International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 19 (2017) pp. 8160-8173
© Research India Publications. http://www.ripublication.com

New System for Adaptive Information Retrieval Based on Fuzzy Sets

Yasser A. Nada[#1], Khaled M. Fouad[*2], Hossam Meshref[#3] and Y. M. Assem[#4]

[#]*Department of computer science, Faculty of Computer Science and Information Technology, Taif University, Kingdom of Saudi Arabia.*

[*]*Department of Information Systems, Faculty of Computers & Informatics, Benha University, Egypt.*

[#1]*Orcid: 0000-0002-5209-1181*

Abstract

Search engines are dealing with huge amount of diverse information in many different disciplines. Many retrieval systems attempt to maintain the personalization of the information retrieved through the used search engines. Unfortunately, there are still many problems such as weak concepts representation, inaccurate of retrieved information, and vague retrieval systems. In this research, the fuzzy theory will be used hence; the fuzzy user model can be created. The fuzzy user model contains fuzzy concepts and fuzzy memberships for the acquired user profile which provide an efficient representation for text in addition to adaptive information retrieval. During the main process of the proposed user model acquiring, the document is represented by using the keyword and key phrase extraction. The acquired fuzzy user model is represented in ontological format to be reusable. Such adaptation is achieved by using user model that is based on fuzzy theory to infer fuzzy user modelling aspects. Finally, many advantages are expected from the proposed system such as flexible and reliable retrieval system, overcoming of traditional search engine problems, and reducing of the information retrieval process time. The system evaluation exploits 206 queries, which are generated during five users are searching and browsing the system. Using evaluation metrics; such as precision, recall, fall-out and f-measure, the experiments prove, that the proposed approach improves the accuracy retrieved information rather than other traditional approaches retrieving information.

Keywords: Domain ontology, Semantic, Adaptive Information Retrieval, Fuzzy Sets.

INTRODUCTION

Information retrieval (IR) [1]is the important work for traffic information service in WWW. And ontology-based semantic information retrieval is a hotspot of current research. The ontology can be used as main resource to understand the textual information contained within the documents. An ontology is shared knowledge of a specific domain among people and systems. It is written in a specific language called a web ontology language (OWL). To achieve efficiency for the proposed ontology, a classic ontology was designed using Protégé OWL. A fuzzy OWL plug-in was then employed to convert the classic ontology into a fuzzy ontology. The fuzzy OWL plug-in is used to assert fuzzy terms in the ontology. The classes, instances, concepts, and axioms of a classic and a fuzzy ontology are the same. A classic ontology cannot handle uncertainty. A fuzzy ontology is generally defined to express vague knowledge using fuzzy concepts. Therefore, this system needs fuzzy ontology based semantic knowledge to handle any kind of situation related to sentiment analysis [2][3].

The ontology population generates the new instance and results in the semantic annotation of document [4]. Ontology is a conceptualization of a domain into a human understandable, machine-readable format consisting of entities, attributes, relationships, and axioms. It is used as a standard knowledge representation for the Semantic Web. The use of ontology to overcome the limitations of keyword-based search has been put forward as one of the motivations of the Semantic Web [5].

Web-based ontology can be seen as a part of implementing the "semantic web". The Stamford medical informatics group suggests their language - PROTEGE-2000 [6] as a means of representing ontology on the web. The central role of ontology in knowledge management is emphasized by [7].

Semantic information retrieval tries to go beyond traditional methods by defining the concepts in documents and in queries to improve retrieval. We saw that there is a current trend toward content based, or semantic, retrieval. In a similar manner, semantic based information retrieval is the next evolution of text IR [8].

One way to facilitate the identification of related concepts and their linguistic representatives given a key concept is to use ontology. They help to build document representations at a higher level of granularity, trying to describe the topical content of documents. These semantic ontology define relations between terms or concepts, or hypernym. However, these relations are generally strict ones, and can therefore be too limited for describing the real world, with all its ambiguity and vagueness, as previously mentioned. Moreover, they must interact with flexible queries, where the user express preferences. Therefore, fuzzy ontology in which relations between terms are weighted or encoded by different kinds of measures, may be used. Many other works use fuzzy logic in IR to solve the ambiguity and vagueness issues, by defining flexible queries or fuzzy indexes. However, the fuzzy ontology used here in ontology to represent fuzzy relational knowledge about words or phrases [9].

To handle uncertainty of information and knowledge, one possible solution is to incorporate fuzzy theory into ontology. Then we can generate fuzzy ontology, which contain fuzzy concepts and fuzzy memberships. The fuzzy ontology can deal with fuzzy knowledge [10], and are efficient in text and multimedia object representation and retrieval [11] [12].

In this paper, user profile plays an important role because of the rapid development of personalized information retrieval systems. Several modules will be applied after acquiring traditional user model. The most important modules in the system are building fuzzy user model. The input of these modules is the keyword (term) weight in the initial user interest generated from the system in ontological form. The generated fuzzy user-model based adaptive information retrieval IR system is developed. The proposed fuzzy user-model is applied to prune the out-of-date interest of user and acquire up-to-date user interest.

RELATED WORKS

A semantic information retrieval approach based on a rough ontology is proposed [13]. Rough ontology in this paper is in the form of an ontology information system. Given a keyword based query, our approach infers the individuals and properties correlated to the query through a procedure of association searches in the rough ontology, and takes properties as equivalence relations to construct an approximation space of rough ontology. Afterward, an algorithm of computing similarity in rough ontology is presented, and approximation space is employed to compute similarity for ranking documents in semantic document indexing space.

In [14] authors presented information retrieval system based on fuzzy ontology framework. The framework includes three parts: concepts, properties of concepts and values of properties, in which property's value can be either standard data type or linguistic values of fuzzy concepts. The framework is the extension of RDF data model "object-property-value", which is the current standard for establishing semantic interoperability on the Semantic Web. The semantic query expansion has been constructed by order relation, equivalence relation and inclusion relation between fuzzy concepts defined in fuzzy linguistic variable ontology, which facilitates the information retrieval at semantic level.

Authors [15] proposed an extended fuzzy ontology for supporting learning evaluation. To improve the accuracy of information retrieval, the proposed model adopted a Fuzzy concept semantic analysis for clustering to generate Learning Evaluation Ontology. Especially there are many search words are blurry in information retrieval, this system can achieve high retrieval accuracy and considerably improve efficiency as compared to common fuzzy ontology, which has been validated through some experiments.

This paper [16] has proposed two extensions for the current generation of digital libraries. First, authors proposed to use ontology to represent scholarly information in digital libraries, thus making the libraries enable to share and exchange knowledge in the Semantic Web environment. Second, fuzzy theory is employed to process uncertain scholarly information as the forms of fuzzy ontology and fuzzy queries. A general architecture of digital libraries in the Semantic Web environment has been presented and an experimental system has also been developed to verify their ideas and techniques.

To achieve fuzzy semantic retrieval for Electronic Commerce, this paper [17] proposes an approach using RDF and fuzzy ontology. Authors have applied RDF data model to represent e-commerce information on the Semantic Web. Then they have presented fuzzy linguistic variable ontology. Introducing new data type referred as fuzzy linguistic variables to RDF data model, the semantic query expansion in SPARQL query language is constructed by semantic relations between fuzzy concepts.

This paper [18] addressed the problem of implementation and query answering of fuzzy ontology on databases. Authors defined a language to create and to query fuzzy ontology databases. They distinguished between fuzzy concepts defined based on membership functions and fuzzy concepts defined using concept forming expressions. They proposed an inferential engine to infer instances and relations between fuzzy concepts defined using concept forming expressions. The fuzzy ontology model defined in this paper is not expressive enough. They project to extend it with other constructors such as existential restriction.

Authors [19] proposed an ontology-based information retrieval model to improve effectiveness of information retrieval. The ontology embedded in the proposal model is a fuzzy taxonomy generated automatically from the documents.

In [20], authors presented a method for constructing learner model that represents the user's interests by analyzing the web-log to extract the interested terms in visited pages by learner. Then the fuzzy clustering is used to extract clusters of output learner profiles the goal of incorporating the semantic web is to build the semantically enhanced user models. Fuzzy technique is used to cluster the extracted data of learner model to classifying the learners for their interests.

New algorithm R-FPC was proposed in [21] to perform fuzzy clustering for large document collections. Comparing with other soft subspace clustering algorithms, R-FPC can group the dataset into soft partitions using a new defined fuzzy membership measurement with non-parameters. We also have presented a new initialization method named R-Greedy that helps the algorithm to choose the robust initial conditions.

The authors in [22] proposed a method for clustering web documents that are represented by vectors of variable size. The fuzzy-based approach for defining the similarity between documents, does not at all consider the vector 's length. For each existing cluster, they first defined a cluster center (centroid) by averaging the vectors already classified in this cluster.

ONTOLOGICAL FORMAT AND XML

Extensible Markup Language (XML) is a set of rules for encoding documents in machine-readable form. The design goals of XML emphasize simplicity, generality, and usability over the Internet. It is a textual data format with strong support via Unicode for the languages of the world. XML is a specification by the World Wide Web Consortium (W3C) for document representations. It initially was developed to represent text documents. Text documents could be memos, letters, and papers. XML is a semi structured format for data with interesting tags. Tags are defined by tag sets called document type definitions (DTDs)[23].Although the design of XML focuses on documents, it is widely used for the representation of arbitrary data structures, for example in web services. Another area that is closely related to XML and Web databases is content management. The Applications of XML is shown in figure 1.

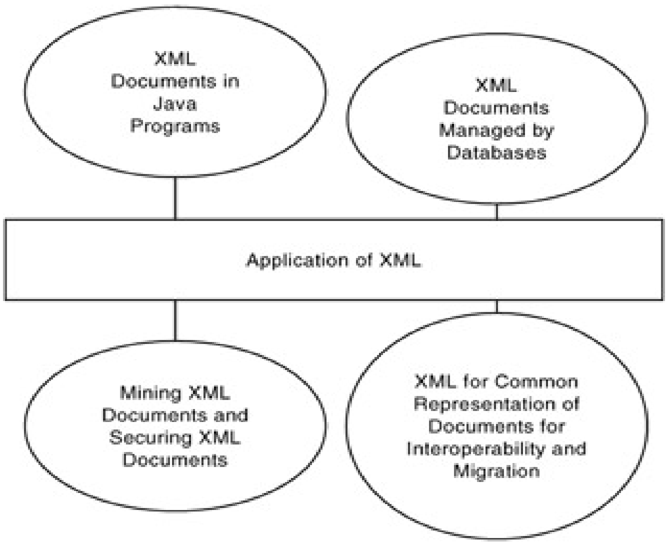

Figure 1: Applications of XML.

FUZZY IN ONTOLOGICAL FORMAT

The fuzzy ontology is based on modification of an existing crisp ontology. Currently there are ontology with an extremely rich set of relations between members, for example component parts of the UMLS have over 80 types of relations between ontology members, ranging from the simple "is a" to such specialized relations as "uniquely_mapped_to" and "developmental_form_of". By preserving these relations, an extremely rich set of relations can and form the framework of the ontology when beginning fuzzification. The modification is entirely incremental, conversion to a fuzzy ontology adds membership values to the currently existing relations, and may also add new entries, in the ontology. The ontology membership is normalized in respect to each of the terms in the ontology that is the sum of the membership value of each term in the ontology is equal to 1. This is because it is primarily concerned with mapping from queries to the ontology. This is justified on the basis that that for each term in a query, only one of the meanings will be required, and that these meanings are exclusive.

Zadeh originally introduced fuzzy logic in 1965 [24], in the context of set theory. He uses the concept of "membership" as an attribute of an item within an ontology. By use of the membership value, a fuzzy logic can be used, modifying the standard Boolean logic as used in classical information retrieval. To replace the "AND" term the fuzzy MIN term is used and to replace the "Or" term the fuzzy MAX term is used. Briefly, the MAX relation involves assigning the highest membership value found in the antecedents [25].

FUZZY ONTOLOGICAL FORMAT AND INFORMATION RETRIEVAL

Combining fuzzy domain ontology [26] with fuzzy linguistic variable ontology, we obtain the three-layered fuzzy ontology framework. The framework comprises the set of concepts and relations, set of properties and set of fuzzy linguistic variable ontology. Process for information retrieval [27] is based on the knowledge ontology, the semantic and concept research can be achieved. Especially, using linguistic value of fuzzy concept, we can construct the research pattern such as: SELECT instance of concept FROM Data source WHERE (property of concept) <comparison operator> "Linguistic value of fuzzy concept", in which the comparison operators includes: equal comparison (=), unequal (<>), less than or equal ($\leq$) and greater than or equal ($\geq$) etc.

Using the fuzzy formal concept analysis, the fuzzy ontology classes can be constructed automatically through the conversion from fuzzy concept lattice. The primitive fuzzy ontology is basis on the taxonomy relation. However, non-taxonomy relation is emerged widely in real world. So, the multiplicity concept relationships and more reasoning rules should be joined to specify the preliminary fuzzy ontology. Figure 2 shows the complete fuzzy ontology model generation framework [28].

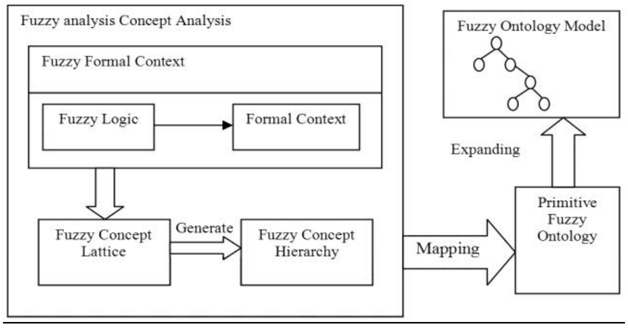

Figure 2: Fuzzy ontology generation framework.

Document clustering [29] is to partition a collection of documents into several clusters, such that the documents in the same cluster are as similar as possible, whereas those in any two different clusters are as dissimilar as possible. It has been applied to many fields such as data mining, information retrieval [30], topic detection and tracking. Usually, document clustering is performed in unsupervised fashion, i.e., only unlabeled documents are handled. However, in real application scenarios, it is often the case that the users have some background knowledge about the dataset, which could be useful in the process of clustering.

THE PROPOSED APPROACH

The research design and methodology must adequate to achieve an improved clustering performance for a given document collection and to improve retrieval performance for information search. In this system, a set of documents (X) are first selected from database. Each document (Xi) is an m-dimensional vector of m elements or m features. Since the m features in each document can have different units, in general, each of the features to a unified scale before clustering had normalized. This knowledge is then manipulated using inference to get a broader picture of the user's knowledge of the domain. We use fuzzy set theory for user model description: the user model representation is based on fuzzy sets and model updating on linguistic rules. Therefore, we must redefine the domain and user representations, considering also the uncertainty of knowledge description.

The approach comprises the set of concepts and relations, set of properties and set of fuzzy linguistic variable ontology. The relation between concept and property is "property of", and the relation between property and fuzzy linguistic variable ontology is "value of", in which property's value can be either standard data type or linguistic values of fuzzy concepts. The framework is the extension of RDF data model "object property- value", which is the current standard for establishing semantic interoperability on the Web. Since considering the essential semantic relationships between fuzzy concepts, the framework facilitates the information retrieval at semantic level. Figure 3 shows the complete architecture of our proposed system.

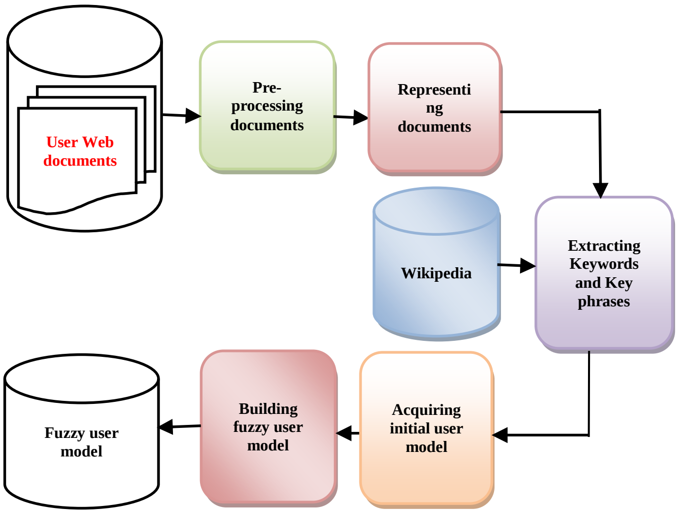

Figure 3: Architecture of our proposed system.

A. *Pre-processing documents*

The task of pre-processing the documents include preparing input Web document for the following processing. The pre-processing aims at partitioning sentences into separate lines and adjusting phrase boundary. The main objective of pre-processing is extracting the text from the Web documents, part of speech (PoS) tagging, removing stop words and stemming words.

1) *Text Extraction:* The first step of pre-processing documents is extracting text from the Web pages. This step is performed by scanning the contents of Web pages and identifying the HTML tags in each page. The tags are excluded; such as table, formatting and image tags (i.e. <HTML>, <BODY>, <IMG>, etc.). Also, all the scripts and codes are excluded; such as JavaScript and VBScript. Then, the text data is extracted from other tags; such as paragraphs, hyperlinks, and metadata tags. The individual text of documents is stored as input for the following steps. The open source high-performance PHP module is used to parse HTML [31].

2) *Part of Speech (PoS) Tagging:* The PoS tagger [32] scan the text and morphological structures to determine the appropriate part-of-speech. The words should be in their original order. PHP interface to Stanford POS Tagger to tag the tokens is used [33].Noun terms is selected for generating user model. To identify nouns, documents are tagged using part-of-speech tagger.

3) *Stop Words Removal:* Stop words are removed from the main text. The proposed approach exploits a fixed list of stop words with PoS data about all tokens. This step removes the words, which are not nouns, verbs or adjectives.

4) *Words Stemming:* The stem is a root-form of the word containing the same meaning and it can appear in various morphological forms. The proposed approach uses the

morphology function [34], which is provided with WordNet [35, 36] that can be used for stemming task.

B. *Representing Documents*

Information retrieval issues can be addressed by using the most common representation model of the documents, which can be called vector space model VSM [37]. VSM is applied in the current proposed approach to effectively achieve documents' representation. In VSM, the document is defined by n-dimensional feature vector. Each dimension corresponds to a specific term (word). Each term in a document vector assigned to calculated weight. The term weight can be calculated as a function of the term frequency (TF). Different weighting approaches may be applied by amending this function. Hence, a document j can be represented by the document vector d_j:

$d_j=(W_{1j}, W_{2j}, \ldots, W_{nj})$, where, W_{nj} is the weight of the n term in the document j.

D_j is a document, which was browsed, where j=1…., n. Term frequency TF_{jk} is the number of how many the specific term T_k found in a document D_j where k=1….,m. The calculation of the terms weight W_{jk} of each term T_k is found in [32, 37]:

$$W_{jk}= TF_{jk} \times idf_k \quad \ldots\ldots (1)$$

where the document frequency df_k is how many documents in the database, which contain the term T_k. The inverse document frequency idf_k is:

$$idf_k= \log_2 n - \log_2 df_k + 1 \ldots\ldots (2)$$

where (n) is the total number of documents in the system database. $tf \times idf$ is a mathematical algorithm [38], which is used to efficiently find key vocabulary that best represents the texts by applying the term frequency and the inverted document frequency together. $tf(T_k, D_j)$ is the term frequency of term T_k that appears in document D_j, and (n) is the total number of documents of the corpus. $df(T_k)$ is the number of the documents, which include the term T_k at least once and represents how often term T_k appears in other documents. $tf \times idf$ for Term T_k is defined as:

$$tf \times idf(T_k, D_j)= tf(T_k, D_j) \times \log(\frac{n}{df(T_k)}) \quad \ldots\ldots (3)$$

For vocabulary with a low or rare appearance frequency, the value of $tf \times idf$ is low, compared to that with a high appearance frequency, thus resulting words successfully classifying the documents. In the term selection process, a list of all terms contained in one document from the text collection is made. Then, the document selection process chooses term T_k that maximizes W(k), which is expressed as a vector for document D_j as follows. Document D_j includes $tf \times idf(T_k, D_j)$, which is $tf \times idf$ for the most appropriate term.

$$W(k) = \sum_{j=1}^{n} tf \times idf(T_k, D_j) \quad \ldots\ldots (4)$$

C. *Extracting Keywords and Key phrases*

Key phrase or keyword extraction [38] aims to achieve the most informative and important terms from specific documents, which represent the main topics of these documents that the user model will be acquired from it. In this section, unsupervised approach to extract key phrases from text, which is extracted from the documents. The proposed approach depends on statistical and linguistic features of terms in the documents. The key phrase extraction should have prepossessing steps such as; text pre-processing, Part-of-speech (POS) tagging, stemming. The key phrase extraction algorithm [39] includes three main steps; preparing dictionary of distinct entries, mapping dictionary entries with Wikipedia titles and ranking entries.

1) Preparing dictionary of distinct items: This step aims at building a hierarchical n-grams of distinct term and/ or co-occurrence with other terms in the input document with their frequencies. The algorithm utilizes LZ78 compression technique [40] for handling words instead of characters. The idea behind this algorithm is clear by considering the following example; let the following two sentences, which are achieved from the input document:

Sentence #1: *"*t7 t8 t9*t9 t4* t7 t8 t9* t4 t5* t6*"*

Sentence #2: *"* t4 t5 * t7 t8 t4 * t9 t6 *"*

Where t4, t5, t6, t7, t8 and t9 represent different terms, and "*" is the boundary of a phrase in a sentence. Table 1 illustrates, how the distinct items dictionary is achieved.

Table 1. The distinct terms and co-occurrence of words

Input sentences (Lines):

Sentence #1:"*t7 t8 t9*t9 t4* t7 t8 t9* t4 t5* t6*"

Sentence #2:"* t4 t5 * t7 t8 t4 * t9 t6 *"

Sequence	Pattern	Status	Action	Dictionary
t7 t8 t9	t7 t8	New	Add to dictionary	t7 t8
t7 t8 t9	t8 t9	New	Add to dictionary	t8 t9
t9 t4	t9 t4	New	Add to dictionary	t9 t4
t7 t8 t9	t7 t8	Exist	Increment frequency	-
t7 t8 t9	t7 t8 t9	New	Add to dictionary	t7 t8 t9
t4 t5	t4 t5	New	Add to dictionary	t4 t5
t6	t6	New	Add to dictionary	t6
t4 t5	t4 t5	Exist	Increment frequency	-
t7 t8 t4	t7 t8	Exist	Increment frequency	-
t7 t8 t4	t8 t4	New	Add to dictionary	t8 t4
t9 t6	t9 t6	New	Add to dictionary	t9 t6
-	-	Null	Quit	

The sentences in table 1are partitioned into sequences based on a boundary of a phrase "*". Each sequence is partitioned initially into patterns of bigram. If the pattern does not have an index in the dictionary, it should be added with a frequency value of "one"; otherwise the frequency of pattern is incremented by "one".

The entries in the dictionary are assigned by two different score values. The first value, which is the entry frequency score, depicts the occurrence time in the document. The second value is the entry influence weight, which is a frequency times calculated according to a grammatical rule by Kumar and Srinathan [41]. The grammatical rule favour noun phrases, which appear earlier or at the end of sentences. The later score is calculated according to the equation 5:

$$0 \leq p_0 < \frac{N_i}{2} \quad Or \quad p_0 > \left(\frac{3\ x\ N_i}{4}\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(5)}$$

Where N_i is a number of words in sentence i,p_0 is an index of first word in phrase p in the sentence.

2) Mapping Dictionary entries with Wikipedia titles: All indexed entries in the dictionary are mapped to Wikipedia titles. If there is matching between an entry and Wikipedia title(s), the entry will be assigned to confidence value equal to one, which means that the dictionary entry is considered as a Wikipedia concept; otherwise it will be assigned to value of zero.

3) Ranking Entries: There are a variety of key phrase ranking algorithms [42]. All ranking algorithms depend solely on key phrase frequency. There are other algorithms such as; n-gram filtration technique [41], which can calculate the influence of key phrase according to number of grammatical rules. The entries are ranked according to equation6 [39].

$$Rank(i) = log\left(p_i\ x\ \frac{TF_i+TI_i}{L} + CF_i\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(6)}$$

Where p_i is the position of dictionary entry *i*. The position is calculated as $p_i = (L - L_s)$, where L is a total number of lines in document andL_s is the first sentence, where a dictionary entry *i* occurs.TF_i,TI_i and CF_iindicate respectively the term frequency, influence weight, and Wikipedia confidence factor for dictionary entry *i*.

D. *Initial User Model Acquiring*

The main purpose of this step is to extract interested items in the web page, then get term frequency that reflects the importance of term. Finally get the weight of terms in the selected page. The output of this step is the weight of terms in selected page that can be used to build user model. Figure 4 shows certain steps to acquire user interest.

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 19 (2017) pp. 8160-8173
© Research India Publications. http://www.ripublication.com

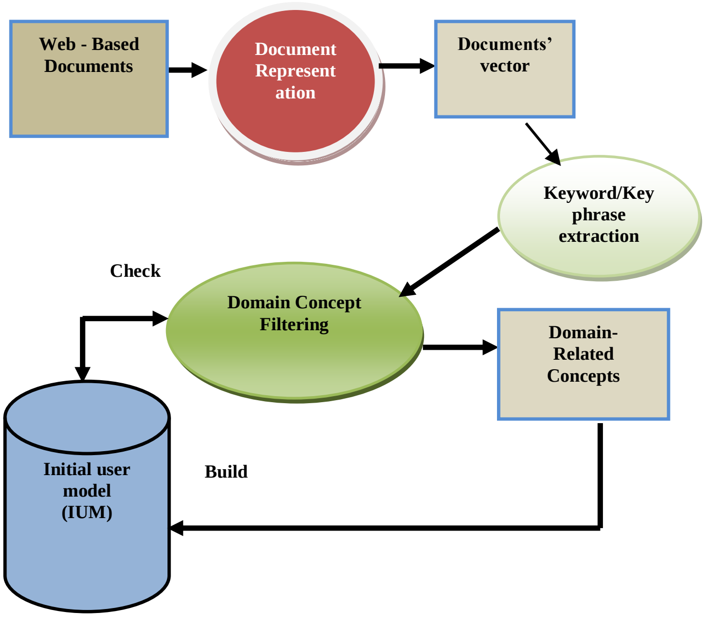

Figure 4: Steps to acquire to build user model.

1) User Model Creating: The user ontology is the specifications of the conceptualization of user information, which defines the various terms relationship related to the user concept, and gives the semantics of the term. User profile ontology mainly describe the interests of the user and display the different attributes and relationships of the user interests. It researches the ontology development process, the ontology life cycle, the methods and methodologies for building ontology, and the tools suite and languages that support them. Knowledge. It describes the software development process, the activities to be carried out, and the techniques that can be used for developing software[43].Figure 5 shows main steps of the ontology development process algorithm.

Input: The Environments, purpose, scope and its intended use.
Output: User model in Ontological format.
Procedures:
1. Analyze the environment.
2. Identify purpose and requirement specification// This identify the ontology purpose, scope and its intended use, that is the competence of the ontology
3. Ontology Acquisition:
  a. Enumerate important concepts and items.
  b. Define Concepts, slots and the relations of concepts.
  If there an Ontology is found
    Get the existing ontology
    Reuse The existing ontology
  Endif
4. Ontology Implementation:
  a. Open Hozo tool
  b. Create new file.
  c. Add the Concepts.
  d. Add relationship between the concepts.
  e. Save the file.
5. Check and Evaluate the created Ontology by the knowledge engineer.
  If there is an Error in the ontology
    a. Open the Ontology file
    b. Update it.
    c. Save the File
  ElseIf There is no Error
    Get out the Formal Ontology
  Endif

Figure 5: Algorithm of the ontology development process

The of purpose of identification and requirement specification concerns to clearly identify the ontology purpose, scope and its intended use, that is the competence of the ontology. Ontology acquisition is to capture the domain concepts based on the ontology competence. The relevant domain entities (e.g. concepts, relations, slots, and role) should be identified and organized into hierarchy structure. This phase involves three steps as follows: first, enumerate important concepts and terms in this domain; second, define concepts, properties and relations of concepts, and organize them into hierarchy structure; third, consider reusing existing ontology.

Ontology implementation aims to explicitly represent the conceptualization captured in a formal language. Evaluation/Check means that the ontology must be evaluated to check whether it satisfies the specification requirements. Documentation means that all the ontology development must be documented, including purposes, requirements, textual descriptions of the conceptualization, and the formal ontology.

E. *Building fuzzy user model*

Fuzzy Ontology, in this context, will be considered as a set of directed graphs where each node represents an item and the edges denote that a term "is related with" another term. A relatedness degree (RD) is associated with each edge to represent the strength of the "is related with" association. The fuzzy ontology is constructed by first calculating the relatedness degree (RD) for each pair of two distinct terms. Then, two tests are applied to select the association that will be incorporated into the ontology.

The most known method of fuzzy clustering is the fuzzy C-means method (FCM). Fuzzy C-means (FCM) is a data clustering technique wherein each data point belongs to a cluster to some degree that is specified by a membership grade while other classical clustering algorithms assign each data point to exactly one cluster. It provides a method that shows how to group data points that populate some multidimensional space into a specific number of different clusters. Most fuzzy clustering algorithms are objective function based: they determine an optimal (fuzzy) partition of a given data set C clusters by minimizing an objective function with some constraints. The most popular fuzzy clustering algorithm is the fuzzy c-means (FCM) algorithm. Even though it is better than the hard K-means algorithm at avoiding local minima, FCM can still converge to local minima of the squared error criterion. The design of membership functions is the most important problem in fuzzy clustering; different choices include those based on similarity decomposition and centroids of clusters. A generalization of the FCM algorithm has been proposed through a family of objective functions. The major difference between FCM and K-means is that FCM employs fuzzy partitioning such that a given data point can belong to several groups with the degree of belongingness specified by membership grades between 0 and 1. In FCM, the membership matrix U can have not only 0 and 1 but also the elements with any values between 0 and1.

FUZZY USER-MODEL BASED ADAPITIVE IR SYSTEM

A. *System Overview*

In this section, generated fuzzy user-model based adaptive information retrieval IR system is developed. The proposed fuzzy user-model is applied to prune the out-of-date interest of user and acquire up-to-date user interest. figure6 shows the web-based GUI of the adaptive IR system. The developed system was implemented by PHP language to be Web based IR system in the commerce domain.

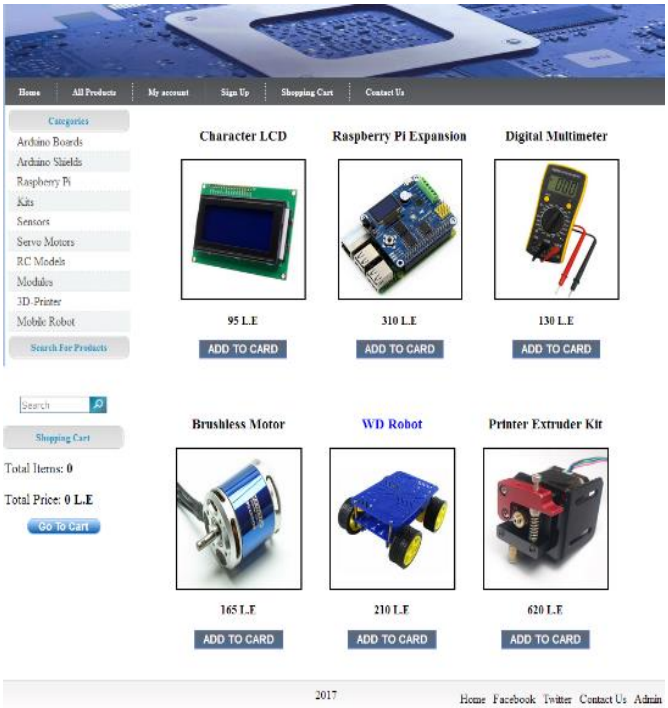

Figure 6: Web-based GUI of the adaptive IR system

For each user, based on his background rating records for browsing the pages in the system, user model is acquired by user interest acquiring module. Several modules will be applied after acquiring traditional user model. The most important modules in the system are building fuzzy user model. The input of these modules is the keyword (term) weight in the initial user interest generated from the system in XML, which is shown in figure 7. The input XML includes the weight of each keyword (term) in the browsed documents in the system.


```
<Key_Word>
<name>Arduino_Shields</name>
<Frequency>6</Frequency>
<tf>0.063157894736842</tf>
<idf>2.7080502011022</idf>
<weight>0.1710347495433</weight>
</Key_Word>
<Key_Word>
<name>Sensors </name>
<Frequency>6</Frequency>
<tf>0.063157894736842</tf>
<idf>2.7080502011022</idf>
<weight>0.1710347495433</weight>
</Key_Word>
<Key_Word>
<name>Raspery_Pi </name>
<Frequency>6</Frequency>
<tf>0.063157894736842</tf>
<idf>2.7080502011022</idf>
<weight>0.1710347495433</weight>
</Key_Word>
<Key_Word>
<name>Arduino_Boards</name>
<Frequency>6</Frequency>
<tf>0.063157894736842</tf>
<idf>2.7080502011022</idf>
<weight>0.1710347495433</weight>
</Key_Word>
<Key_Word>
<name>Servo_Motors</name>
<Frequency>6</Frequency>
<tf>0.063157894736842</tf>
<idf>2.7080502011022</idf>
<weight>0.1710347495433</weight>
</Key_Word>
```


Figure 7: The weights of the user in initial user model in ontological format

These modules perform the general steps of fuzzification, fuzzy inference rules and defuzzification. The fuzzification is applied on the weight of each interest items in the user model. Figure 8 shows the output of the fuzzification module in XML format.


```
<fuzzy>
<fuzzy word="arduino" Weight="5">
<fuzzy_value>low</fuzzy_value>
<ratio>50%</ratio>s
</fuzzy>
<fuzzy word="kits" Weight="7">
<fuzzy_value>High</fuzzy_value>
<ratio>70%</ratio>
</fuzzy>
<fuzzy word="sensor" Weight="9">
<fuzzy_value>Very High</fuzzy_value>
<ratio>90%</ratio>
</fuzzy>
<fuzzy word="shell" Weight="2">
<fuzzy_value>Very low</fuzzy_value>
<ratio>20%</ratio>
</fuzzy>
<fuzzy word="rasperypi" Weight="10">
<fuzzy_value>Very High</fuzzy_value>
<ratio>100%</ratio>
</fuzzy>
</fuzzy>
```


Figure 8: The output of fuzzification module

The fuzzy inference module aims to exploit the extracted rule for each fuzzy term to build the fuzzy inference. Figure 9 shows the output of this module. Figure 10 shows the output of the defuzzification module.

```
<Users>
<User>
<name value="Amgad ali hassan" interset="electronics,murano"></name>
<Interest degree="Low" value="weight less than 0.25" Keywords="amazon,
this,safavieh,for,silverado,swarovski,jewelry,mom,birthstone,christmas,crystals,rin
gs,anniversary,grandmother,ring,valentines,to,day,mother,perfect,avoid,egypt,ship
s,imported,cotton">Interest</Interest>
<Interest degree="Medium" value="weight between 0.25 &amp; 0.5"
Keywords="book,south,america,colombia,florida,davis,award,jammer">Interest</I
nterest>
<Interest degree="High" value="weight between 0.5 &amp; 0.75"
Keywords="">Interest</Interest>
<Interest degree="Very High" value="weight greater than 0.75"
Keywords="electronics,murano">Interest</Interest>
</User>
<User>
<name value="fahmy nazmy" interset="electronics,murano"></name>
<Interest degree="Low" value="weight less than 0.25"
Keywords="to,accelerometer,contentsony,ion,mpsoundalert,memoryram,gbcard,g
binternal,cpuocta,li,push,documentation,headphones,charger,disclaimer,see,emailb
atterybattery,compact,mobile,snapdragon,simcpuchipsetqualcomm,colors,mtouch,f
eaturesflashsecondary,gbdisplayscreen,typeips,specificationsdevice,density,size,m
pprimary,photosprimary,simsim,phone,typemobile,typenano,up,screenyes,multitou
chvideo,email,storage,connectivitybluetoothv,marshmallow,system,ltesoftwareope
rating,document,direct,fiwi,photo,glonasswireless,wi,fi,camera,amazon,and,game,
play,android,find,plus,fire,controller,full,minecraft,pocket,edition,asphalt,featuring
,kick,integrated,wireless,paid,affordable,compatible,book,channel,identifiers,fanta
sy,browse,gametrackers,follow,your,easily,ring,crystals,jewelry,christmas,silverad
o,valentines,anniversary,birthstone,swarovski,mom,rings,grandmother,of,tablets,p
hones,sports,day,the,gilead,puritans,genesis,states,republic,united,tale,handmaid,m
other,safavieh,for,this">Interest</Interest>
<Interest degree="Medium" value="weight between 0.25 &amp; 0.5"
Keywords="height,pile,japanese,rich">Interest</Interest>
<Interest degree="High" value="weight between 0.5 &amp; 0.75"
Keywords="">Interest</Interest>
<Interest degree="Very High" value="weight greater than 0.75"
Keywords="electronics,murano">Interest</Interest>
</User>
</Users>
```


Figure 9: The output of fuzzy rules


```
<Users>
<User>
<name>Amgad ali hassan</name>
<ID>3</ID>
<keyword>amazon</keyword>
<weight>0.11836619462528</weight>
<Mu>0.19</Mu>
<keyword>this</keyword>
<weight>0.13553161367866</weight>
<Mu>0.06</Mu>
<keyword>safavieh</keyword>
<weight>0.14412876827768</weight>
<Mu>0.09</Mu>
<keyword>for</keyword>
<weight>0.14412876827768</weight>
<Mu>0.23</Mu>
<keyword>silverado</keyword>
<weight>0.1710347495433</weight>
<Mu>0.05</Mu>
<keyword>swarovski</keyword>
<weight>0.1710347495433</weight>
<Mu>0.03</Mu>
<keyword>jewelry</keyword>
<weight>0.1710347495433</weight>
<Mu>0</Mu>
</User>
</Users>
```


Figure 10: The output of defuzzification

B. *Experiment*

In this section, the experiments are performed to evaluate the performance of the user model for the adaptive IR system based on fuzzy sets and keyword extraction. The system evaluation includes two parts. The first one is about the user model based on fuzzy set. The second experiment aims to evaluate the user model based on fuzzy set and keyword extraction. The results of each experiment show that the proposed approach can improve the performance of adaptive IR system in commerce systems.

1) *Experiment Setup:* The experiment is performed by using set of data to be used in the evaluation section.

a) Data set: The experiment is implemented on dataset of 300 documents of electronically devices and components for commerce system. It includes 5users for searching and browsing the system. Each user browsed the system for number of sessions to acquire his model, then he used number of different queries to search the specific component. Table 2 shows the main characteristic of the dataset.

Table 2: The Experimental Dataset

User	No of query	Total number of documents
User 1	**39**	**300**
User 2	**50**	
User 3	**54**	
User 4	**42**	
User 5	**21**	
Total number of queries	**206**	

b) Evaluation metric: After a test experiments have been finalized, at any session a user may submit a query derived from one of his interested documents to the proposed IR system, the outcome list of retrieved documents is obtained. The measures of the system's effectiveness are performed using the relevance judgments for that documents. The IR literature is overcrowded with different evaluation measures to evaluate different aspects of retrieval performance. Precision (P), Recall (R), Fall-out and F-measure [44] are standard evaluation metrics used in information retrieval research. The comparative performance evaluation is carried out by comparing the values of P, R, Fall-out and F-measure for the proposed IR approach and other approaches.

2) Evaluatinguser model based on fuzzy set: By considering the history data of 5 users sessions for two approaches, the difference between the user model based on fuzzy set and traditional user model is evaluated. The evaluation metrics; Precision (P), Recall (R), Fall-out and F-measure for the two approaches are performed. Table 3 shows sample of the evaluation between the two approaches; user model based on fuzzy set and traditional user model using the four metrics during 5 users were using the system.

3) Evaluating user model based on fuzzy set and keyword extraction: By evaluating the proposed approach, which is acquiring user model based on fuzzy set and keyword extraction, the comparison of the four metrics is applied between the two approaches; user model based on fuzzy set and user model based on fuzzy set and keyword extraction. Table 4 shows sample of the evaluation between the two approaches using the evaluation metrics; Precision (P), Recall (R), Fall-out and F-measure.

4) Experiments' Results: From the previous comparison, the proposed system, which is using user model based on fuzzy sets and keyword extraction can improve the information retrieval performance by increasing the metrics; precession, recall and f-measure. The user model based on fuzzy set improves the precession by the ration 19.36% against traditional user model. The user model based on fuzzy set and keyword extraction improves the precession by the ration 15.87%against user model based fuzzy set. The user model based on fuzzy set improves the recall by the ration 15.73% against traditional user model. The user model based on fuzzy set and keyword extraction improves the recall by the ration 14.81%against user model based fuzzy set. The user model based on fuzzy set improves the f-measure by the ration 6.83% against traditional user model. The user model based on fuzzy set and keyword extraction improves the f-measure by the ration 17.87 % against user model based fuzzy set. The figures 11, 12 and 13 show the difference between the three approaches for each metric. Table 5 shows the meaning of the labels used in figures 11 and 12.

Table 3: Sample of the evaluation between the two approaches

		With user model				With Fuzzy user model			
Query	**User**	**Precision**	**Recall**	**Fall-out**	**F-measure**	**Precision**	**Recall**	**Fall-out**	**F-measure**
Cameras	User 1	0.04	0.30	0.23	0.07	0.05	0.40	0.23	0.09
Capacitors	User 1	0.04	0.50	0.24	0.07	0.05	0.67	0.24	0.10
Cases	User 1	0.09	0.38	0.15	0.15	0.13	0.54	0.14	0.21
Casters	User 1	0.06	0.25	0.13	0.10	0.04	0.17	0.13	0.06
Cell Phone Accessories	User 1	0.03	0.20	0.17	0.06	0.05	0.30	0.17	0.09
Gauges, Automotive	User 2	0.01	0.05	0.25	0.02	0.07	0.30	0.24	0.11
Gears	User 2	0.07	0.29	0.24	0.11	0.09	0.38	0.23	0.15
Handles	User 2	0.08	0.29	0.20	0.12	0.17	0.62	0.20	0.26
Hardware	User 2	0.04	0.17	0.21	0.06	0.08	0.33	0.20	0.13
Headers	User 2	0.06	0.20	0.14	0.09	0.19	0.67	0.13	0.29
Switches - Pushbutton	User 3	0.01	0.04	0.50	0.01	0.05	0.28	0.50	0.08
Switches - Reed	User 3	0.10	0.50	0.39	0.17	0.05	0.21	0.29	0.08
Switches - Rocker	User 3	0.08	0.23	0.20	0.12	0.14	0.52	0.32	0.22
Switches - Rotary	User 3	0.11	0.32	0.19	0.16	0.11	0.47	0.45	0.18
Switches - Slide	User 3	0.13	0.40	0.16	0.19	0.22	0.32	0.05	0.26
Tape	User 4	0.07	0.35	0.46	0.12	0.32	0.63	0.13	0.42
Telecom / Networking Accessories	User 4	0.18	0.19	0.01	0.19	0.07	0.37	0.40	0.12
Telephone (Cell Phone)	User 4	0.05	0.23	0.57	0.09	0.04	0.15	0.19	0.06
Terminal Strips	User 4	0.12	0.35	0.17	0.17	0.11	0.54	0.35	0.19
Terminals, Crimp	User 4	0.01	0.02	0.54	0.01	0.11	0.60	0.31	0.19
Enhanced Mid-Range 8-bit	User 5	0.11	0.45	0.35	0.17	0.08	0.43	0.35	0.14
IC Sensors	User 5	0.03	0.18	0.40	0.05	0.09	0.52	0.39	0.15
Zilog Z84C0006PEG Z80 CPU	User 5	0.08	0.26	0.40	0.12	0.17	0.66	0.39	0.27
Operational Amplifiers	User 5	0.07	0.25	0.28	0.11	0.12	0.27	0.20	0.16
Logic circuits	User 5	0.09	0.17	0.17	0.12	0.24	0.49	0.15	0.33

Table 4: Sample of the evaluation between the two approaches

		With Fuzzy user model				With Fuzzy user model & Keyword			
Query	**User**	**Precision**	**Recall**	**Fall-out**	**F-measure**	**Precision**	**Recall**	**Fall-out**	**F-measure**
Cameras	**User 1**	**0.06**	**0.50**	**0.23**	**0.11**	**0.21**	**0.80**	**0.22**	**0.33**
Capacitors	**User 1**	**0.05**	**0.67**	**0.24**	**0.10**	**0.14**	**0.89**	**0.23**	**0.24**
Cases	**User 1**	**0.15**	**0.62**	**0.14**	**0.24**	**0.22**	**0.47**	**0.13**	**0.30**
Casters	**User 1**	**0.14**	**0.58**	**0.13**	**0.23**	**0.32**	**0.67**	**0.11**	**0.43**
Fuse Holders	**User 2**	**0.09**	**0.58**	**0.22**	**0.16**	**0.16**	**0.56**	**0.21**	**0.24**
Fuses	**User 2**	**0.11**	**0.50**	**0.21**	**0.18**	**0.22**	**0.50**	**0.20**	**0.30**
Gauges, Automotive	**User 2**	**0.08**	**0.35**	**0.24**	**0.13**	**0.24**	**0.64**	**0.24**	**0.35**
Gears	**User 2**	**0.13**	**0.52**	**0.23**	**0.20**	**0.20**	**0.46**	**0.22**	**0.28**
Switches - Slide	**User 3**	**0.02**	**0.05**	**0.18**	**0.03**	**0.83**	**1.00**	**0.12**	**0.91**
Switches - Snap Action	**User 3**	**0.09**	**0.32**	**0.38**	**0.14**	**0.63**	**0.88**	**0.14**	**0.73**
Switches - Tilt, Mercury	**User 3**	**0.22**	**0.35**	**0.07**	**0.27**	**0.51**	**0.60**	**0.14**	**0.55**
Switches - Toggle	**User 3**	**0.05**	**0.23**	**0.36**	**0.08**	**0.48**	**0.70**	**0.21**	**0.57**
Microphone	**User 4**	**0.11**	**0.33**	**0.38**	**0.17**	**0.40**	**0.64**	**0.33**	**0.49**
Thermal Switches	**User 4**	**0.09**	**0.53**	**0.54**	**0.15**	**0.23**	**0.67**	**0.52**	**0.35**
Thermal Sensor	**User 4**	**0.10**	**0.48**	**0.57**	**0.17**	**0.13**	**0.45**	**0.57**	**0.20**
Pressure Switches	**User 4**	**0.08**	**0.26**	**0.37**	**0.12**	**0.21**	**0.39**	**0.35**	**0.28**
IC Sensors	**User 5**	**0.12**	**0.36**	**0.33**	**0.18**	**0.13**	**0.45**	**0.39**	**0.20**
Zilog Z84C0006PEG Z80 CPU	**User 5**	**0.10**	**0.44**	**0.40**	**0.17**	**0.13**	**0.33**	**0.40**	**0.18**
Operational Amplifiers	**User 5**	**0.11**	**0.38**	**0.25**	**0.17**	**0.28**	**0.79**	**0.28**	**0.42**
Logic circuits	**User 5**	**0.07**	**0.29**	**0.24**	**0.11**	**0.30**	**0.51**	**0.20**	**0.38**

Table 5: Meaning of the labels used in figures 11 and 12

Label	Meaning
Precision 2-1	Difference between precision of user model based on fuzzy set and traditional user model
Precision 3-2	Difference between precision of user model based on fuzzy set and keyword extraction and user model based on fuzzy set
Recall 2-1	Difference between recall of user model based on fuzzy set and traditional user model
Recall 3-2	Difference between recall of user model based on fuzzy set and keyword extraction and user model based on fuzzy set
F-Measure 2-1	Difference between f-measure of user model based on fuzzy set and traditional user model
F-Measure 3-2	Difference between f-measure of user model based on fuzzy set and keyword extraction and user model based on fuzzy set

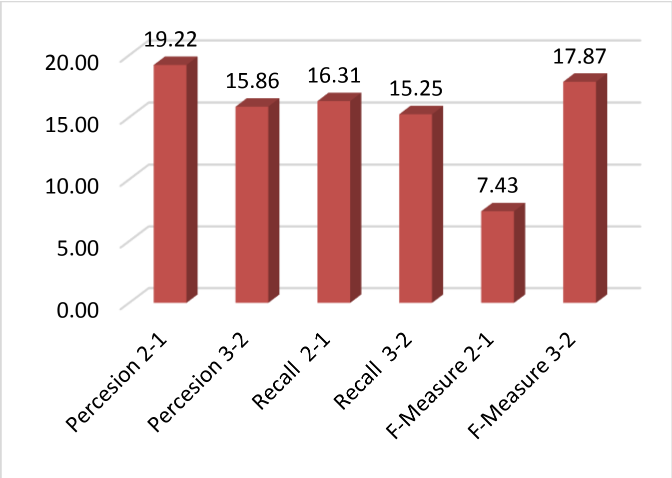

Figure 11: Difference of ration between proposed approach and others.

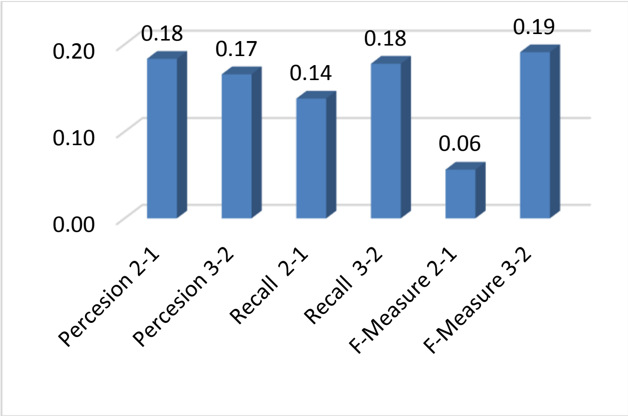

Figure 12: Difference of average between proposed approach and others.

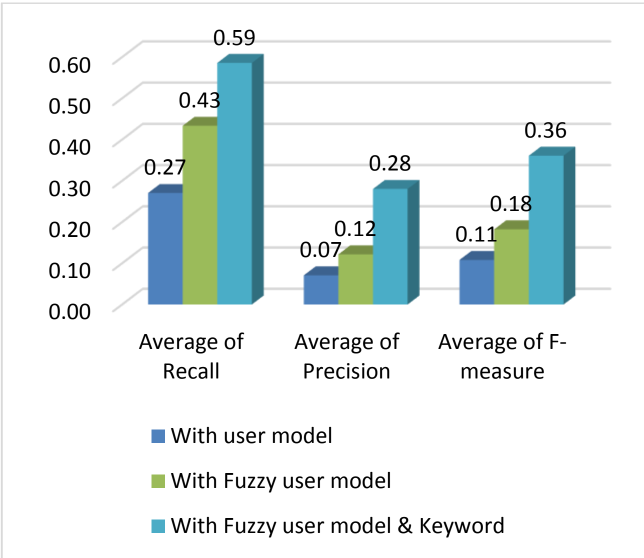

Figure 13: Comparison of the three approaches for the average of each Metric.

CONCLUSION

Search engines dealt with huge amount of diverse information in many different disciplines. Many retrieval systems attempted to maintain the personalization of the information retrieved through the used search engines. Unfortunately, there were still many problems such as weak concepts representation, inaccurate of retrieved information, and vague retrieval systems. In this paper, fuzzy user model is generated and represented in ontological format, in addition to developing of an adaptive information retrieval IR system. The experiments are performed to evaluate the performance of the fuzzy user model for the adaptive IR system based on fuzzy sets and keyword extraction. The system evaluation includes two parts. The first one is about the user model based on fuzzy set. The second experiment aims to evaluate the user model based on fuzzy set and keyword extraction. The results of each experiment show that the proposed approach can improve the performance of adaptive IR system in commerce systems.The fuzzy user model comprises fuzzy concepts and fuzzy memberships for the user profile, which provides an efficient representation for data, in addition to adaptive information retrieval. Such adaptation was achieved by using a model that was based on fuzzy theory to infer user modeling aspects. Finally, many advantages were expected from the proposed system such as flexible and reliable retrieval system, overcoming of traditional search engine problems, and reducing the information retrieval process time.

REFERENCES

[1] M. Ramachandra. (2010). Information Retrieval. In: Web-BasedSupply Chain Management and Digital Signal Processing: Methodsfor Effective Information Administration and Transmission . 182-194 pp. IGI Global.

[2] Farman Ali , Daehan Kwak , Pervez Khan , S. M. Riazul Islam , Kye Hyun Kim , K. S. Kwak (2017).Fuzzy Ontology-based Sentiment Analysis of Transportation and City Feature Reviews for Safe Traveling. ArXiv.

[3] Bobillo, F. and Straccia, U., (2011). Fuzzy ontology representation using OWL 2. International Journal of Approximate Reasoning, 52(7), pp.1073-1094.

[4] Sridevi U K . (2014) . Ontology based optimization techniques for information retrieval. Shodhganga: a reservoir of Indian Theses.

[5] Z. Jun, C. Yan, and C. Yan. (2008). Semantic Information Retrieval Based on Fuzzy Ontology for Intelligent transportation Systems. 1-4244-2384-2/08, IEEE.

[6] N. Noy, et, (2001) . Creating Sematic Web contents with Protégé -2000, IEEE Intelligent Systems 16 (2) (March/April), 60-7 1.

[7] S. Staab, A. Maedche, (2001). Knowledge portals: Ontologieast work, A1 Magazine 22, 63-85.

[8] R. Fuji. (2009). Advanced Information Retrieval. Electronic Notes in TheoreticalComputer Science 225 (2009) 303–317. Elsevier B.V.

[9] B. Mustapha, B. Mohand, L. Yannick, and P. Henri. (2007). Fuzzy Logic and Ontology-based Information Retrieval. In: Studies in Fuzziness and Soft Computing, 2007, Volume 215/2007, 193-218, DOI: 10.1007/978-3-540-71258-9_10.

[10] Widyantoro D. H., Yen J.(2001). A fuzzy Ontology-based Abstract Search Engine and its User Studies. Proceedings of the 10th IEEE International Conference on Fuzzy Systems, Melbourne, Australia, , pp. 1291-1294.

[11] John W.T. Lee, Alex K.S. Wong. (2004). "Information retrieval based on semantic query on RDF annotated resources". in: Proceedings of the 2004 IEEE International Conference on Systems, Man and Cybernetics, pp. 3220-3225.

[12] Parry D.,(2004) . A Fuzzy Ontology for Medical Document Retrieval", Proceedings of the Second Workshop on Australasian Information Security, Data Mining and Web Intelligence, and Software Internationalization, Dunedin, New Zealand, pp. 121-126.

[13] Jun Hu, Xinzhou Lu and Chun Guan, (2014). A Semantic Information Retrieval Approach Based on Rough Ontology . The Open Cybernetics & Systemics Journal, 2014, 8, 399-404.

[14] Z. Jun, W. Qinglian, and L. Miao. (2008). Application of Fuzzy Ontology Framework to Information Retrieval for SCM. 2008 International Symposiums on Information Processing. 978-0-7695-3151-9/08, IEEE.

[15] W. Qing, C. Wei, W. Bin. (2009). Fuzzy ontology generation model using fuzzy clustering for learning evaluation. Granular Computing, 2009, GRC '09. IEEE International Conference. 682 – 685. IEEE.

[16] Q. Tho, H. Siu, and C. Tru . (2007). Ontology-Based Fuzzy Retrieval for Digital Library. ICADL 2007, LNCS 4822, pp. 95–98. Springer-Verlag Berlin Heidelberg.

[17] Z. Jun, S. Yamei. (2010). Semantic Retrieval Based on SPARQL and Fuzzy Ontology for Electronic Commerce. 2010 International Conference of Information Science and Management Engineering. 978-0-7695-4132-7/10, IEEE.

[18] B. Afef, B. Rafik, G. Faıez. (2009). Fuzzy Ontology Implementation and QueryAnswering on Databases. The 28th North American Fuzzy Information Processing Society Annual Conference (NAFIPS2009) Cincinnati, Ohio, USA - June 14 - 17, 978-1-4244-4577-6/09, IEEE.

[19] C. Been-chian, H. Chih-hung, j. Ming-yi. (2007). IntelligentInformation Retrieval Applying Automatic Constructed Fuzzy Ontology. Proceedings of the Sixth International Conference on Machine Learning and Cybernetics, Hong Kong, 19-22 August 2007. 1-4244-0973-X/07, IEEE.

[20] F. Khaled, H. Mofreh, G. Shehab, M. Nagdy. (2010). Adaptive E-Learning System based on Semantic Web and Fuzzy Clustering. (IJCSIS) International Journal of Computer Science and Information Security,Vol. 8, No. 9, December 2010.

[21] C. Lifei, W. Shengrui, J. Qingshan. (2009). A Robust Algorithm for Fuzzy Document Clustering. 2009 International Conference on Advanced Information Networking and Applications Workshops. 978-0-7695-3639-2/09, IEEE.

[22] F. Menahem, K. Abraham, S. Moti, L. Mark, S. Bracha, E. Yuval and Z. Omer. (2004). A Fuzzy-Based Algorithm for Web DocumentClustering. 0-7803-8376-1/04, IEEE.

[23] Bhavani Thuraisingham.(2002). XML databases and the semantic web. CRC PRESS Boca Raton London New York Washington , D.C.Library of Congress Cataloging -in-Publication Datan.ISBN 0-8493-1031 -8 (alk. paper).

[24] Zadeh, L. (1965). Fuzzy Sets. Journal of Information and Control 8: 338-353.

[25] P. David, (2004). A fuzzy ontology for medical document retrieval. Appeared at The Australasian Workshop on DataMining and Web Intelligence (DMWI2004), Dunedin.Conferences in Research and Practice in InformationTechnology, Vol. 32.

[26] Z. Jun,C. Yan, and C.Yan,(2008). Semantic Information Retrieval Based on Fuzzy Ontology for Intelligent Transportation Systems. 1-4244-2384-2/08, IEEE.

[27] Bordogna, G. and G. Pasi (2000). Application of Fuzzy Set Theory to extend Boolean Information Retrieval. Soft Computing in Information Retrieval: Techniques and Applications. F. Crestani and G. Pasi. Heidelberg,Physica-Verlag: 21-47.

[28] C. Wei , Y. Qing,Z. Li, W. Bin, (2009). Research on Automatic Fuzzy Ontology Generation from Fuzzy Context. 978-0-7695-3804-4/09, IEEE .

[29] H. Guobiao, Z. Shuigeng, G. Jihong, H. Xiaohua. (2008). Towards effective document clustering: A constrained K-means based approach. Information Processing and Management 44 (2008) 1397–1409, Elsevier Ltd.

[30] W. Thaung, M. Lin. (2010). Document Clustering by Fuzzy C-Mean Algorithm. 978-1-4244-5848-6110, IEEE.

[31] An HTML DOM parser. It allows you to manipulate HTML. Find tags on an HTML page with selectors just like jQuery. https://github.com/paquettg/php-html-parser, 10-05-2017.

[32] Fouad, K., Khalifa, A., Nagdy, N., Harb, H..(2012) Web-based Semantic and Personalized Information Retrieval. IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 3, No 3.

[33] PHP interface to Stanford NLP tools (POS Tagger, NER, Parser), https://github.com/agentile/PHP-Stanford-NLP, 10-05-2017.

[34] www.foxsurfer.com/wordnet/wordnet.php.tgz.

[35] B. Rujiang, W. Xiaoyue, L. Junhua. (2010). Extract Semantic Information from WordNet to Improve Text Classification Performance. AST/UCMA/ISA/ACN 2010, LNCS 6059, pp. 409–420. Springer-Verlag Berlin Heidelberg.

[36] F. Christiane. (2010). WordNet. In: R. Poli et al. Theory and Applications of Ontology: Computer Applications, (pp. 231-243). 231-243, DOI: 10.1007/978-90-481-8847-5_10, Springer Science + Business Media B.V.

[37] H. Harb, K., Fouad, N. Nagdy. (2011). Semantic Retrieval Approach for Web Documents. (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No. 9.

[38] Jean-Louis, L., Zouaq, A., Gagnon, M., Ensan, F. (2014). An Assessment of Online Semantic Annotatorsfor the Keyword Extraction Task. PRICAI 2014, LNAI 8862, pp. 548–560.

[39] Amer, E. and Fouad K. (2017). AKEA: An Arabic Keyphrase ExtractionAlgorithm, A.E. Hassanien et al. (eds.), Proceedings of the International Conferenceon Advanced Intelligent Systems and Informatics 2016, Advances in Intelligent Systems and Computing. PP:137-146. Springer International Publishing.

[40] Pu, m. (2006). Fundamental data Compression, ELSEVIER, 1st edition.

[41] Kumar,N.,Srinathan,K. (2008). Automatic keyphrase extraction from scientific documents using N-gram filtration technique . pp. 199-208, proceeding of the eighth ACM symposium on Document engineering.

[42] Aliaa, A., Ghalwash, Y., Amer,E. (2011).KPE:An Automatic Keyphrase Extraction Algorithm, IEEEproceeding of International Conference on Information Systems and Computational Intelligence (ICISCI 2011), pp. 103 -107.

[43] H. Yun, J. Xu, M. J. Xiong. (2009). Development of Domain Ontology for E-learning Course, 978-1-4244-3930-0/09, IEEE.

[44] Alhabashneh, O., Iqbal, R., Doctor, F., Aminm S. (2015). Adaptive Information Retrieval System Based on Fuzzy Profiling. IEEE International Conference on Fuzzy Systems (FUZZ-IEEE).