Balazinska, A. Deshpande, M. J. Franklin, P. B. Gibbons, J. Gray, M. Hansen, M. Liebhold, S. Nath, A. Szalay, and V. Tao, “Data management in the worldwide sensor web,” *IEEE Pervasive Computing,* vol. 6, no. 2, pp. 30-40, 2007.
[15] A. Gangemi, N. Guarino, C. Masolo, A. Oltramari, and L. Schneider, "Sweetening ontologies with DOLCE." pp. 166-181.

Some defects in tools used by the main components in firewall

بعض العيوب في الوسائل المستخدمة في المكونات الاساسية للجدار الناري

Dr. Saleh N. Abdullah
Associate Prof., Dept. of Comp. Science, Khawlan College, Sana'a University, Yemen.
Email:saleh.alasali97@yahoo.com

الملخص

الجدار الناري من المصطلحات المهمة في علم الحاسوب وهو يتحكم في تدقيق المعلومات التي تمر بين اجهزة الحواسيب في شبكة داخلية LAN والإنترنت، على غرار جهاز التوجيه، وفكرته تشبه فكرة نقطة التفتيش التي تسمح بمرور أناس، وتمنع مرور آخرين، بناءً على تعليمات مسبقة.

إن أي جدار ناري يتألف من مجموعة من المكونات، كل مكون له مهام أو مهمة خاصة، وتتفاعل تلك المكونات لتحقق الهدف المطلوب من الجدار الناري.

وكل مكون له مجموعة من الميزات ومجموعة من العيوب، ولا يمكن أن يدعي أي شخص بأنه قد وصل إلى الحماية المطلقة.

وعند معرفة بعض المميزات وبعض العيوب لبعض مكونات الجدار يمكن تقديم المشورة المناسبة دون مبالغة في زيادة لأمنه وبالتالي زيادة التكلفة، أو إهمال بعض العيوب لتوفير بعض التكلفة.

إن معظم الثغرات والاختلالات تأتي من استخدام بعض الوسائل غير المناسبة للتطبيقات الآمنة، والتي قد تكون مناسبة نظرياً ولكن في الواقع تحتاج إلى وسائل غير متاحة أو غير مدركة.

1 المقدمة

مع تطور علم الاتصالات اصبح التواصل من بعد له أهمة قصوى في انجاز بعض المهام من مكان الإقامة في أي مكان في العالم الذ اصبح كالقرية الصغيرة، ولذلك اصبح لجدار الناري من المتطلبات المهمة، حيث انه يتحكم في تدقيق المعلومات التي تمر بين اجهزة الحواسيب المتواجدة في شبكة داخلية LAN والإنترنت، ولذي يعمل كنقطة التفتيش التي تسمح بمرور أناس، وتمنع مرور آخرين، بناءً على تعليمات ضمن سياسة مسبقة.

إن أي جدار ناري يتألف من مجموعة من الأنظمة، كل نظام له وسائله ومهامه الخاصة، وتتفاعل تلك المكونات لتحقق الهدف المطلوب من الجدار الناري.

إن معظم الثغرات والاختلالات تأتي من استخدام بعض الوسائل غير المناسبة للتطبيقات الآمنة، والتي قد تكون مناسبة نظرياً ولكن في الواقع تحتاج إلى وسائل غير متاحة أو غير مدركة.

1.1 مشكلة البحث:

وجود بعض الصعوبات تتعلق بالوسائل المستخدمة في بعض المكونات الأساسية لأنواع الجدار الناري ومنها استخدام public key في التحقق من الهوية.

1.2 أهداف البحث:

تقديم بعض الحلول للعيوب في الوسائل المستخدمة في بعض المكونات الأساسية للجدار الناري.

في هذا البحث سوف يتم توضيح بعض العيوب في استخدام public key المستخدم في بعض المكونات الأساسية للجدار الناري، وسوف يتم تقديم مقترح لحل بعض العيوب في استخدام public key المستخدم في الجدار الناري.

المكونات الأساسية للجدار الناري2

يتكون الجدار الناري من مجموعة من الأنظمة، تسمى المكونات الاساسية، كل مكون ينجز مجموعة من المهام، اهم هذه المكونات على النحو التالي:

A packet filter 2.1

يقوم هذ المكون بفحص كل *packet* يطب السماح بالدخول الى الشبكة الداخلية و يكون الفحص على IP addresses و أنواع محددة من TCP port numbers، ولا يتم فحص محتوى كل ال packet

ويتم الفحص وفق السياسة المرسومة من قبل المؤسسة التي يتبعها الجدار الناري، وبناء على نتيجة الفحص يتم السماح بالمرور أو الرفض ، وبالطبع فانه لن يتم فحص محتويات كل ال *packet* ، فان هذا بدوره سوف يعيق الاعمال، ولكن يتم التركيز بالفحص على IP addresses وكذلك يتم التركيز بالفحص على أنواع محددة من TCP port numbers، ويتم

منع مجموعة من ال protocols such as SMTP, NetBios session، الذين يشتبه في سلوكهم. امثال هذه ال protocols تمر عن طريق بعض الTCP port numbers و غالبا ما تكون ارقامها عالية، ويمكن منعها بسهولة.

هذا المكون لا يمكن الاستغناء عنه لان من خلاله تتم معرفة الجهة المرسلة (source)، وكذالك خلاله تتم معرفة ال protocols التي تمرعن طريق بعض الTCP port numbers .

2.2 الوكيل*proxy server*A

يقوم هذ المكون بالعمل نيابة عن الأصيل حيث يستقبل الطلبات ويتم فحص الطلبات وفق السياسة المرسومة من قبل المؤسسة ، وبناء على نتيجة الفحص يتم اتخاذ القرار المناسب، فاذا كانت نتيجة الفحص مطابقة للقواعد المطلوبة، يتم طلب التوصيل من الوكيل نيابة عن الأصيل أو رفض طلب التوصيل تماما اذا كانت نتيجة الفحص غير مطابقة للقواعد المطلوبة. في معظم الجدر النارية، يقوم هذ المكون بإنجاز معظم الوظائف ، حيث يحتوي على أنظمة تطبيقات متعددة. يعتمد عمل هذ المكون على فحص المحتوى نفسه، ومعرفة ما اذا كان المحتوى نص ام كود برمجي، ويمنع كل executable program codes، وبشكل خاص JAVA applets، وذلك لوجود احتمال لتسرب احصنة طروادة المؤذية، او بعض الفيروسات، ولهذا السبب فان هذ المكون يحتاج الى استخدام كثيرا من الأنظمة المساعدة والمهارات العالية، وبالتالي تكون قيمة هذ المكون هي الأكثر كلفة اذا قورنت بالمكونات الأخرى. هذا المكون لا يمكن الاستغناء عنه لان من خلاله تتم معرفة هوية صاحب الطلب، ماهية الطلب ونوعه، ومحتواه.

2.3 Statues packet filtering

يقوم هذ المكون بفحص حالة كل packet يطب السماح بالدخول الى الشبكة الداخلية و يكون الفحص معتمدا على الجدول الخاص بمكونات الجلسات الخاصة لطلبات التوصيل عند بداية اية نوع من التواصل. من خلال هذ المكون يتم التأكد من ان أي packet لن يرسل الا الى الجهة الصحيحة، ولن يتم ارساله الى جهة أخرى. هذا المكون مهم لان من خلاله يتم منع الجهات التي تحاول تقمص هويات آخرين غير المرسلين الحقيقين.

2.4 Network address translation NAT

يقوم هذ المكون بعمل قناع على محتويات الشبكة الداخلية حتى لا تتعرض للمشاهدة من خارج الشبكة الداخلية، وبالتالي لا يمكن مهاجمتها، ويتم التعامل مع محتويات الشبكة الداخلية فقط عن طريق عنوان الجدار الناري. مثل هذا الاجراء يعمل على تقليل الخصوم، وبالتالي الحصول على بعض الامن لبعض محتويات الشبكة الداخلية. تكمن أهمية هذا المكون لان من خلاله يتم تقليل الخصوم.

Virtual private network VPN 2.5

يقوم هذ المكون بعمل قناة مشفرة للتواصل بين محتويات مقاطع الشبكة الداخلية عن طريق الشبكة العامة حتى لا تتعرض المحتويات للمشاهدة من خارج الشبكة الداخلية، وبالتالي لا يمكن الاستفادة منها. هذ المكون يمكن الاستغناء عنه اذا كانت محتويات مقاطع الشبكة الداخلية لا تحتاج الشبكة العامة للتواصل.

2.6 اهم متطلبات الجدار الناري

لكي يقوم الجدار الناري بوظائفه لا بد من توفر مجموعة من المتطلبات اهمهن على النحو التالي:

2.6.1 Random number generator RNG مولدا للأرقام العشوائية، وذلك لعمل قناع على محتويات الشبكة الداخلية، وعمل الجدول الخاص بمكونات الجلسات الخاصة بفحص الحالة لكل packet

2.6.2 نظام المفتاح العام public key ، توليد وتشفير وفك تشفير، وذلك للتأكد من الهوية، ولتبادل المفاتيح السرية، او لعمل توقيع الكتروني على وثيقة الكترونية.

2.6.3 نظام (متماثل) تشفير وفك تشفير، وذلك لإخفاء المعلومات التي لا ينبغي كشفها عند التراسل.

2.6.4 نظام عمل توقيع الكتروني والتعرف على توقيع الكتروني في وثيقة الكترونية

2.6.5 نظام توليد مفتاح الجلسة session key وذلك لإخفاء المعلومات التي لا ينبغي كشفها عند التراسل .

2.7 اهم المحددات للجدار الناري

عند التفكير بالجدار الناري يجب ان يؤخذ بعين الاعتبار ان أنشطة المؤسسات مختلفة، وبالتالي لا يمكن تعميم مواصفات جدار ناري على الجميع وانما يجب ان يؤخذ بعين الاعتبار التالي:

2.7.1 حجم الشبكة الداخلية في المؤسسة ، فاذا كان الشبكة الداخلية صغير (4 أجهزة مثلا) فلا تحتاج الى جدار ناري، وانما يتم تخصيص جهاز للتواصل مع الانترنت

2.7.2 نوع الشبكة الداخلية في المؤسسة، كما نعلم ان هناك أنواع مختلفة من الشبكات، كل نوع يتناسب معه نوع من الجدر النارية

2.7.3 نوع نظام التشغيل المستخدم في الشبكة الداخلية، كل نظام تشغيل يتناسب معه نوع من الجدر النارية

2.7.4 نوع النشاط او الأنشطة الممارسة في المؤسسة، كل نشاط يتناسب معه نوع من الجدر النارية

2.7.5 الميزانية المخصصة للحصول على جدار ناري للمؤسسة، لان أسعار الجدر النارية متفاوتة.

2.8 جودة الجدار الناري

من المعلوم ان الجودة تختلف من جهة الى أخرى ، ولكن هناك مجموعة من المعاير من خلالهن يتم الحكم على الجودة المناسبة للجدار الناري .

2.8.1 اهم المعاير التي تقاس بها جودة الجدار الناري اهمهن على النحو التالي:

1. نوع الخوارزمية المستخدمة في التشفير المتماثل وكذلك طول المفتاح المستخدم
2. نوع الخوارزمية المستخدمة في ال Hash function وكذلك طول code المستخدم
3. عدد خانات المعامل n المستخدمة في Public key algorithm ، فكلما كان عدد خانات اكبر كلما زادت الجودة.

4. كيف يتم التعامل مع الاعداد الكبيرة مثل 100 digits and more فما فوق
5. كيف يتم توليد واختبار الاعداد الأولية الكبيرة مثل 100 digits and more فما فوق
6. كيف يتم التعامل مع الدوال الاسية المكونة من اعداد كبيرة مثل 100 digits and more فما فوق
7. نوع خوارزمية توليد الاعداد العشوائية ، وماهي دورتها؟ وما نوع الاختبارات التي اجرية عليها؟
8. كيف يتم حماية كلمة السر عند التعامل من بعد؟
9. عند التعامل مع مفتاح الجلسة session key كم عدد خانات المعامل p المستخدم ، وكم عدد خانات المعامل g المستخدم .

2.9 التقييم النسبي لبعض مكونات الجدار الناري

يمكن عمل تقييم نسبي لبعض المكونات الأساسية للجدار الناري وذلك بالرجوع الى مجموعة المهام التي ينجزها كل مكون، بالإضافة الى complexity (الزمن اللازم لإنجاز المهمة ، الذاكرة اللازمة ، تعقيدات اللغة المستخدمة)، وأيضا المتطلبات اللازمة لإنجاز المهمة، ونحن في هذا البحث سوف نركز فقط على مجموعة المهام التي ينجزها كل مكون، والمتطلبات اللازمة لإنجاز المهمة، ولن نتطرق الى complexity. والجدول التالي يوضح ذلك:

مسلسل	المهمة المنجزة	المكون الذي ينجز المهمة	متطلبات الإنجاز	الوزن النسبي
1	Packet filtering	Packet filtering	Inspection	1
2	Packet filtering status	Packet filtering status	Inspection	1
3	Encrypted /decrypted tunnel	Virtual private network VPN	Encryption/decryption algorithms	2

2	Random numbers generator	Network address translation	IP address masking	4
1	Inspection	Proxy agent	Inspect the contents	5
3	Random numbers generator and Encryption/decryption algorithms	Proxy agent	ntity authentication and Password protection	6
3	Session key algorithm and Encryption/decryption algorithms	Proxy agent	Making accounts	7
1	Nothing	Proxy agent	Lists modification	9
1	Nothing	Proxy agent	Requesting connection/ disconnection	10
2	Encryption/decryption algorithms	Proxy agent	Encryption /decryption	11

من الجدول يتضح ان ال proxy agent هو الذي يقوم بإنجاز اكثر المهام، فهو يتطلب حوالي 65%

2.10 فتح حساب لمستخدم

يتم فتح حساب لمستخدم عن طريق الجدار الناري، يحمل الحساب اسم المستخدم ويمنح رقم من قبل الجدار الناري، ثم يتم توليد مفتاح جلسة session key ، ثم يتم اختيار كلمة السر من قبل المستخدم ويتم تشفيرها مستخدما خوارزمية محددة، و مفتاح الجلسة session key ،ثم يتم ارسالها من قبل المستخدم الى الجدار الناري، ويقوم الجدار الناري بفك الشفرة مستخدما نفس الخوارزمية المحددة، ونفس مفتاح الجلسة session key ،ويتم وضع كلمة السر الخاصة بالحساب، وتكون هذه هي كلمة السر الخاصة بحساب المستخدم، ويمكن استخدام خوارزمية public key لتشفير وفك الشفرة، وفي كلتا الحالتين تكون الحاجة ماسة للتعامل مع الاعداد الكبيرة مثل 100 digits فما فوق.

3 بعض الصعوبات المتعلقة ببعض الوسائل لاستخدام بعض المكونات الأساسية للجدار الناري

توجد بعض الصعوبات تعلق ببعض الوسائل لاستخدام بعض المكونات الأساسية للجدار الناري مما يؤدي الى بعض الثغرات والاختلالات، حيث تأتي الثغرات والاختلالات من استخدام بعض الوسائل غير المناسبة للتطبيقات الآمنة، والتي قد تكون مناسبة نظرياً ولكن في الواقع تحتاج إلى وسائل غير متاحة أو غير مدركة أهمها على النحو التالي:

عند فتح حساب لمستخدم عن طريق الجدار الناري يتم اختيار كلمة السر من قبل المستخدم ويتم تشفيرها اما باستخدام خوارزمية محددة، و مفتاح جلسة session key ،يتم توليده عند الطرفين ، او عن طريق استخدام خوارزمية public key ، وكلا الطريقتين تعتمد على التعامل مع اعداد اولية كبيرة مثل 100 digits and more فما فوق ، وهنا يظهر بعض الاسئلة على النحو التالي:

1. كيف يتم التعامل مع الاعداد الكبيرة مثل 100 digits فما فوق
2. كيف يتم توليد واختبار الاعداد الأولية الكبيرة مثل 100 digits فما فوق
3. كيف يتم التعامل مع الدوال الاسية المكونة من اعداد كبيرة مثل 100 digits فما فوق
4. ما هو نوع خوارزمية توليد الاعداد العشوائية ، وماهي دورتها؟ وما نوع الاختبارات التي اجرية عليها؟

هذا البحث سوف يقدم بعض الطرق للتعامل مع الاعداد الكبير والأولية والدوال الاسية المكونة من اعداد كبيرة مثل 100 digits فما فوق.

3.1 التعامل مع الدوال الاسية المكونة من اعداد كبيرة Exponential function computation

يمكن التعامل مع الدوال الاسية المكونة من اعداد كبيرة مثل 100 digits فما فوق، والتي تكون الدالة فيها على هيئة $x=a^p \bmod n$ ، وبما ان اعلى قيمة ل x فإنها لن تصل الى قيمة n ، وبالتالي فلا داعي لإنجاز العمليات الحسابية a^p وخاصة عندم تكون كلا من قيمتي a,p كبيرتين، فلتعامل مع الدوال الاسية المكونة من $x=a^p \bmod n$ فيمكن تحويل قيمة p الى النظام الثنائي، ثم يتم استخدام الخوارزمية التالية لحساب قيمة x.

3.2 التحقق من هوية المستخدم

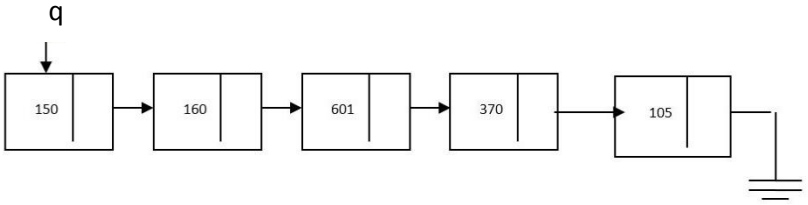
للتحقق من هوية المستخدم يتم الاتي:

1. يرسل المستخدم طلب السماح بالمرور الى الجدار الناري، الطلب يحمل اسم المستخدم ورقمه.
2. الجدار الناري يفحص طلب المستخدم (يفحص رقم المستخدم واسمه)، فاذا كان غير مطابق يرفض، اما اذا كان مطابقا فينتقل الى الفقرة 3 .
3. يولد رقم عشواءي ، يحتفظ بنسخة منه، ثم يشفره مستخدما خوارزمية محددة، وكلمة السر الخاصة بحساب المستخدم كمفتاح ثم يرسله الى المستخدم
4. يقوم المستخدم بفك الشفرة مستخدما نفس الخوارزمية المحددة، وكلمة السر الخاصة بحسابه كمفتاح ثم يرسلها الى الجدار الناري
5. يقوم الجدار الناري بمطابقة ما وصله مع النسخة المحفوظة من ا لرقم العشوائي، فاذا كان غير مطابق يرفض، اما اذا كان مطابقا فيتم السماح بالمرور.

3.3 الفكرة الأساسية للبحث

الفكرة الأساسية للتعامل مع الاعداد الكبيرة مثل 100 digits فما فوق

يمكن التعامل مع الاعداد الكبيرة مثل 100 digits فما فوق عن طريق بناء قوائم متصلة linked lists اذات اتجاه واحد و ذات اتجاهين، بحيث تتكون كل قائمة متصلة من مجموعة من nodes ويتم اسناد كل مجموعة من الارقام n-digits الى node، ثم يتم اجراء العمليات المطلوبة على محتوى ال node الواحدة وترحيل اثر العملية الى ال node التالية، وهكذا حتى آخر node من القائمة المتصلة، ومن ثم يتم التعامل مع الاعداد الكبيرة وكأنها اعداد مجزئة ولكنها مترابطة.



3.3.1 جمع الاعداد الكبيرة

يتم بناء ثلاث قوائم متصلة linked lists ذوات اتجاهين، الأولى للرقم الأول والثانية للرقم الثاني والثالثة للمجموع بحيث تتكون كل قائمة متصلة من مجموعة مناسبة من nodes ، ويتم اسناد الرقم الأول للقائمة الأولى، ويتم اسناد الرقم الثاني للقائمة الثانية كل node تحوي مثلا 9 من الارقام 9-digits ،

يتم اجراء عملية الجمع بين محتويات node الأولى من القائمة الأولى و محتويات node الأولى من القائمة الثانية، ويتم وضع نتيجة الجمع في node الأولى من القائمة الثالثة، ويتم ترحيل الحمل اذا وجد الى العملية التالية، ثم تتم عملية الجمع بين محتويات node الثانية من القائمة الأولى و محتويات node الثانية من القائمة الثانية مع الحمل اذا وجد ويتم وضع نتيجة الجمع في node الثانية من القائمة الثالثة ، وتكرر هذه العمليات حتى تنتهي احدى القائمتان الأولى او الثانية، ويتم جمع الحمل اذا وجد مع محتويات node القائمة المتبقي فيها nodes ويتم وضع نتيجة الجمع في node المقابلة من القائمة الثالثة، وتستمر هذه العمليات حتى تنتهي القائمة المتبقي فيها nodes

3.3.2 طرح الاعداد الكبيرة

يتم بناء ثلاث قوائم متصلة linked lists ذوات اتجاهين، الأولى للرقم الأول (المطروح منه) والثانية للرقم الثاني (المطروح) والثالثة لنتيجة الطر، بحيث تتكون كل قائمة متصلة من

مجموعة مناسبة من nodes ، ويتم اسناد الرقم الأول للقائمة الأولى، ويتم اسناد الرقم الثاني للقائمة الثانية كل node تحوي مثلا 9 من الارقام 9-digits ،

يتم اجراء عملية مقارنة بين محتويات node الأولى من القائمة الأولى و محتويات node الأولى من القائمة الثانية، فاذا كانت محتويات node الأولى من القائمة الأولى اكبر من محتويات node الأولى من القائمة الثانية، فتتم عملية الطرح، ويتم وضع نتيجة الطرح في node الأولى من القائمة الثالثة، اما اذا كانت محتويات node الأولى من القائمة الأولى اصغر من محتويات node الأولى من القائمة الثانية، فيتم إضافة اكبر قيمة يمكن ان تستوعبها n-digits + 1 ، ثم تتم عملية الطرح، ويتم وضع نتيجة الطرح في node الأولى من القائمة الثالثة ويتم ترحيل الإستلاف (1) ليتم تنقيصه من العملية التالية، وتكرر هذه العمليات حتى تنتهي القائمتان الأولى و الثانية، او الثانية، ويتم طرح الاستلاف اذا وجد من محتويات node القائمة الأولى ويتم وضع نتيجة الطرح في node المقابلة من القائمة الثالثة، وتستمر هذه العمليات حتى تنتهي القائمة الأولى.

3.3.3 ضرب الاعداد الكبيرة

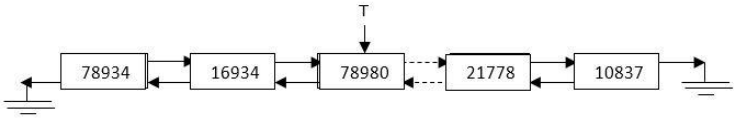
لإجراء عملية الضرب، يتم بناء ثلاث قوائم متصلة linked lists ذوات اتجاهين، الأولى للرقم الأول (المضرب) والثانية للرقم الثاني (المضرب فيه) والثالثة لنتيجة الضرب، بحيث تتكون كل قائمة متصلة من مجموعة مناسبة من nodes ، ويتم اسناد الرقم الأول للقائمة الأولى، ويتم اسناد الرقم الثاني للقائمة الثانية كل node تحوي مثلا 4 من الارقام n-digits ، ويتم اسناد الرقم صفر لكل node في القائمة الثالثة.

يتم اجراء عملية الضرب بين محتويات node الأولى من القائمة الأولى و محتويات node الأولى من القائمة الثانية، ويتم جمع نتيجة الضرب مع محتويات node الأولى من القائمة الثالثة ووضع النتيجة في node الأولى من القائمة الثالثة، ويتم ترحيل الزيادة عن 4-digits الناتجة من عملية الضرب ليتم جمعها الى العملية التالية، وتكرر هذه العمليات حتى تنتهي القائمة الأولى، ثم يتم الانتقال الى node الثانية، من القائمة الثانية، ثم تتم اجراء عملية الضرب بين محتويات node الثانية من القائمة الثانية، وبين محتويات كل node في القائمة الأولى ووضع النتيجة في node المقابلة من القائمة الثالثة بعد ان تتم عملية إزاحة بمقدار node نحو اليمين بين محتويات القائمة الثالثة.

، وتستمر هذه العمليات حتى تنتهي القائمة الثانية.

3.3.4 قسمة الاعداد الكبيرة

يتم بناء اربع قوائم متصلة linked lists ذوات اتجاهين، تخصص الأولى للرقم الأول (المقسوم) وتخصص الثانية للرقم الثاني (المقسوم عليه) وتخصص الثالثة لحساب مؤشر لنتيجة القسمة ، وتخصص الرابعة لنتيجة القسمة ، بحيث تتكون كل قائمة متصلة من مجموعة مناسبة من nodes ، ويتم اسناد الرقم الأول للقائمة الأولى(المقسوم)، ويتم اسناد الرقم الثاني للقائمة الثانية (المقسوم عليه) كل node تحوي مثلا على n من الارقام n-digits ، ويتم اسناد القيمة صفر لكل node في القائمة الثالثة التي سوف تحتوي على مؤشر لنتيجة القسمة.



من المعروف ان عملية القسمة بين مجموعة من الاعداد على مجموعة اخرى من الاعداد تتم بأخذ عدد مناسب من خانات المقسوم، تكون قيمتها اكبر او تساوي قيمة المقسوم عليه، ثم تتم عملية القسمة، ويتم وضع خارج القسمة في مكان ، كما يتم ترحيل باقي القسمة الى يسار الاعداد التي لم تتم اجراء عملية القسمة عليها، ثم تكرر نفس العمليات السابقة الى ان تغطى كل خانات المقسوم، ولكن هذا يتم عند التعامل مع الاعداد الصغيرة ، ولا يمكن التعامل مع الاعداد الكبيرة مثل 100 digits فما فوق بهذه الطريقة، لان اجراء عملية القسمة تتم بتوزيع عدد مناسب من خانات المقسوم على كل خانات المقسوم عليه دفعة واحدة ، ولكن الحاسوب لا يستطيع ان يتعامل مع كل خانات المقسوم عليه دفعة واحدة اذا كانت كبيرة، وعلى ذلك لا بد من عمل خوارزمية للوصول الى المطلوب دون الحاجة الى التعامل مع كل خانات المقسوم عليه دفعة واحدة.

3.4 ملخص الفكرة

ملخص الفكرة هو ان من الممكن تحويل القسمة الى عمليات طرح، ولكن كم عدد عمليات الطرح المطلوبة؟ وخاصة اذا كان عدد خانات المقسوم عليه كبيرة، وكذلك عدد خانات المقسوم

كبيرة جدا، في هذه الحالة نحتاج الى عدد كبير من عمليات الطرح، قد يكون من المستحيل انجازهن .

فهل يمكن تحويل القسمة الى عملية ضرب واحدة وبضع عمليات طرح؟ الجواب نعم، ولكن بوجود مؤشر قسمة مناسب، ولكن ما هو مؤشر القسمة أولا ؟

ان العوامل الأكثر تأثيرا في تحديد نتيجة القسمة هي قيمة الخانات التي في اقصى اليسار من المقسوم، وكذلك قيمة الخانات التي في اقصى اليسار من المقسوم عليه، فاذا اخذة قيمة مجموعة من الخانات من يسار المقسوم، ثم قسمة على قيمة مجموعة من الخانات من يسار المقسوم عليه، تكون هذه العملية الأكثر تأثيرا في تحديد نتيجة القسمة، وكلما كان عدد الخانات المأخوذة كبير، كلما كانت نتيجة القسمة اكثر دقة وصحيحة.

وعلى وجه العموم يجب ان يكون حاصل ضرب مؤشر القسمة في قيمة المقسوم عليه اقل او تساوي قيمة المقسوم، ولكن كما ذكرنا سابقا ان قيمة مؤشر القسمة تأتي من قسمة قيمة مجموعة من الخانات من يسار المقسوم على قيمة مجموعة من الخانات من يسار المقسوم عليه، وليس من قسمة قيمة المقسوم على المقسوم عليه، وبالتالي لا يمكن ضمان ان يكون دائما حاصل ضرب مؤشر القسمة في قيمة المقسوم عليه اقل او تساوي قيمة المقسوم، الا تم اضافة 1 الى قيمة المقسوم عليه ، ثم يتم حساب قيمة مؤشر القسمة، وفي هذه الحالة يمكن ضمان ان يكون حاصل ضرب مؤشر القسمة في قيمة المقسوم عليه اقل او يساوي قيمة المقسوم، وتكون نتيجة القسمة صحيحة. والمثال التالي يوضح ذلك.

3.4.1 اذا كان لدينا هذا العدد 36984527381787654 المقسوم، هذا العدد 53646312416المقسوم عليه، فهل يمكن تحويل القسمة الى عملية ضرب واحدة وبضع عمليات طرح؟ وما هو مؤشر القسمة ؟

اذا اخذنا قيمة ال 5 خانات من يسار المقسوم في متغير ما مثلا $x=36984$، و اخذنا قيمة ال 5 خانات من يسار المقسوم عليه 53646 في متغير اخر مثلا y، اضافة 1 الى قيمة $y=53647$ ، ثم يتم قسمة x على y ، ويتم وضع نتيجة القسمة في متغير ما مثلا $z=0.6894141594$.

الفرق بين عدد خانات المقسوم وعدد خانات المقسوم عليه هو 6، يتم ضرب قيمة المتغير z في 10^6، فتكون قيمة $z= 689414.1594$، ويكون هذا هو مؤشر القسمة.

عند ضرب قيمة المتغير z ، مؤشر القسمة، في قيمة المقسوم عليه 53646312416، تكون نتيجة الضرب هي 36984518830565485 .

عند تنقيص 36984518830565485 من 36984527381787654، تكون النتيجة هي 8551222199 ،

عند مقارنة نتيجة التنقيص 8551222199 مع قيمة المقسوم عليه 53646312416 تجدها اصغر من قيمة المقسوم عليه، وبالتالي تكون نتيجة القسمة صحيحة.

والمثال التالي سوف يوضح ان نتيجة القسمة غير صحيحة اذا لم يتم اضافة 1 الى قيمة الخانات المأخوذة من يسار المقسوم عليه. فاذا كان لدينا بعض التعديل في المثال السابق العدد 35235169494المقسوم، هذا العدد 362109468 المقسوم عليه، فاذا اخذنا قيمة ال 5 خانات من يسار المقسوم في متغير ما مثلا x=35235، و اخذنا قيمة ال 5 خانات من يسار المقسوم عليه 36210 في متغير اخر مثلا y، y=36210 ، ثم يتم قسمة x على y ، ويتم وضع نتيجة القسمة في متغير ما مثلا z=0.9730737365.

الفرق بين عدد خانات المقسوم وعدد خانات المقسوم عليه هو 2، يتم ضرب قيمة المتغير z في 10^2، فتكون قيمة z=97.30737365، ويكون هذا هو مؤشر القسمة.

عند ضرب قيمة المتغير z ، مؤشر القسمة،97.30737365 في قيمة المقسوم عليه 362109468، تكون نتيجة الضرب هي 35234948235 ، وهي اكبر من قيمة المقسوم 35235169494 ، وبالتالي تكون نتيجة القسمة غير صحيحة.

3.4.2 مقترح لخوارزمية لإجراء قسمة الاعداد الكبيرة

وهذا مقترح لخوارزمية لإجراء قسمة الاعداد الكبيرة مثل 100 digits فما فوق دون الحاجة الى التعامل مع كل خانات المقسوم عليه دفعة واحدة، وهو على النحو التالي:

1. يتم وضع كل محتويات المقسوم في قائمة متصلة linked list ذات اتجاهين، مثلا القائمة صفر، ويتم تقسيمها الى blocks ، كل block عدد خاناته تقريبا عدد خانات المقسوم عليه او يزيد عنها 1 او اكثر قليلا،
2. يتم وضع ال block في قائمة متصلة linked list ذات اتجاهين، مثلا القائمة الأولى، و يتم وضع المقسوم عليه كاملا في قائمة أخرى متصلة linked list ذات اتجاهين، مثلا القائمة الثانية.

3. يتم وضع قيمة ال 10 خانات من يسار المقسوم في متغير ما مثلا x، و يتم وضع قيمة ال 10 خانات من يسار المقسوم عليه في متغير اخر مثلا y،
4. يتم اضافة 1 الى قيمة y ، ثم يتم قسمة x على y ، ويتم وضع نتيجة القسمة في متغير ما مثلا z
5. يتم ضرب قيمة المتغير z في 10 اس (الفرق بين عدد خانات المقسوم وعدد خانات المقسوم عليه) اذا كان هناك فرق بين عدد الخانات، ويتم وضع نتيجة الضرب في المتغير z ، ويتم وضع نسخة من نتيجة الضرب في القائمة الرابعة.
6. يتم ضرب قيمة المتغير z في محتوى القائمة الثانية، ويتم وضع نتيجة الضرب في القائمة الثالثة.
7. يتم تنقيص محتويات القائمة الثالثة من محتويات القائمة الأولى، ويتم وضع النتيجة في القائمة الثالثة ،
8. يتم مقارنة محتويات القائمة الثالثة مع محتويات القائمة الثانية ، فاذا كانت محتويات القائمة الثالثة اكبر من محتويات القائمة الثانية يتم الانتقال الى الخطوة 9 اما إذا كانت محتويات القائمة الثالثة اصغر من محتويات القائمة الثانية فيتم الانتقال الى الخطوة 10
9. يتم تنقيص محتويات القائمة الثانية من محتويات القائمة الثالثة، ويتم وضع النتيجة في القائمة الثالثة ، ويتم إضافة واحد الى محتويات القائمة الرابعة ، ثم يتم الانتقال الى الخطوة 8 ،
10. يتم الانتقال الى block المقطع التالي من القائمة صفر إذا كانت هناك بقية من محتويات القائمة صفر، ويتم وضع قيمة محتويات القائمة الثالثة غير الصفر الى يسار ذلك المقطع ، او يظهر النتيجة النهائية لخارج القسمة الموجودة في القائمة الرابعة، وكذلك باقي القسمة الموجودة في القائمة الثالثة .

3.4.3 هنا بعض الأمثلة لعمليات القسمة

enter the no of digits in upper

85

enter the no of digits in lower

77

x= 5.61782e+06 Y= 201176 D= 279.249 w = 279 k= 6

0561782303462245036515913310529576757662330297501244769766980
4616071714034547605
0201175862932579155328631129761283910920614302710169933045300
44728504764382785
enter the no of digits in upper 88
enter the no of digits in lower 80

L1H
0561782303462245036515913310529576757662330297501244769766980461607
171403454760502121163 2

L2H
0561782303462245036515913310529576757662330297501244769766980461607
17140345476050

L1H
0561782303462245036515913310529576757662330297501244769766980461607
171403454760502121163 2
L3H
0561781601234365708709617665637938595691383219588372893211018252881
594394490506181549375 0
L5H
0000000702227879327806295644891638161970947077912871876555962208725
5770089642543205717882
L2H
0561782303462245036515913310529576757662330297501244769766980461607
17140345476050

enter the no of digits in upper
88
enter the no of digits in lower
83

0100184315349303150336462820002336862072927228414009966735744659851701995434069545316214

0528476095569541955309190030983741059849668274577769525470990952799717764234276278856

L1H
0100184315349303150336462820002336862072927228414009966735744659851701995434069545316214
L3H
0100183213437118068467963154173587792715701614811707768943535754922242473458917541827319 2
L5H
0000001101912185081868499665828749069357225613602302197792208904928927726084489412704302 2

enter the no of digits in upper
27
enter the no of digits in lower
23

L1H 0100184315349303150336462820

L2H 0528476095569541955309 19

L1H 0100184315349303150336462820
L3H 0100146220110428200531091505
L5H 0000038095238874949805371315

enter the no of digits in upper
27
enter the no of digits in lower
23

0528476095569541955309 19
L1H 0100184315349303150336462820

L2H 0528476095569541955309 19

L1H 0100184315349303150336462820
L3H 0100146220110428200531091505
L5H 0000038095238874949805371315
L1H 0100184315349303150336462820

Last one

enter the no of digits in uper
31
enter the no of digits in lower
25

L1H 0561782303462245036515913310 5295

L2H 056178230346224503651591 33

L1H 0561782303462245036515913310 5295
L3H 0561782303462245036515913300 0000
L5H 0000000000000000000000000010 5295

3.4.4 تحليل الخوارزمية

عند تحليل الخوارزمية لإجراء قسمة الاعداد الكبيرة مثل 100 digits فما فوق، نجد أنه لا حاجة الى التعامل مع كل خانات المقسوم دفعة واحدة، وانما نحتاج الى التعامل مع كل خانات المقسوم عليه في عملية ضرب وعملية طرح، وفي كلتا العمليتين يمكن التعامل مع

كل الخانات مهما كان عددهن، أي انه تم تحويل عملية القسمة الى عملية ضرب واحدة وبضع عمليات طرح.

في هذه الخوارزمية نجد أنه يمكن التعامل مع كل خانات المقسوم عليه مهما كان عددهن، وبالتالي يمكن إجراء قسمة الاعداد الكبيرة ، عن طريق تجزْئتهن الى مجموعات، عدد خانات المجموعة معتمدة على عدد خانات المقسوم عليه تكون نتيجة القسمة صحيحة.

ان قسمة عدد من فئة n خانة على عدد اخر من فئة m خانة يحتاج الى الاتي:

n*mعملية ضرب احادي الخانات (ضرب خانة في خانة)

n*mعملية جمع احادي الخانات (جمع خانة مع خانة)

nعملية طرح احادي الخانات (طرح خانة من خانة)

15n موقع في ذاكرة الحاسوب، وهذا كله من وجهة نظر الحاسوب شيء يسير ويمكن الحصول عليه بسهولة ويسر.

References:

[1]. Bruce Schneier (2001) "Applied Cryptography" 2nd Edition John Wiley & Sons. (ASIA) Pvt. Ltd., 2 Clementi Loop # 02-01, Singapore 129809.

[2]. David M. Burton (2003) "Elementary Number Theory" 2nd Edition Universal Book Stall New Delhi India.

[3]. Deborah Russell and G. T. Gangemi Sr (2001) " Computer Security Basics" O'Reilly & Associates, Inc., New York.

[4]. Dennin, Dorothy E.(1983) "Cryptography and Data Security" Library of Congress Cataloging in Publication Data. Addison-Wesley. whit Mobile Code"[1999] Wiley Computer Publishing USA.

[5]. Douglasr. Stinson (2002) "Cryptography: Theory and Practice" Department of Combinatory and Optimization University of Waterloo, Waterloo, Ontario Canada. 2nd Edition, Chapman & Hall/CRC.

[6]. Ellis Horowitz, University of Southern California, Sartaj Sahni, University of Florida and Sanguthevar Rajasekran, University of Florida, (1999) Computer Algorithms. Galgotia Publication Pvt. Lid, New Delhi, India.

[7]. Richard E. Smith.(2002) Internet Cryptography, Addison-Wesley.

[8]. Rothke-B 1999 DES is dead! Long live? Information-Systems Security. Vol..7, No..1; Spring 1998; p.57-60 1998 Journal-article.

[9]. Thomas H. Cormen, Charles E. Leiseroin and Ronald L. Rivest Clifford Stein (2002) "Introduction to Algorithms" 2nd Edn. Prentice,Hall of India, Pvt. Ltd., New Delhi-110 001-2002.

[10]. Tjahjono-BPN; Fulcher-J; Pieprzyk 1999 Implementation of private key data encription Electron. Book Syst., Singapore journal-of-Electrical-and-Electronics Engineering,-Australia. Vol..19, No..4;Dec. 1999; p.201-11 1999 Journal-article.

[11]. William Stallings (2003) "Cryptography and Network Security: Principles and Practice" 3rd Edn. India Reprint. Agrawal-M IETE-Technical-Review.

[12]. Wright-MA in 1999 Network-Security. Nov. 1999; p.11-14 PY: 1999 RT: Journal-article.

An Ontology-based Automatic Text Summarization: Survey

Bakeel Azman
Department of Computer Science
Sana'a University, Sana'a, Yemen
Bakeel056@gmail.com

Ibrahim Ahmed Al-Baltah
Department of Information Technology
Sana'a University, Sana'a, Yemen
albalta2020@gmail.com

Abstract- In recent year, text summarization has gained its importance due to the data overflowing on the web. Huge amount of information is available in all over the places and topics. So it is difficult to understand the main content of the document without reading the entire document. It takes much significant time. Due to that, summarization models emerge to construct a concise and sufficiently accurate approximate representation of data summary or synopsis promptly. Furthermore, summarization systems receive the importance because of their variety of applications like summaries of newspaper articles, books, magazines, events, weather forecast, stock market, news, etc. However, this paper presents a survey of automatic text summarization models developed in the recent years. The uniqueness of our survey is the specific consideration on ontology-based text summarization approaches. One of the drawbacks that we recognized in the reviewed models is the shortage in rhetorical figures cognition and processing them, as well as lacks of paraphrase the selected verbosity sentences.

Keywords—ontology; automatic summarization; summarizer

Introduction

Information overload increases in a great demand for more capable and dynamic text summarizers. As information is available in abundance for every topic on the internet, condensing the important information in the form of summary would benefit a number of users. Therefore, the need of text summarization has emerged to provide users with the predominant and summarized information. As the information communication technologies are growing at a high speed, a large number of electronic documents are available on-line and users are facing difficulty to find relevant information. Four main summarization needs are considered by Huang et al. [1] : coverage of information,

information significance, redundancy in information and cohesion in text.

The big problem related to information is the overload. For instance, 1.39 Billion URLs catalogued by Google, which disperse the user and make the process of accomplishing his demand is very difficult [2]. As a result of this complexity and due to the user necessity, a noticeable number of works have proposed different approaches. Some of these approaches are information retrieval, document clustering, information extraction, visualization, question answering, and text summarization. This paper will focus only on automatic text summarization based on ontology.

Abstractive summarization is an automatic summarization that builds an internal semantic representation and then use natural language generation techniques to create a summary that is closer to what a human might generate. Such a summary might contain words not explicitly present in the original [3]. An automatic text summarization takes various types of approach (e.g. single-document vs. multi-document, extractive vs. abstractive, generic vs. query-focused, supervised vs. unsupervised, indicative vs. informative, sentence-based vs. concept-based, multi-lingual vs. cross-lingual vs. mono-lingual summarization) [4].

The abstractive summarization is an efficient form of summarization compared to extractive summarization, as it retrieves information from multiple documents to create a precise summary of information. There are two broad abstractive summarization methods namely, structured based approach and semantic based approach. The former approach encodes most vital data from the document(s) through psychological feature schemas like templates, extraction rules and alternative structures like tree, ontology, lead and body, rule, graph based structure [5]. The latter approach is employed to feed into a natural language generation system. This technique specialize in identifying noun phrases and verb phrases by processing linguistic data. According to [5], the methods of semantic based are namely, multimodal semantic, semantic graph based, information item based, and semantic text representation model. Single-document summarization deals with one data file trying to get more accuracy in summarizing.

It is worth nothing to point out that, when humans attempt to summarize a document they usually go through some sequence steps starting from reading the text, understanding it, and finally writing the summary. In a contrast, the automatic text summarization system generates a summary, i.e. short length text that includes all the important information of the document. This encourages researches to develop different approaches in order to get more accurate and fast summarizing system automatically. Therefore, researchers have been trying to improve techniques that aim to generate summaries that could match with the human made summary. The ultimate goal of automatic text summarization is to create summaries that are similar to human-created abstracts.

To achieve this goal, researches have adapted the ontology to improve the process of text summarization. The predominant role of the ontology at summarization is to enrich semantically the concepts and joint relations between them more other method. Identifying the corpus concepts and disambiguating uncertain relations is one of the most important tasks in ontology-based summarization. It gives an effective way for modeling the context in which summaries are generated in different application domains, as well semantic dependency of textual context [3]. Since ontology is assumed to be specific domain, it is especially more focuses on bounding of related concepts and properties that domain. Ontology, with abundant concise concepts and rich domain-related information, can capture the hidden semantic information [6]. The roles of linguistic ontologies are dual role: The first one is to present and define the vocabulary used. This is achieved by a dictionary which list all the terms actually used in language. Secondly, linguistic ontology is the result of a terminology agreement between users' community [3]. This agreement defines which term is used to represent a concept in order to avoid ambiguity. This process is called vocabulary normalization. When a concept could be described by two synonym terms, the normalization process selects one of those to be the preferred label of the concept. Anyway, a summarization should produce a summary with preserve important information and shorter than original document(s).

This paper surveys some of the well-known and recent ontology-based automatic text summarization models in some details. The aims of this survey paper are to provide an overview of the current state of the research, to identify the challenges that are still standing as barriers in this topic, as well as, to provide insights into the trends of ontology-based automatic text summarization topic.

The rest of the paper is organized as follow: section II presents the related works, section III analyzes the surveyed models, section IV presents the challenges and research trends, and the conclusion of this study is presented in section V.

literature review

Text summarization is not a new topic; it has been in the research since the time of rapid growth of the information, and the urgent user need for the summary of huge information. As a result of that, many studies have focused on proposing and developing different approaches to accomplish the summarization task. On the other hand, some studies have focused on analyzing the existing works, in which will be further analyzed in this study.

From the automatic text summarization point of view, researchers in their study [4] explore some related methods. The focus of their study was on the empirical methods of extractive summarization approach. In many extractive summarization systems, several algorithms have been adopted which employ machine learning and statistical techniques. On one hand, some studies employ naïve-Bayes methods [7] which assume the independence between the features used while others do not make the same assumption. On the other hand, some other studies adapt different techniques such as decision trees, hidden markov models, log-linear models and neural networks. For single-document summarization, the process of producing the summary is usually straightforward and it could be accomplished by choosing the highest ranked sentences according to their scores. For multi-document summarization, the process is usually more complex as it involves checking for redundancy, diversity and relevance to the user's specific needs. The natural language analysis [8] can be classified as entity-level analysis and discourse level analysis. Entity level analysis is usually performed by building an internal

representation of text through the use of text entities and modeling the relationships between them. Discourse level analysis on the other hand models the overall structure of the text and its relation to the goal of the summary. Other approach [9], sentence compression is defined as follows: given a sequence of words $W = w_1 w_2 ... w_n$ that constitute a sentence, and a subsequence $w_{i1}\ w_{i2} ... w_{ik}$, with $1 \leq i_1 < i_2 < ... i_k \leq n$, that is a compressed version of W. Note that there are 2^n possibilities of output. The measure that [6] adopted is a statistical method for both single document and multi documents. Extractive method is flowed. The researcher [6] mentioned that the big difficulty encounters automatic evaluation is systems development for assessing the automatic summarizers.

B. N. R. Kasture. et al. [5] formed an overall idea about the extractive and concentrated on the abstractive summarization methods. Extractive summaries are formulated by weighting the sentences as a function of high frequency words. The most frequently occurring or the most favorably positioned text is considered to be the most important. The methods used for determining the weights of the sentences are cue method, location method and title method. Abstractive summarization is an efficient form of summarization compared to extractive summarization as it retrieves information from multiple documents to create precise summary of information. This has gained its popularity due to the ability of developing new sentences to tell the important information from text documents. An abstractive summarizer displays the summarized information in a coherent form that is easily readable and grammatically correct. Readability or linguistic quality is an important catalyst for improving the quality of a summary. Abstractive summarization techniques can be further classified into two categories, as mentioned above, structured based and semantic based methods.

A. Structured based method.

This method encodes the most important information from document(s) through cognitive schemas such as:

1) Tree based method: It uses a dependency tree to represent the text/contents of a document. Different algorithms are used to perform content selection process for summary e.g. theme intersection algorithm or algorithm that uses local alignment across pair of parsed sentences.

The technique uses either a language generator or an algorithm for generation of summary [10].

2) Template based method: This method takes a template to represent a whole document. Linguistic patterns or extraction rules are matched to identify text snippets that will be mapped into template slots [11].

3) Rule based method: in this method, the documents to be summarized are represented in terms of categories and a list of aspects. Content selection module selects the best candidate among the ones generated by information extraction rules to answer one or more aspects of a category. Finally, generation patterns are used to generate the summary sentences.

4) Ontology based method: It prompted many researchers to use ontology (knowledge base) to improve the process of summarization. From Web perspective, most documents on the web are domain related because they discuss the same topic or event. Each domain has its own knowledge structure and that can be better represented by ontology. Fuzzy ontology with fuzzy concepts is introduced for Chinese news summarization [12] to model uncertain information and better describe the domain knowledge.

B. Semantic based method

In this method, semantic representation of document(s) is used to feed into natural language generation system. This method focuses on identifying noun phrases and verb phrases by processing linguistic data [13]. In fact, some models have adapted the idea of semantic based method such as:

1) Multimodal semantic model: the key idea of the semantic model is to capture concepts and the relationships among them. Then it uses these concepts and the relationships to semantically represent the contents (text and images) of multimodal documents. Consequently, the important concepts are rated based on some measure and finally the selected concepts are expressed as sentences to form summary [14].

2) Semantic graph based method: It aims to summarize a document by creating a semantic graph called Rich Semantic Graph (RSG) for the original document, reducing the generated semantic graph, and then

generating the final abstractive summary from the reduced semantic graph [15].

However, C. Jishma Mohan M. et al. [3] covered an important summarization aspect based on ontology methods. They discussed various works carried out using ontology for abstractive text summarization. Furthermore, they introduced the ontology roles that may contribute in more accuracy summarizer, and the objectives are gained by applying ontology in abstractive summarization. The models that have been covered in [3] and [15] are either specific-application [16] or less coupled with summarization [17],[18]. Therefore, this study will cover the most known and accepted models that depend on ontology method.

reviewed models

In fact, there are many studies that have proven and shown the success of ontology using in different field, such as, artificial intelligence, semantic Web…etc. Text summarization is not an excepted field, thus, some of the current studies have attempted to use ontology to accomplish the automatic text summarization successfully. Therefore, this study, unlike other studies, will focus on surveying only the ontology-based automatic text summarization models.

As we have observed, most of the current ontology-based models are constructed from common layers, namely, preprocessing layer, knowledge base constructing layer, extracting features and ranking layer, summary sentence generation or selection layer.

Elena baralis et al. [19] suggested summarization system relies on the YAGO[1] ontology for summarizing multi-documents to carry out a summary concedes that corpus the corresponding user query. The proposed system has three main modules:

1) First module concerns the recognition and disambiguation of the entity (one or more words). The ontology method comes to solve these statements through the use of YAGO ontology (which is constituent from Wikipedia[2], GeoNames[3] and Wordnet[4]). It disambiguates all entities by comparing two contextual indicators

(corpus entity and ontology entity) for corresponding context of use. The output of this module benefits also from ontology entity metadata like entity popularity and other features; therefore, it computes the importance of a sentence in concerned corpus and ranks each entity.

2) Second module receives its input from the previous module output in which encompasses with each candidate entity its related scores. Therefore, the computation of sentences scores will be performed using the sentence importance with the required query. The importance of this process is because the summary includes only the most pertinent and semantically meaningful document content.

3) Third module is the common procedure at last most summarization system. It is responsible for selection of the sentences that have most ranking in documents and least similarity with preceded summary sentences. Baralis model [19] is a good system fairly, but it doesn't consider a threshold, hence, it is difficult for users to control the summary quantity.

R. Ragunath and and N. Sivaranjani [20] and Hennig et al. [21] adopted concept terms technique along with based on ontology. The multi-documents summarization, after preprocessing, depends on Term Frequency and Inverse Document Frequency (TFIDF) method to extract the concepts. Due to some reasons, such as, the ontology is mainly specific-domain, and the extracted concepts are multi-domain, the clustering is presented.

K-mean algorithm is adopted to construct the related concepts to groups. Then, embed all document concepts with cosine similarity of the other documents. They used the protégé tool for building the hierarchical ontology, and encode the ontology with a tree structure. Each node includes the concepts represented by the node's children. The root of the ontology will always get the highest score, while nodes in the second level, which represent subtopics, will get different scores. Creating bag-of tags for each sentence by collecting the nodes computed by the hierarchical classifier. If a sentence is mapped to multiple sub trees in the taxonomy, it will include all nodes from every sub tree. Starting at the root node, the distance measure algorithm computes the similarity of a sentence to all child nodes, then determines

the mean μ and standard deviation σ of the resulting similarities, and selects all nodes for further exploration whose similarity to the sentence sim (sentence, node) > μ +ασ. This model relies on hierarchical classifier' nodes within specific ontology level, as well as the classifier's confidence weight to compute the subtree overlap for all sentences.

Summarization process is not limited to the English language or to any specific language, which means it could be applied to Arabic language. Ibrahim Imam et al. [22] proposed a method for Arabic text summarization based on ontology (OSSAD) to research field. They used the ontology method in expansion of both the user's query by Arabic WordNet and knowledge base that extracted from corpus. This model relies on C_value algorithm [23] for extracting and recognizing the concepts counter to the most models that rely on frequency techniques. Arabic language lacks to tools and systems comparing to other languages (e.g. English language), due to that a researchers are compelled to adopt other tools to serve Arabic language processing. Aliane [24] model is adopted to extract the relations from the documents. The middle component is responsible for the construction of knowledge. Every sentence is attached a set of features scores having a feature vector (e.g. number occurrence of words from original query and extended query and knowledge base). The sentences are also labeled an importance and related tag (related, semi related, informative…) to the concern domain. The model created for the C4.5 algorithm is applied to each group of sentences representing a single file. According to the required summary length, the appropriate number of sentences is displayed to the user.

Chang-Shing Lee et al. [12] extend the domain ontology into fuzzy ontology, hence domain ontology incapacitates toward the join degree of concept. They employed the ontology method in more than one procedure in their system starting in term classifier, then in news corpus summarization. The fuzzy ontology concepts are constructed by news domain experts and embedded with membership degree. The summarization process adapts a temporary fuzzy ontology as container for corresponding the output terms from documents preprocessing with

the constructed fuzzy ontology. The terms that will be included in the temporary ontology only which are related with the fuzzy ontology's concepts or relationships. This procedure aims to extract the sentences path. Then, the workflow comes to sentence generator module to trace back the temporary fuzzy ontology nodes and arcs in order to combine the sentence and store the whole sentences in a set. Computing of the possibility of the relevance a term with a concept (e.g. "{Temperature; 0.04, 1.0, 0.64}," the membership degrees of the fuzzy concept "Temperature" for weather events " E1: Typhoon,""E2 : Cold current," and " E3: Rain" are 0.04, 1.0, and 0.64, respectively). That set of sentences pass in final process which is sentences filter for eliminating of the redundant, and ignore other sentences that don't add an additional semantic meaning in summary. Extension domain ontology that Lee proposed can enhance of power ontology method for dealing with uncertainty reasoning problems and gives flexibility in knowledge representation. Arabic summarization and whole language processing is poor in several models, system modeling, researches, so that it doesn't expect an efficient summarizer system.

challenges and trends

After conducting this survey, the best of our knowledge, researchers should paid more consideration on the linguistic analysis and tools development to assist the context understanding of the text. Therefore, the produced automatic text summarizers will perform the required task efficiently, which result a high quality summary. The summarizer should summarize the text equal or better than the summaries that done by references (human). The main factor for generating high quality text summary is the understanding of sentences context. Nevertheless most of the current techniques and methodologies are suffering from critical short comings in cognizance of the text context. One of the most challenges is the automatic process for building ontologies by extracting and perceiving serialized concepts in the documents before considering the use of ontology itself as a tool to assist in text summarization.

It is practically impossible to manually evaluate multiple documents for obtaining an unbiased view. Due to the fact that, summary evaluation is not an easy task for human to judge the quality of the

summarization and the information that should be presented in the summary. Therefore, there is a need for mutual and reliable metrics to evaluate the automatic summarization [25]. Another challenge is the dynamic information changes in the dynamic environments such as the Web, and the different user needs for summary from the same data. Furthermore, paraphrasing is a common phenomenon in language, and it could be done frequently from different users which makes the automatic summarization more difficult.

Nowadays, ideal summarizer system is not arised, for instance state-of-the-art summarizers incapable of facing up to complex rhetorical relations and speech discourse.

As a potential research opportunity, we suggest to integrate speech discourse into text summarizer to obtain high quality summary. Another opportunity could take the advantage and capability of visualization field, in which input is in the text format and output can be represented through statistics, tables, graphics, visual rating scales, etc to provide the result to user in a simple manner.

There is change in the type of summaries to adapt to changing user requirements. Initially generic single document summaries were generated will not be desirable now because of availability of large amount of data in different formats and different languages. Due to fast development of technology, multi-document, multi-lingual, multimedia summaries have gained popularity. Another direction are being attempted to capture and harness the power of paraphrases to improve summarization systems and increase results accuracy.

conclusion

This paper attempts to enrich the state of the art by providing a specific survey of the text summarization from ontology-based perspective. Also, it gives researchers a recent and clear picture of the reviewed approaches. We reviewed variant text summarization models, which adapted the ontology methods as a favorable module for obtaining precise summarization system. By using the domain-related ontology, summarizer can better capture the semantic relevance among the concepts within documents, and thus lead to better summarization

results. The most common techniques that have been reviewed in this study are, query-based, mono-language, sentence-oriented summarization. As a recommendation of this study, text summarization still need further research to reach its maturity level. Compressions involving lexical substitutions, paraphrase and reformulation is difficult. Furthermore, distinct summaries that should satisfy user requirements from multimedia are up to date issue. Our future work will enhance ontology-based summarization by including rhetorical figures interpretation ability.

References

[9] L. Huang, Y. He, F. Wei1, and W. Li "Modeling document summarization as multi-objective optimization," Intelligent Information Technology and Security Informatics (IITSI), 2010 Third International Symposium on. IEEE, 2010.

[10] Abdullah Bawakid, "Automatic Documents Summarization Using Ontology based Methodologies," Ph.D , University Of Birmingham, 2012.

[11] Jishma Mohan M, Sunitha C, Amal Ganesha, Dr.Jaya A, "A study on ontology based abstractive summarization," Fourth international conference on recent trends in computer science & engineering Procedia Computer Science 87 (2016) 32 – 37. doi: 10.1016/j.procs.2016.05.122.

[12] Das, Dipanjan and Andre F.T. Martins, " A survey on automatic text summarization,". Literature Survey for Language and Statistics II course at CMU 4 (2007): 192-195.

[13] N. R. Kasture1, N. Yargal , N. N. Singh, N. Kulkarni and V. Mathur, "A survey on methods of abstractive text summarization," International Journal For Research In Emerging Science And Technology, vol-1, ISSUE-6, November-2014.

[14] Li, Lei, et al. "Ontology-enriched multi-document summarization in disaster management." Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval. ACM, 2010..

[15] Kupiec, J., Pedersen, J., and Chen, F. , "A trainable document summarizer," In Proceedings SIGIR '95, pages 68-73, New York, NY, USA,(1995).

[16] Nenkova, A. “Automatic text summarization of newswire: Lessons learned from the document understanding conference,” In Proceedings of AAAI 2005, Pittsburgh, USA.

[17] Knight, K. and Marcu, D. , “Statistics-based summarization - step one: Sentence compression,” In AAAI/IAAI, pages 703-710, 2000.

[18] R. Barzilay and K. R. McKeown, "Information fusion in the context of multi-document summarization," in Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics, 1999, pp. 550-557.

[19] S. M. Harabagiu and F. Lacatusu, "Generating single and multi-document summaries with gistexter," in Document Understanding Conferences, 2002.

[20] C.-S. Lee, Z.-W. Jian and L.-K. Huang, "A fuzzy ontology and its application to news summarization," Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, vol. 35, pp. 859-880, 2005.

[21] H. Saggion and G. Lapalme, "Generating indicative-informative summaries with sumUM," Computational Linguistics, vol. 28, pp. 497-526, 2002.

[22] C. F. Greenbacker, "Towards a framework for abstractive summarization of multimodal documents," ACL HLT 2011, p. 75, 2011.

[23] I. F. Moawad and M. Aref, "Semantic graph reduction approach for abstractive Text Summarization," in Computer Engineering & Systems (ICCES), 2012 Seventh International Conference on, 2012, pp. 132-138.

[24] Zhang, X., Cheng, G., Qu, Y. , “Ontology summarization based on rdf sentence graph”. In: 16th Inter. World Wide Web Conference Banff, Alberta, Canada, May 8-12 (2007).

[25] Thanh Tran, Philipp Cimiano, Sebastian Rudolph and Rudi Studer,” Ontology-based interpretation of keywords for semantic search” Institute AIFB, Universität Karlsruhe, Germany.

[26] Peroni, S., Motta, E., d'Aquin, M.: “Identifying key concepts in an ontology through the integration of cognitive principles with statistical and topological measures”. In: 3rd AsianSemantic Web Conference, Bangkok, Thailand (2008).

[27] E. Baralis, L. Cagliero, S. Jabeen, A. Fiori and S. Shah, "Multi-document summarization based on the Yago ontology," Elsevier, Expert Systems with Applications 40 (2013) 6976–6984.

[28] R. Ragunath and N. Sivaranjani, "Ontology based text document summarization system using concept terms," ARPN Journal of Engineering and Applied Sciences, VOL. 10, NO. 6, April 2015.

[29] L. Hennig, W. Umbrath and R. Wetzker, "An ontology-based approach to text summarization," 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology.

[30] I. Imam, A. Hamouda and H. Abdul Khalek, "An ontology-based summarization system for arabic documents (OSSAD)," International Journal of Computer Applications (0975 – 8887), Volume 74– No.17, July 2013.

[31] S. Ananiadou and H. Mima, *"*An application and evaluation of the C/NC-value approach for the automatic term recognition of multi-word units in Japanese*"*, International Journal of Terminology, Vol 6, No. 2, pp. 175–194, 2000.

[32] A. C. Mazari, H. Aliane, Z. Alimazighi. "Automatic construction of ontology from Arabic texts", ICWIT, Vol 867, pp. 193-202. 2012.

[33] A. Khan, N. Salim and Y.J. Kumar, "Genetic semantic graph approach for multidocument abstractive summarization," IEEE, ISBN: 978-1-4673-6832-2, 2015.

Cloud Computing Framework for Yemeni Universities

Mohammed Ahmed Al_shaibani
Department of Engineering and Information Technology
Shaibani2007@gmail.com
Master Thesis, 1437-2016

Nibras Ali Alazazi
Department of Engineering and Information Technology
AL Andalus University

Yahya Abdulrazaq Kotran
Department of Computer Science
Sana'a University, Yemen.

I. ABSTRACT:

Cloud Computing is considered in three out of ten top strategic technology trends for 2014 year according to Gartner [1]. In my opinion it is the green field for all IT researcher to develop our Higher Education (HE). The Universities in this time is becoming completely dependent with the IT on their teaching-learning, communication, service delivery, and business requirements.

whereby shared resources, software and information, are provided to computers and devices on-demand, like the electricity grid. With the rapid growth of Cloud Computing, the use of Clouds in educational settings can provide great opportunities for reduction of in-house data centers and the delegation of a portion or all of the Information Technology infrastructure capability to a third party, giving the opportunity for students and teachers to quick access to various applications, systems and resources through the Internet, and share files and documents, and the exchange of duties and projects among students "improving their learning outcomes". Meanwhile, all of services on Cloud Education can be accessed using any devices such as mobile phones, computers, and tablet computers. Despite all of advantages of cloud, computing it still not fully adopted in academic institutions sector.

It is only 4% of cloud computing used in education and the other 96% is related to industrial sectors and services [2].

Hence, The paper discusses the advantages of cloud computing for educational institutions, The research finding shows that Cloud Computing is the better Information and Communications Technology (ICT) utilization mechanism for Education. This paper identifies the reasons for the slow rate of adoption of cloud computing at university level in Yemeni universities,

discusses the challenges faced and proposes a cloud computing adoption framework that contains strategic guidelines to overcome the major challenges identified and a roadmap called CCFYU (Cloud Computing Framework for Yemeni Universities) for the successful adoption of cloud computing by universities, case study Alandalus university, sana'a university and science and Technology university. The Data sources are direct observation of the employees of universities, student and teacher and make interview with the engineer in Yemen net that is the ISP for Yemen. The researcher engaged the heads of the various companies selected for this study likes Yemen soft and Natco companies to establish relationship with them and explain the benefits of this research as well as scheduling for the interview appointments that dial with the cloud computing services and interview with the IT support for university and make questionnaires "online and offline" of student, some employees and some teachers, And is analyzed through SPSS statistical Software all that for Evaluation the readiness of Yemeni universities to use cloud computing and to help overcome the challenges of transition for secure cloud computing or safety migration through CCFYU and five stages as a roadmap help the universities and any organization to do self assessment before adoption to the cloud computing.

Keywords: cloud computing, Adoption, Framework, Roadmap, CCFYU, Higher Education (HE).

II. INTRODUCTION

Frequent advancing in information technology has becomes a great challenge in every academic institution in providing necessary ICT infrastructures. Constantly updating of ICT infrastructures in academic institutions for education process, research and development of training activities is becoming a big issue in this crucial financial crisis facing by every national economy.

In such situation a relatively new concept and constantly evolving cloud technology is starting introduce across the world in academic institutions.[3]

Cloud computing will change the way we work and business will not be as usual anymore both at work and for personal life.[4]

Cloud computing offers opportunities to improve the quality of education by offering flexibility and accessibility through the Internet. This can enable more dynamic and interactive learning experiences and allow students and teachers in multiple locations to collaborate and communicate more

effectively[5]. In addition, cloud-based services can offer users and academic institution cost savings and access to scalable computing power[6][7]. Research highlighted by Forbes outlines that by 2020 large American corporations can "achieve annual energy savings of $12.3 billion" and small to medium sized companies could reduce their energy consumption by "90%" should they move to the cloud.[8]

At present, as many universities are trying to update their

IT infrastructure and data, but they are facing few challenges

which can be solved by cloud computing. Students will have access to all software anytime, anywhere and any technological devices connected internet by suggested cloud structure. Also, students will have access to development platform, and develop their applications, and store on university infrastructure. In this way, lecturers will focus their basic tasks and not lose their workforce. With suggested cloud structure, universities can open their technology infrastructures to businesses and industries for research

advancements and develop university-industry collaboration[26]. In this research, we aim to suggest a Cloud Computing infrastructure scenario to using universities in Yemen.

III. Problem Statement:

Although the predictions on the increase in the rate of adoption of the cloud technology, the acceptance level for higher education is still very low. Current higher education settings are traditional and often unable to face the demand of rapid changes in technologies in an interactive learning environment, Nowadays a lot of universities or users need to process a long amount of information data or need to do some complex operations, for example some mathematical models calculations, and they need a high amount of process power to resolve it. Maybe the power of a personal computer can be not enough to finish in a determinate time one task. Also sometimes a user wants to use a program for few times, for example for making some practices in university, and maybe it's pointless to make install this user a program, so for both we have to bring to the user the necessary tools to be able of executing a problem in the cloud of the university

Every day that goes by, the research and educational needs of universities' change with developing technology; All the software and hardware of universities' must be renewed in accordance with the changes.

Accordingly The process of purchasing, maintaining, and administering computing assets requires a large investment of financial and manpower resources for university, and make all that is in an availability or accessibility. One option that centralizes computing assets and can lower costs and manpower requirements for these organizations is the use of centralized computing assets provided as Cloud Computing.

Currently, many universities are interested in using cloud computing capabilities, but they do not know how can be moved. So this research attempts to answer the following

research questions:

1. What are the factors which organizations take into account when deciding about adopting Cloud Computing and how important are these factors?
2. To what extent Yemeni Universities are aware of Cloud Computing?
3. Is it suitable for Yemeni Universities to adopt of Cloud Computing?
4. What measures can be put in place to ensure a safer and more reliable transition to the cloud?
5. What is the roadmap and framework for the successful adoption of cloud computing.

IV. Objectives of the Research

1. Study the knowledge of cloud computing in the universities of Yemen and study the infrastructure of its current system.
2. Investigate the factors that affect adoption of cloud computing in Yemeni Universities.
3. propose the architecture or framework for cloud in Yemeni Universities including various service and deployment models.
4. provide strategies and recommendations for effective implementation cloud environment in Yemeni Universities and for successful and efficient migration of its traditional system to cloud based system.

V. Related Works:

1- Saidhbi(2012) in the research Cloud Computing Framework for Ethiopian Higher Education Institutions proposed the implementation of a central hybrid cloud computing infrastructure that combines both the current local infrastructure of the universities as the private cloud and public cloud to

enable the sharing of educational resources and collaboration within all universities in Ethiopia and the global educational community, so that Ethiopian higher institutions can enjoy the benefits of ICT in an efficient and affordable way (Saidhbi, 2012). The research further states that by deploying the proposed hybrid cloud model, the risks of privacy and other security challenges can be avoided as critical and sensitive data will be housed in a private cloud.

2- In a study on Cloud Computing in Higher Education in Jordan by Massadeh and Meslah (2013) suggested that Jordanian universities consider adopting cloud computing as a way of meeting the growing demands of IT services and managing the tight budget due to very limited financial support from the government. The researchers believe that implementing cloud computing will be a strategy to offer good business models for the Jordanian universities as they do not have sufficient resources to manage the required IT support for development, educational, and research activities that should be provided in an ideal higher education environment.

3- In article "Cloud Computing Adoption Model for Universities to Increase ICT Proficiency",2014, by Safiya and others identifies the reasons for the slow rate of adoption of cloud computing at university level, discusses the challenges faced and proposes a cloud computing adoption model that contains strategic guidelines to overcome the major challenges identified and a roadmap for the successful adoption of cloud computing by universities. The model was tested in one of the universities and found to be both useful and appropriate for adopting cloud computing at university level. the article proposes a cloud computing adoption model consists of :

1.Strategic guidelines to overcome security and privacy Concerns.

2. Strategic guidelines to overcome reliability concerns of the Cloud Service Providers.

3. A roadmap for the successful adoption of cloud Computing.

The proposed roadmap has seven stages: planning, choosing the right deployment model, choosing the most suitable service models, vendor selection, negotiating the SLA, migration, and integration.

4- In the research The "Adoption of Cloud Computing Technology in Higher Education Institutions: Concerns and Challenges (Case Study on Islamic University of Gaza "IUG")",2013, by Ahmed Jameel This research aims at showing the concerns and challenges of the adoption of Cloud Computing technology in Higher Education Institutions, case study Islamic university of Gaza (IUG).

Using the descriptive analytical method to study the effects of the main five dimensions (Top management support, Support and integration with university Services, Skills of IT human resources, Security effectiveness and Cost reduction) on the adoption of Cloud Computing technology. This

research focuses on IUG as a case study of the academic institutions of Palestine which is the first from among other universities in terms of modern technology utilizing in its operations.
Moreover, it's used several services of Cloud Computing technology for example IUG Gmail, Facebook, Fliker, and IUG Tube …etc.
The researcher used a questionnaire as a data collection tool. The results showed that there is a significant relationship between the adoption of Cloud Computing and the five independent variables at level of significance α= 0.05. The research recommended that IUG can adopt Cloud Computing technology in its operations, if it is interesting on the side of IT human resource through training, scientific missions, and innovations,..etc.

5- In research "Cloud Computing: Strategies for Cloud Computing Adoption",2010, by Faith Shimba, Dublin Institute of Technology this research project aims at developing a roadmap called ROCCA (Roadmap for Cloud Computing Adoption), which provides organizations with a number of steps for adopting cloud computing and building trust. An associated framework called ROCCA achievement Framework (RAF) is also proposed. RAF is a framework that uses the criteria in the ROCCA to build a framework for measuring the adherence level to the proposed roadmap. It presents in detail the technological factors key to a successful cloud computing adoption, and it introduces
The technology underlying cloud computing, and describes different cloud computing delivery and deployment models.

VI. Research Methodology:

The methodology that is used in this research consists of two section:

First Section: The research methodology and methods, where this section will describe the methodology used, the methods of data collection, identify the research population. In addition to explaining the steps for setting up a research tool, which are represented In three parts. The first part involved interviewing managers of Information Technology of cloud service supports and providers. This method enabled the researcher to gain firsthand information and direct communication to the top management of the three companies used. These face to face interviews allow the researcher to delve deep into the experiences and/or knowledge of the participants in order to gain maximum data from the interview process.

The second part of the data collection process consisted of interviews for IT support of universities.

The third part is survey questionnaires for employees, teachers, and student of universities. The questionnaires will be online survey and offline "hard paper".

Second Section: This section describes statistical analysis for each dimension of the questionnaire, where we'll help the researcher to build the framework and the roadmap for cloud computing "cloud computing adaption" in the universities in Yemen, and analyze the responses of research population members on the scale used by the statistical system SPSS.

VII. Cloud Computing Service:

There are three main cloud computing service; Software as a service (SaaS), Platform as a service (PaaS) and Infrastructure as a Service (IaaS) [10].

A. Software as a service(SaaS) : This type of services focus on the software and service of IT systems such as recording applications, finical applications and reports applications in various domains. Therefore, SaaS provide online user's interfaces as connection points between end users and organization services.

B. Platform as a service (PaaS): The main purpose of PaaS is to manage, control, process and operate the gather information between different applications and systems storages and infrastructures. Thus, PaaS considered cloud computing operating system to operate the gather information between SaaS and IaaS sides. There are many bold PaaS examples such as online programming language debugger, tasks interruptions management and automated daily operations.

C. Infrastructure as a Service (IaaS): It is the higher services layer in cloud computing services structure; SaaS represent the infrastructure that responsible about store systems operating systems, applications, data and information and the network requirement to connect between cloud services. Therefore, SaaS considered as systems hardware and storages resources. Figure 1 illustrate the structure of cloud computing services.

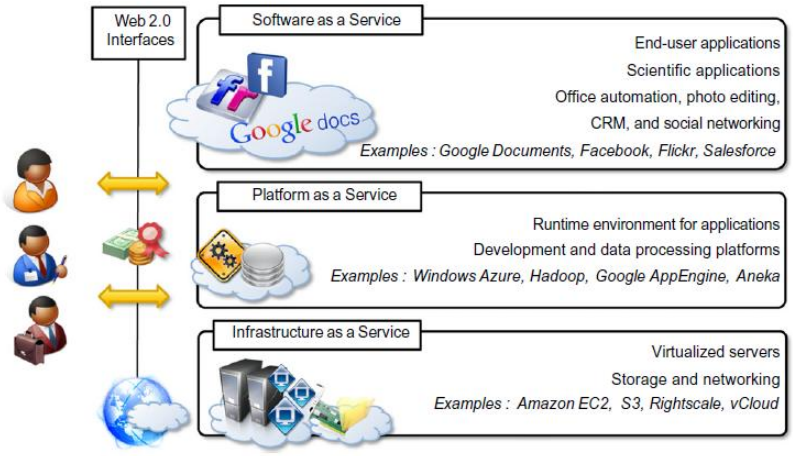


Figure 1: Cloud computing Services[32]

VIII. Cloud Deployment Model "Types of clouds"

There is four different types of cloud and each of cloud computing deployment approach has its own characteristics such as costs, security and availability, and the organization can decide the suitable deployment approach for its businesses based on working environment.

A. Public Cloud:

Public cloud is deployment approach that allows organizations or users to use the cloud resources. Therefore, public cloud uses WWW networks as infrastructure to communicate between customers and cloud resources. The main advantage of this approach is globalizing the business markets which give the organization the opportunities to maximize their consumers and publish their products widely.[11][12][13]

B. Private Cloud:

In this model the cloud resources are not shared by unknown third parties. The cloud resources in this model may be located within the client organization premises or offsite. In this model the clients security and compliance requirements are not affected though, this offering does not bring the benefits associated with reduced capital expenditure in IT infrastructure investments.[14] For example, a company may host email in their own private cloud, but archive email in a provider's public cloud.

c. Hybrid or heterogeneous clouds:

It is a combination of public and private cloud models or more clouds that try to address the limitations of each approach, but are bound together by standardized or proprietary technology that enable data and application

portability. In hybrid cloud, part of service infrastructure runs in private cloud while the remaining part runs in public clouds. Hybrid cloud offer more flexibility than both public and private clouds.[11][12][13]

D. Community Cloud: A community cloud contains features of both the public and private cloud models. Like a public cloud, the community cloud may contain software, data storage, and computing resources used by multiple organizations. Where this model differs from the public model is that the infrastructure is used exclusively by a group of organizations known to each other. [11][12][13]

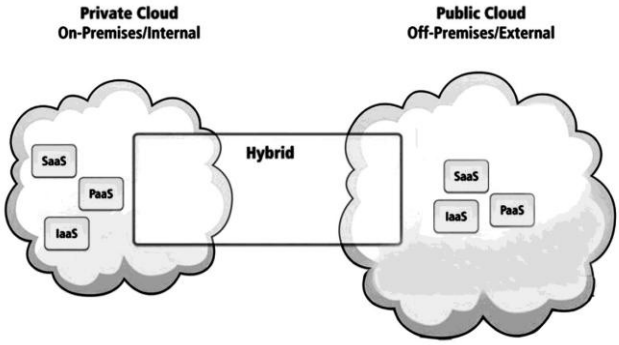


Figure 2: Cloud Deployment Model

IX. Cloud Computing in HE:

The educational cloud computing industry is still in the beginning. Nevertheless, cloud computing has potential to play a wide and significant role in education technology in the near future. It is because now cloud computing spreads fast in all industries.[29] In recent days, many research institutes are struggling to adapt Cloud Computing for solving problems that are continuous increasing computing and storage. There are three main factors interests in Cloud Computing[21]:

1- Rapid decrease in hardware cost and increase in computing power and storage capacity, and the advent of multi-core architecture and modern supercomputers consisting of hundreds of thousands of cores. The education industry now faces a new set of challenges that is driving a fundamental transformation across the education. Their customer demographics, behaviors and expectations have changed. They face revenue pressures along with increasing sensitivities from their customers on the return on investment for dollars spent on higher education. [28]

2- The exponentially growing data size in scientific instrumentation simulation and Internet publishing and archiving.

3- The wide-spread adoption of Services Computing and Web 2.0 applications. The Cloud Computing trend of replacing software traditionally installed on campus computers (and the computers themselves) with applications delivered via the internet is driven by aims of reducing universities' IT complexity and cost [22]. Cloud Computing could be a technological innovation that both reduces IT costs for the college and eliminates many of the time-related constraints for students, making learning tools accessible for a larger number of students [23]. There are many benefits of cloud computing for educational institute and below are listed a few of them:

- With cloud computing, universities can open their technology infrastructures to businesses and industries for research advancements.
- The efficiencies of cloud computing can help universities keep pace with ever-growing resource requirements and energy costs.
- The extended reach of cloud computing enables institutions to teach students in new, different ways and help them manage projects and massive workloads. When students enter the global workforce they will better understand the value of new technologies [24].
- Cloud computing allows students and teachers to use applications without installing them on their computers and also allows access to saved files from any computer with an Internet connection [25].

X. the overcoming Concerns of Cloud Adoption:

1-security and privacy concerns:
the greatest barrier to cloud adoption continues to be security. In fact, a recent poll shows that 65% of organizations list security[16].
5 Steps for Building a Secure Cloud:
Step 1: Plan: The best way to approach cloud security is to integrate it with your overall cloud planning early in the process(understand your risk tolerance, identify the best deployment models for your specific needs, and detect potential exposure points for sensitive data and processes. With a private cloud implementation, since you own the infrastructure, security is under your own control. You establish security and compliance standards, as well as enforce and measure them.[16]

Step 2:Protect Your Hardware and Infrastructure: A best practice for cloud implementation is to layer technologies to develop a strong security net that protects your data, applications and platforms, and network at all levels.

Step 3:Secure Your Data: Of course, as you move workloads to the cloud, your priority is to keep your data out of the wrong hands. the main targets to hit when you implement a data security plan are Data Loss Protection, Email Encryption, and Data Encryption.

Encryption*:* Encryption is an effective, well-established way to protect sensitive data. This is the process of changing or transforming information into a form that cannot be understood by any unauthorized person. By using this technique, the data are translated into a secret code that cannot be understood by anyone else except those who have the code or password to decrypt the encrypted information.[15] The provider and vendor allows you to encrypt your data in your owns private environment, after that you can upload the data, so nobody can decrypt or see your data even the provider itself.

Encryption is the main method used to ensure the security of data stored in the cloud.[15]

***Digital signature*:** an electronic signature used to authenticate the identity of the user of the services provided over the cloud, by using this technique, the user must provide the appropriate login or access credentials before they can have access to the information or application they want to use. This will help to ensure the authenticity, accountability, and integrity of data in the cloud.[15]

Step 4: Gradual sequence of migration: The migration toward cloud should not be done all at once, as much as the cloud offers cost savings, increased agility and efficiency caution should be taken and the movement gradual with low risks applications going first. This will give the university time to see whether the cloud project or the chosen vendor is worthy if so, then the other applications can be moved step by step.

Step 5: Choose the Right Cloud Service Provider: The chosen cloud provider(s) security measures and the kind of security mechanisms, infrastructure and configuration in place to ensure the safety of data stored on their cloud, should be thoroughly investigated. Also, the plan for security events by the service provider detailing the responsibilities and actions to be taken in the event of a security breach should be understood, analyzed, and ensured to be in line with the required standards set by cloud computing bodies like the NIST (National Institute of Standards and Technology) and the CSA (Cloud Security Alliance) to ensure that the right level of security is provided by the cloud vendors to their clients and there are appropriate backups in place in case of any problems or the occurrences of disasters like fire, flood, or earthquake [17]. This investigation is necessary by adopters of the cloud technology in the educational sector to ensure constant availability and

confidentiality because of the level of privacy that is required to preserve research results and other confidential information.

Finally in another word Cloud providers must maintain confidentiality, integrity and availability (CIA) by establishing security requirements to satisfy educational cloud computing systems. Some of these requirements are identification and authentication accounts for students, faculty members and administration staff to verify and validate each individual by username and password. Some need control permissions, priorities and resource ownership (authorization). Encryption techniques should be employed to protect sensitive data of institution such as exams, grades, etc. from tampering or unauthorized access. There is also need to ensure non-repudiation is some circumstances which means the transactions cannot be denied using time stamps, digital signatures and confirmation receipts.[34][35][36][37]

2-Reliability of the service provider concerns: The reliability of the service provider, in other words trust, is a major obstacle and barrier to the adoption of cloud computing.

Reliability has also been an issue for cloud users. For example, in February 2008, Salesforce.com customers were without service for 6 hours while Amazon's S3: simple storage service and EC2 experienced 3 hours outage in the same month a few days later and 8 hours outage in July.[38]

Cloud service providers needed to be reliable and trusted so that adopters of the cloud technology can be confident to entrust their vital information to them for safekeeping. The need to look up, properly research, and make a comparison between different cloud vendors to ensure reliability of the chosen vendor(s) cannot be overemphasized.

This is because information is the live wire of universities, and all efforts should be put in place to ensure that moving to the cloud brings true relief, solutions, and improved services. These can only be achieved, and the promises of the cloud enjoyed if the cloud vendor used is reliable and trustworthy. To achieve this and overcome the barrier posed by lack of trust and reliability of cloud service providers, the following guidelines should be followed:

Reliability and availability: To ensure that the services in the cloud are reliable and up and running well at all times, reliable cloud vendor is required. The chosen cloud service provider should be one that guarantees service level, uptime and availability 99% of the time. It was mentioned that an 100% availability is impossible.[36][37] The chosen vendor should have redundancy of power, cooling systems, security system, servers, storage, excellent Internet connection, and fire suppression systems among other things to ensure that the required services are consistently and constantly available.

Reputation of good track record: The chosen cloud vendor should be one with good track record; this can be ascertained by how well their services function properly without frequent downtimes and when they occur, how long they last

before service is fully restored. The reasons behind downtimes and frequency of occurrence should be considered, this information can be gotten by finding out who their customers are and how well they have enjoyed or not enjoyed the services of the vendor. This should be carefully considered before data are moved to any vendors' data center.

Providing security in the cloud: The university must be assured of tight, well-defined security services in the cloud before they employ the services of any vendor. These security services include identity management, access control as well as authorization and authentication mechanisms to ensure the right level of control within the cloud environment and that only authorized personnel can make any changes or additions to the data and applications in the cloud as a way of ensuring the security, privacy, and confidentiality of data. The service provider should have a comprehensive security infrastructure in place at all levels of the services they provide.

SLA management: The cloud vendors should give guarantee by providing service levels for all services they are offering and ensure to meet the requirements of the SLA. The SLA should be negotiated to meet the expected level of service quality and should include refund guarantees or some kind of penalties if the promised service level is not delivered. This will keep the service providers on their toes to meet up with the terms and requirements of the SLA and the clients assured of quality service delivery. Also, the copyright laws as contained in the vendors' SLA and that of the location where the vendors' infrastructures are located should be carefully considered before commitments are made.

Transparency and code of practice: The choice of the cloud vendor should be one that is transparent in their dealings with clients and follow the code of practice unique to them as an organization and also that provided by cloud regulatory bodies to ensure the provision of highly secured and efficient services to their customers and be willing to explain any ambiguities and provide clarity to their clients when needed. The service provider's chosen should be ones that are accountable and live up to their claims and promises.

3- Bandwidth:

Internet bandwidth is the backbone of the internet-based educational services. The quality of service relies on the connection speed, which can require investment in the network infrastructure.[35][37]

4- **Lack of Skills, Knowledge and Expertise:** It's different in the cloud, and many IT organizations may not have the necessary tools or resources to implement, monitor and manage cloud solutions. It's not what they are geared to do. Educating staff about new processes and tool sets, or hiring staff with new skills, may be necessary. increasingly so as more of your operations and applications move to the cloud over time. Selecting the right service provider will definitely help ease the transition and fill gaps.

Integration with Existing Infrastructure: This is a difficult yet essential piece of maximizing the value of cloud services. Frankly, after time it will be not a

problem because we need with time to develop the university's software and hardware and also not problem gains to the benefits from cloud services into an organization.

.

XI. When should you avoid the cloud?

In the following cases:

- if the provider either don't support higher security or prevents you from hosting even encrypted data.
- An application requires greater reliability or speed than the Internet.
- You want control over your assets, including physical possession of the hardware your data resides on. A private cloud offers one solution if you still want to take advantage of cloud benefits.

XII. A Proposed Work:

A) Strategic Planning for the successful adoption

One of solution for ensuring a successful transition to the cloud is to follow this 5 stages, that we proposed as shown in figure(3), we will discuss it below.

This five stages is like a roadmap help the universities and any organization to do self assessment before adoption to the cloud computing.

Every cloud implementation should be as unique as the campus it serves. When undertaking any major technology initiative, it is necessary to carefully define objectives and requirements, aligning them with the business needs as well as the technology architecture and strategy. It is also imperative to understand the risks and develop plans for mitigating them. This approach is critical in order to maximize success and return on investment.

it good idea for ensuring a successful transition to the cloud is to turn to a professional services group" University IT experts" for assistance in developing a strategy that best suits your college or university's needs. This team of experts will work with university IT teams to create a wish list for making the transition to the cloud easier. Items on the wish list may include[18]:

- Choice of vendors, partners, and cloud-computing solutions.
- Alignment with business strategy and goals.
- A comprehensive, architectural approach.
- A full service and solution offering with robust security.
- Measurable benefits such as time-to-market.

because it will save a lot of costly mistakes from happening, and the university can call at least one external cloud expert who will give professional, expert, and practical advice as that is their area of expertise.

As mansion The five (5) stages in the adoption of cloud computing project are: analysis, planning, adoption, migration and management.

Stage1: preparation:

1. In this initial phase, decision makers will Start by discovering how much cloud computing is already taking place in their education institution, and understand cloud service offerings, benefits, the risks, and best practices.
2. The costs, benefits, and operational changes required to successfully migrate to a cloud computing model should also be evaluated, and It is at this stage that the initial requirements, feasibility, project scope, costs and initial plan will be developed. University IT experts should evaluate their cloud services and architecture security risks, focusing on protecting access and providing on-demand security options within a services catalog for their users. in order to analyze the strengths, weakness, opportunities and threats of existing systems a SWOT matrix is useful[40]
3. The university should work with experts who have extensive experience in multiple technology areas, such as virtualization, service orchestration, automated provisioning, and the security that underpins network architectures
4. Formal and informal training to introduce cloud concepts should be offered and, where possible, tied to early projects. The human pillar is very important for the framework to work. The administration of university should take it upon themselves to develop the personnel that will be identified to work in the university cloud.
   The personnel should be trained in all the aspects of datacenter technology, which include grid computing, virtualization, clustering, they should be trained in most of the operating system. The training will enable the university employees to support the cloud computing efficiently.
   Training also should be diversified to vendor specific platform; it should be the responsibility of all vendors to train the personnel on the specific products they
   install in the datacenter.
5. Existing Web workloads are ideal candidates for this early phase of adoption such as web applications that are not linked to sensitive institutional data.

Stage2: Planning and Design:

1- After the decision to adopt the cloud technology has been taken, The university will begin to more systematically explore how to use the cloud, the University IT experts build detailed architecture design, data-center–specific expertise, and security designed from end to end.

The University IT experts build prototype for cloud [19]:

1- Choosing the right deployment model, so As it is recommended, the hybrid model is the best deployment model because it combines the strengths of the public and private clouds and handles their weaknesses more efficiently. The hybrid cloud provides scalability without boundaries; it is more cost-effective, gives the needed security, and offers great flexibility by giving its users the opportunity to explore different operational avenues.

2- Choosing the most suitable service models: For instance, using IaaS delivers everything: servers, storage, space, and networking equipment as a service. PaaS provides the platform for application developers to build and host their applications whereas SaaS provides complete applications to the end user of the cloud service; it only requires that the user has a web browser and is connected to the Internet.

3- Vendor selection: it is extremely important to investigate into the selected vendor before outsourcing.

After investigating the reliability of the service provider, the next line of consideration is the location where the data will be stored; this is important because of legal compliance that comes with different locations. Different countries have different laws on the right to information in the land, information security, privacy, data protection, and different levels of restrictions; therefore, it is important to understand the local data protection requirement of the country or location where the data will be housed before commitments are made.

4- Negotiating the SLA: The SLA is a very important document; it is a binding contract between the university and the cloud service provider. The terms of the SLA should be negotiated and agreed upon by the university and chosen vendors before the deal is signed. Among other things, the agreements should include the following:

-Planned downtimes by the vendor to check bugs, do maintenance, and updates should be scheduled at times when the impact will be least felt. For example, at midnight of the university's location and preferably over the weekends; such times should be communicated to the university beforehand.

-In the case of service transfer to another vendor, a seamless transfer without delay, downtime, and data loss should be assured.

- Besides the service costs, all hidden costs within the documents for extra charges of any kind and terms and conditions written in tiny fonts should be made clear and properly understood by the university before the SLA is signed to avoid any misunderstandings in the future.

- The chosen vendor should assure 99.9% availability and have an immediate data recovery plan in case of any disaster.

The university Decide The services that needs migration and the services that needs to kept with the institute are decided.

The resulting of this stage is including an end-to-end architecture for example a migration roadmap, a common control framework, a security technology

framework, physical safety and security, and the future cloud services evolution, particularly with respect to integration with internal systems.

Stage3: Integration:

1-This stage will mark a growing acceleration towards cloud adoption. The systems/ application integration is done to ensure that the candidate applications will be able to function with the internal applications that are not migrated to the cloud and also with the cloud infrastructure of choice.

2-This stage will see the initial development of organizational structures, as well as service design and development to support the growing needs of cloud environments. Outsourcing strategies are decided upon and the benchmarks developed in the planning stage. collaboration with vendors is crucial in establishing SLA agreements and different security policies and best practices to ensure compliance and trust as it is palned in the second stage. The last thing in this stage is contract development and signing that meets the user requirements for using cloud services.

Stage4: Migration:

1- Migrate the data and applications to cloud.
2- The transition to the cloud may be achieved gradually starting from testing a pilot project in cloud and the finalizing the application chosen for the cloud.
3- Selecting the Vendor based on set benchmarks as it is planned in second stage, and Put the plane stage into practice.
4- Support and provide adequate training to all users for successful migration.
5- contract development and signing that vendor that meets the user requirements for using cloud service.
6- In some instances, financial or unusual technical limitations may necessitate continuing to serve needs with local infrastructure, though over time this will become a very small proportion of IT services "smooth migration from your existing environment to a cloud utility computing architecture, while helping ensure adherence to plans and enabling on-time delivery of a fully implemented cloud-computing model".
7- enterprises determine which applications to move into the cloud first, which applications to move later, and which applications should remain in-house, sometimes we need to Categorize databases to determine what can move as a service. Moving a collection of applications and their associated databases may be easier than moving component parts of the service.
8- Migrating systems without evaluating or adjusting resource needs will limit the financial benefit of migration[20].
9- University might replace a technology that does not support public cloud deployments[20].

10- There will be services that don't make sense to move to the cloud or that need to have at least some on-premises component (e.g., DHCP, DNS, and Authentication).

Finally it is also important to understand these important basics regarding cloud providers[20]:

- You should consider the sensitivity of your data, and decide if and how you will encrypt your data while it is in transit and while it is at rest.
- You can set highly granular permissions to manage access of a user within your organization to specific service operations, data, and resources in the cloud for greater security control.
- Security should be designed in at every layer of the environment. Institutions should require the use of two-factor authentication where possible and use a least privilege design model.

Stage5: Management, Optimization & Maintenance:

1- The project now should be fully operational in the cloud.
2- This stage is for Contact Management, Vendor management, Ongoing maintenance, user support, Architectural reviews, Security audits and Process improvements.
3- University will need to understand how backups, snapshots, and data restores to development and test systems map from current practice to the cloud.
4- Monitor and control the project to ensure successful migration.

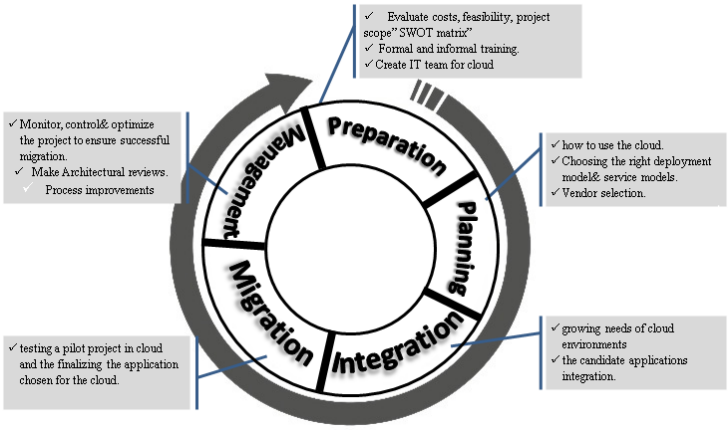


Figure 3: Roadmap to Cloud Computing Adoption

B) CCFYU(Cloud Computing Framework for Yemeni Universities)

1- **Hybrid Cloud** is one of the Cloud Computing deployment models. It provides the ability to access, manage, and use third-party (vendors) resources from multiple Cloud Service Providers (CSPs) and combines them within in-house infrastructure the framework needs to deploy a hybrid cloud as the most suitable deployment model for universities.

Deploying the hybrid cloud will offer cost benefits to the management of the university. with the hybrid model, the fear of privacy and other security-related issues are avoided, as very critical and sensitive data whose confidentiality cannot be compromised will be hosted on premise in the university's datacenter, which serves as a private cloud, and the private cloud is owned and managed by the university and its accessed is limited only to students and faculty, staff of the university.

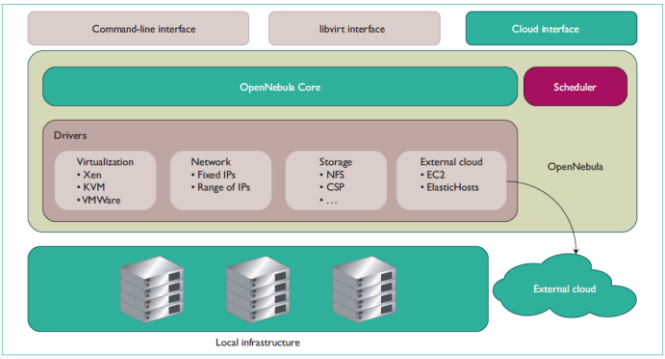
2- **We propose** that the public cloud should combine the services of different cloud service providers or third-party to serve students and staff of the university more efficiently as well as avoid the problem of vendor lock-in. Thereby Using such a model allowed us to avoid lock-in and was blocked with one CSP by allowing mix and match services from different CSPs.[27]

3- **The framework's candidate** is an open-source project called OpenNebula which can support on-demand VMs provisioning, pre-configured, and manage groups of interconnected VMs; thus, OpenNebula enhances the integration of external providers (CSPs) to enable the selected model of deployment. OpenNebula is growing very fast to meet the industry and developer requirements.

Using OpenNebula gives the ability to manage the local infrastructure and establish the first step toward hybrid Cloud solution by interfacing with a remote Cloud site.

OpenNebula's main role is to manage the VMs.[30]

The following figure shows the architecture of OpenNebula:

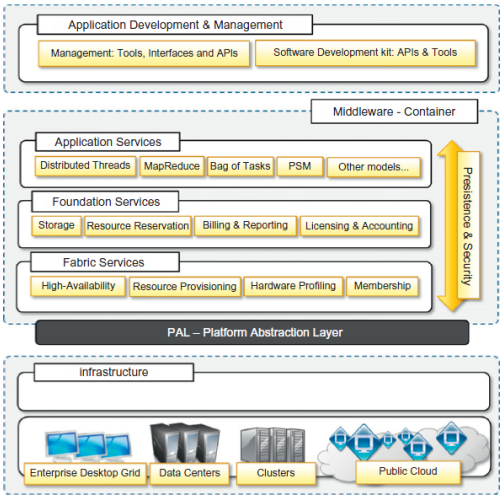


OpenNebula virtual infrastructure engine components[31]

4- **using Aneka** as a reference platform, It supports a collection of programming abstractions for developing applications and a distributed runtime environment that can be deployed on heterogeneous hardware (clusters, networked desktop computers, and cloud resources).[32]

Aneka provides software infrastructure for scaling applications using broad collection of APIs for the developers to design and implement applications. Aneka gives developers the ability to run their application on a local or remote distributed infrastructure which supports the hybrid

Cloud deployment model. The figure shows Aneka framework architecture:[32]



As shown in the above figure, the Aneka framework architecture contains three different layers corresponding to the basic service layers of the Cloud Computing easily integrated with the external Cloud. Aneka enables the execution of the application on its runtime environment by using the underlying Cloud infrastructure for either private or public Clouds. It provides management tools; administrators can easily start, stop, and deploy any application. The Aneka platform contains three classes of services which characterize its middleware layer:

1. Execution Services: Their primary responsibility is scheduling and executing deployed applications.

2. Foundation Services: They represent the collection set of management services, such as metering applications and resource allocation and updating the service registry whenever needed.

3. Fabric Services: They present the lowest level of middleware services classes. They provide access to Cloud resource management to enable resource provisioning which will scale the allocated resources to the applications to achieve the required QoS.[33]

5- <u>Layer of proposed cloud :</u>

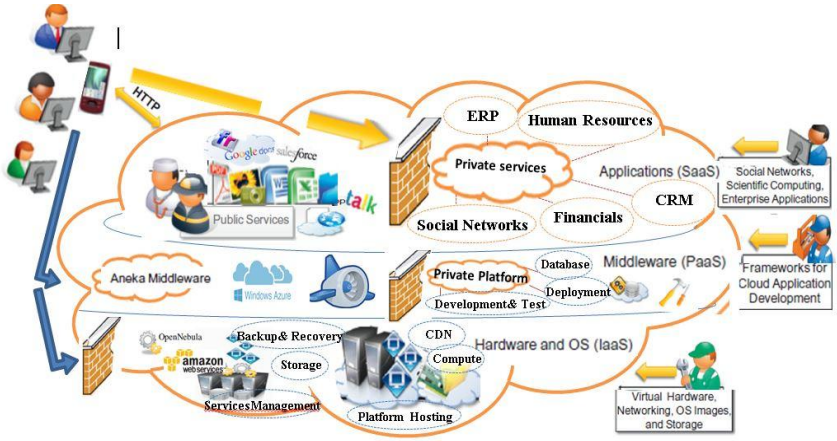
a) The SaaS layer: The SaaS layer represents the top layer of the cloud. It offers its services in a software model of web-based applications and serves a large number of users. User will connect to internet and has many services such as e-learning systems, admission process, research, faculty and student corners, admin, digital library, email , account and financial processes, student life cycle and information system, classroom

management system, and other administrative processes can be hosted on the SaaS layer for easy access and convenience for students, academic as well administrative staff of the university. With SaaS, the user does not need to install or run any applications on their devices or worry about maintenance, upgrades, and support. This layer divided into two sections that are the public cloud and private cloud, the most sensitive applications such as the admission process, account and financial services, and any other services considered critical and requires very high confidentiality should be hosted on a private cloud. The SaaS services on the public cloud can be provided by cloud vendors such as Salesforce.com, Google, and Sky-Drive. Using SaaS will help to overcome the problems of limited modern teaching and learning resources and lack of e-learning scalability.

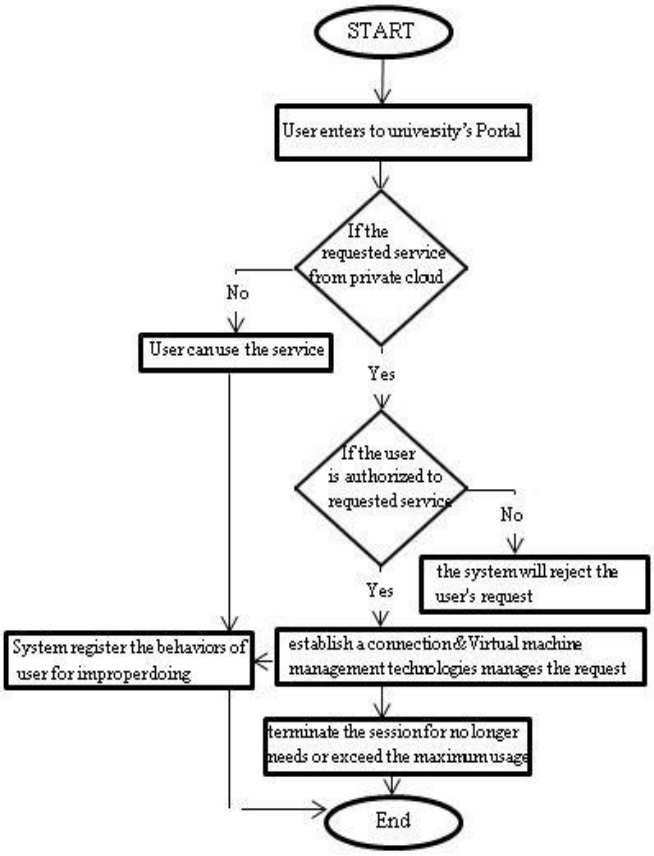
b) The PaaS layer: The PaaS layer is the middle layer between the SaaS and IaaS. Universities member can Develop, test, deploy, and manage applications hosted in The PaaS layer. The PaaS layer facilitates the development and deployment of applications without the developers worrying about Provisioning and management cloud infrastructure and middleware for the platform consumers and provides development, deployment, and administration tools to platform consumers because it is provided by the cloud vendor. This layer provides access to different platforms and programming languages, thereby making it possible for developers and programming students to easily do their work by simply connecting to the cloud. This layer also divided into two sections private and public cloud, many public services can be hosted on a public cloud using Force.com, Microsoft Azure, Google App Engine, and other and the private system to the university is required to authentication and control access.

c) The IaaS layer: It is the bottom layer. It is referred to everything as a service layer provides basic computing resources including servers, storage, hardware, and networking equipment. University Creates/installs, manages, and monitors services for IT infrastructure operations. The IaaS layer gives more flexibility when dealing with the Hardware layer but through the virtualization. This layer includes the operating system, bandwidth, and virtualization technology needed to manage the resources. Universities can deploy and run arbitrary software. The users of IaaS can be system developers, system administrators, and information technology(IT) managers who are interested in creating,

installing, managing and monitoring services for IT infrastructure operations, thereby this layer is sensitive need to access authentication.
One of the key players in the IaaS layer is Amazon E2C (Amazon Elastic Compute Cloud), it reduces the time needed to get and boot up new servers to minutes, it allows quick scaling capacity back and forth as demands change and provides a computing environment and resources that quickly and cost-effectively process large amounts of research data. Other providers of this service are HP, GoGrid, Rackspace, and so on.[19]



Proposed cloud computing structure schematic diagram



Cloud Computing Framework for Yemeni Universities

As shown in the tow figures above shows an overview of how the content could be in the Cloud:

1- The user sends a request using the University portal.
2- If the user request is for public service the user can enjoy with service.
3- If the user request is for private service, The verification of the authorization level will be checked using the user profile private Cloud.
4- If the user is unauthorized to request such services, the system will reject the user's request; otherwise, the request will be sent to the virtual infrastructure manager (OpenNebula) or Aneka to redirect the request to the appropriate service.
5- The system will establish a connection between the requested service from the Cloud and the user.
6- As long as the user needs the resource and does not exceed the maximum usage period, the system synchronizes the service delivery between the user and the resource.

7- When the user is done and no longer needs the requested resource, the system will terminate the session and disconnect the user from the target Cloud.

XIII. Conclusion

Cloud Computing paradigm is a new approach to produce a solution for old problems.

In a NY State CIO Conference in July 2009, the Vice President of EDUCAUSE, Richard N. Katz, spoke about "The Tower and the Cloud: Higher Education in the Age of Cloud Computing."[39]. He cited a poll that sensed how people felt about the role of cloud computing. In a nutshell, 74% of those polled thought that cloud services will have a great effect on higher education while 75% thought the same for their IT organization. [39]

It will become an essential source of e-learning in order to give the opportunity for students and teachers and the hole university department to quick access to various applications, systems and resources through the Internet, and share files and documents, and the exchange of duties and projects among students, a challenge that must overcome the comprehensive coverage of the service and fast Internet access , to enable the student to take advantage of this technology applications.

Cloud computing is an excellent alternative for educational institutions which are especially under budget shortage in order to operate their information systems effectively without spending any more capital for the computers, network devices and others.

From the points of advantages provided by cloud, there is a great advantage for university IT staff to take them away the responsibility of the maintenance burden in the university.

As any new technology develops and becomes popular, new issues have to be faced.

Cloud computing is not an exception, however Security, privacy and service-level agreements issues continue to be the biggest concern on cloud computing that limits its adoption in practice.

This condition poses challenges for confidential data, which organizations can't afford to reveal.

They continue to raise many questions for their CSPs such as where is their data located and who manages and accesses it, why is their personal information requested and who uses it and what is the fate of their data in case of disasters or when the CSP went out of business.

It is not surprising that much of the future work in cloud computing will focus on developing approach that are able to address its security issues.

It is very important to Make sure that the vendor has adequate staff that possesses legal and contracting skills to control your risk.

However, it is strongly recommended that early adopters plan the transition carefully and have a planning team in place to plan and oversee the cloud project , in order to ensure a uniform and smooth transition.

The research shows that hybrid cloud computing is a better choice for deployment in the universities since it gives the combined benefit of private and public clouds.

We have proposed roadmap and framework for adoption to cloud computing smoothly.

Acknowledgement:
I would like to thank 1st scientific Conference on IT& Network-November 1-2-2016 at Alandalus University for Science and Technology at Sana'a for their support in carrying out the research work and for allowing me this opportunity to explore my research.

IX. REFERENCES:

[1] Cearley,Cearley, “Cloud Computing Basics.” ,2014
[2]Bittman, T. Cloud Computing Inquiries at Gartner, 2009
[3] Pushparani Devi and others, Cloud Technology in Higher Teacher Education Institutions: A Solution to ICT Infraucture adoption problem,International Journal of Advanced Studies in Computer Science and Engineering, IJASCSE Volume 3, Issue 4, 2014.
[4] David, C. W. (2009). 2010-07 IBM Business of Goverment Cloud Computing in Goverment
[5] Alabbadi, M., 2011. Cloud computing for education and learning: Education and learning as a service (ELaaS). 14th International Conference on Interactive Collaborative Learning (ICL2011), (September).
[6] Buyya, R., Broberg, J. and Goscinski, A., n.d. Cloud Computing Principles and Paradigms.
[7]Motta, G., Sfondrini, N. and Sacco, D., 2012. Cloud computing: An architectural and technological overview. Proceedings - 2012 International Joint Conference on Service Sciences, Service Innovation in Emerging Economy: Cross-Disciplinary and Cross-Cultural Perspective, IJCSS 2012.
[8] Forbes, “ Cloud Computing's Hidden 'Green' Benefits,” 2011, [online] , the website is visited at 10-8-2016, Available at:
http://www.forbes.com/sites/joemckendrick/2011/10/03/cloud-computings-hidden-green-benefits/#605938645403
[10] Sumit Khurana, and Anmol Gaurav Verma, Dept. of CSE, Surya World, Punjab, India, Comparison of Cloud Computing Service Models: SaaS, PaaS, IaaS, IJECT Vol. 4, Issue Spl - 3, April - June 2013.
[11] Brad McMillan, Cloud Computing: What It Can Do for Your Business, Entrepreneurial Practice Review, Vol 2, No 1, 2011.
[12] Qi Zhang, and Lu Cheng, Cloud computing: state-of-theart and research challenges, Journal of Internet Services and Applications , 2010.
[13] Bhaskar Prasead Rimal, Eunmi Choi, and Lan Lumb, I., ATaxonomy and Survey of Cloud Computing Systems, Networked Computing and Advanced Information Management, Fifth International Joint Conference on INC, IMS and IDC, pp. 44-51, 2009
[14] DUSTIN AMRHEIN, P. A., ANDREW DE ANDRADE, JOE,ARMSTRONG, E. A. B., JAMES BARTLETT, RICHARD BRUKLIS,KEN CAMERON, et al. (2010): Cloud Computing Use Cases White Paper.Version 3.0 ed., Cloud Computing Use Case Discussion Group.
[15] Safiya Okai and others, 2014, Cloud Computing Adoption Model for Universities to Increase ICT Proficiency,
[16] 2013, white paper for partnership with Cisco and Intel®—Cloud Computing Survey Results:
http://www.govconnection.com/~/media/PDFs/Brands/C/Cisco/Survey/25240_PCC_CloudSurvey.pdf

[17] Cloud Security Alliance. (2010). Top threats to cloud computing V1.0. Retrieved from https://cloudsecurityalliance.org/topthreats/csathreats.v1.0.pdf
[18] Developing a Cloud-Computing Strategy for Higher Education, White Paper, 2012 Cisco and/or its affiliates
[19] Cloud Computing Adoption Model for Universities to Increase ICT Proficiency",2014, by Safiya and others
[20] Asbed Bedrossian and others,November 2014,CLOUD STRATEGY FOR HIGHER EDUCATION Building a Common Solution
[21]Foster, I., Zhao, Y., Raicu, I., & Lu, S. (2008). Cloud Computing and Grid Computing 360-Degree Compared. Grid Computing Environments Workshop (pp. 1 - 10). Austin: GCE.
[22] Sasikala, S., & Prema, S. (2010). Massive Centralized Cloud Computing (MCCC) Exploration in Higher Education. Advances in Computational Sciences and Technology , 3 (2), 111–118.
[23] Behrend, T. S., Wiebe, E. N., London, J. E., & Johnson, E. C. (2011). Cloud computing adoption and usage in community colleges. Behaviour & Information Technology , 30 (2), 231–240.
[24] IBM. (n.d.). Cloud Computing. Retrieved March 20, 2011, from IBM Academic Initiative, accessed at 10-8-2016: https://www.ibm.com/developerworks/university/cloud/
[25] Siegle, D. (2010). Cloud Computing: A Free Technology Option to Promote Collaborative Learning. Gifted Child Today , 33 (4), 41-45.
[26]Cloud Computing For Distributed University Campus: A Prototype Suggestion, Mehmet Fatih Erkoç, Serhat Bahadir Kert, Yildiz Technical University (Turkey),2013
[27] Eman A. Aldakheel, A CLOUD COMPUTING FRAMEWORK FOR COMPUTER SCIENCE EDUCATION, December 2011, MASTER OF SCIENCE, Bowling Green State University
[28] Cognizant Cloud Services "Harnessing Cloud Computing In Education",2014
[29] CLOUD COMPUTING ARCHITECTURE FOR HIGHER EDUCATION IN THE THIRD WORLD COUNTRIES (REPUBLIC OF THE SUDAN AS MODEL), Mohmed Sirelkhtem and others, International Journal of Database Management Systems (IJDMS) Vol.7, No.3, June 2015
[30] Sotomayor, B., Montero, R. S., Llorente, I. M. & Foster, I. (2009) 'Virtual Infrastructure Management in Private and Hybrid Clouds ', IEEE Internet Computing, September/October 2009, 13 no. 5.
[31] Peng, J., Zhang, X., Lei, Z., Zhang, B., Zhang, W. & Li, Q. (2009) 'Comparison of Several Cloud Computing Platforms ', Information Science & Engineering (ISISE), 2009 Second International Symposium on 26-28 Dec. (2009)
[32] Mastering Cloud Computing Foundations and Applications Programming, Rajkumar Buyya and others,2013

[33] A CLOUD COMPUTING FRAMEWORK FOR COMPUTER SCIENCE EDUCATION, Eman A. Aldakheel, Bowling Green State University, December 2011
[34] Abdul Razak, S.F., 2009. Cloud computing in malaysia universities. 2009 Innovative Technologies in Intelligent Systems and Industrial Applications, CITISIA 2009, (July), pp.101–106.
[35] Ketel, M., 2014. E-Ieaming in a Cloud Computing Environment.
[36] Sultan, N., 2010. Cloud computing for education: A new dawn? International Journal of Information Management, 30(2)
[37] Mathew, S., 2012. Implementation of Cloud Computing in Education - A Revolution. International Journal of Computer Theory and Engineering.
[38] An Overview of Cloud Services Adoption Challenges in Higher Education Institutions,Abdulrahman Alharthi and others,Electronics and Computer Science, University of Southampton, Southampton, United Kingdom,2013
[39] The Need for Strategic Planning Toward Adopting Cloud Computing in Higher Education, Samir Tout, Eastern Michigan University, Ypsilanti, MI, USA,2014
[40] SWINTON, L. (2004) How to do a SWOT analysis; CSM, http://www.customerservicemanager.com/swot-analysis.htm, 20/2/2010. The website is accessed at 10-8-2016

تعريف بالمؤتمر:

وهو مؤتمر علمي في مجال الحاسوب وتقنية المعلومات والشبكات نطمح ان يعقد سنوياً، ويهدف الى تشجيع الباحثين في مجال الحاسوب وتقنية المعلومات للالتقاء وتبادل الافكار العلمية والبحثية، وخلق نوع من التواصل بين الباحثين والموسسات ذات العلاقة. وكذلك منح فرصة لأبناءنا الطلاب لعرض ابداعاتهم من خلال عرض ابحاث التخرج المتميزة، وكذلك يمنح هذا المؤتمر فرصة لطلاب الدراسات العليا لعرض مشاكلهم البحثية امام نخبة من الباحثين والمتخصصين. و من خلال هذا المؤتمر تحقق كلية الهندسة وتقنية المعلومات جانباً من اهدافها متمثلاً في تشجيع البحث العلمي في الكلية.

الجمهورية اليمنية-صنعاء-دارسلم
هـــاتف: +967 1 675567
فاكـــس: +967 1 675885
صندوق بريد: 37444

جامعة الأندلس
للعلوم والتقنية
القيمة الكاملة للتعليم...

جامعة الأندلس للعلوم والتقنية
Alandalus University For Science & Technology
AUST