International Journal of Innovative Computing, Information and Control
Volume **10**, Number **4**, August **2014**

ICIC International ©2014 ISSN 1349-4198
pp. 1351–1367

ARABIC DOCUMENT SUMMARIZATION USING FA FUZZY ONTOLOGY

El-Sayed Atlam[1,2] and Omnia El-Barbary[2]

[1]Department of Information Science and Intelligent System
University of Tokushima
Tokushima 770-8501, Japan

[2]Computer Science Division
Faculty of Science
Tanta University
Tanta, Egypt
satlam@yahoo.com

Received September 2013; revised January 2014

Abstract. *Ontology is the basis of sharing and reusing knowledge on the semantic web. The fuzzy ontology is an extension of the domain ontology for solving the uncertainty problems. Although many earlier methods tried to create a fuzzy ontology and applied it to documents summarization, none of the previous methods have discussed Arabic document summarization using FA fuzzy ontology which could benefit future research. This paper presents a new technique for Arabic document summarization using a fuzzy ontology, which is a fuzzy linguistic variable ontology and Field Association (FA) words. It is more suitable to describe fuzzy linguistic variable ontology than domain ontology for solving the uncertainty reasoning problem. At first, the domain ontology with various events Arabic language is predefined. The document preprocessing mechanism generates the meaningful terms based on Arabic corpus and Arabic language dictionary defined by the domain expert. Then, the meaningful terms have been classified according to an FA term classifier algorithm. Every fuzzy concept has a set of membership degrees associated with various events of the domain ontology. In addition, some process based on the fuzzy ontology is also developed for Arabic document summarization. From the experimental results, it is clear that the average of summarization accuracy achieved 69%, 68% and 68% in terms of Recall, Precision and F-measure. The results show that the Arabic document based on FA words and fuzzy ontology can effectively operate for summarization. Moreover, our new method using FA fuzzy ontology (FAFO) technique could be improved by 30-35% in terms of Recall, Precision and F-measures than traditional methods.*
Keywords: Ontology, Fuzzy ontology, Field association words, Document summarization, Arabic information retrieval, Latent semantic analysis (LSA), Local and global properties (LGP)

1. Introduction.

Modern naval battle forces generally include many different platforms e.g., ships, planes, helicopters. Each platform has its own sensors, e.g., radar, electronic support measures, and communications. The sharing of information measured by local sensors via communication links across the battle group should allow for optimal or near optimal decisions. The survival of the battle group or members of the group depends on the automatic real-time allocation of various resources. The FA fuzzy ontology algorithm has been developed than automatically allocates electronic attack (EA) resources in real-time.

Ontology is a conceptualization of a domain into a human understanding, but machine-readable format consists of entities, attributes, relationships and axioms [10,13]. With the support of the ontology, both user and system can communicate with each other by the shared and common understanding of a domain [36]. There are many ontological applications that have been presented in various domains. For example, the fuzzy ontology is capable of dealing with fuzzy knowledge [20,21,40], and is efficient in text and multimedia object representation and retrieval [29]. Also, a method of extracting information from unstructured documents based on application ontology is presented in [11,12]. The Artequakt that automatically extracts knowledge about artists from the web based on ontology is proposed in [3]. An algorithm to create a fuzzy ontology and apply it to news summarization has been proposed in [25,27]. This work is based on their previous work on ontology-based fuzzy event extraction agents for Chinese news summarization [26]. The ontology learning capability to extract relevant domain terms from a corpus of text has been defined in [30]. Moreover, a Fuzzy Ontology Generation frAmework (FOGA) for the fuzzy ontology generation of uncertainty information is discussed in [31,33,38] using a system designed for content-based information retrieval in [16,35,36,38]. They combined an ontology driven content-matching mechanism with moderately expressive representation formalism. Furthermore, a fuzzy ontology framework in which a concept descriptor is represented as a fuzzy relation which encodes the degree of a property value using a fuzzy membership function is discussed in [2]. A new fuzzy extension of description logics called the fuzzy description logics with comparison expressions (FCDLs) is illustrated in [22]. The goal of text summarization is to take the abstract from the extracted content and present the most important message for the user in a condensed form. Moveover, automatic document summarization is developed in this area [17,19,25,29]. Although many earlier methods tried to create a fuzzy ontology and applied it to documents summarization, none of the previous methods have discussed Arabic document summarization using FA fuzzy ontology which could benefit future research.

Due to the morphological complexity of the Arabic language, Arabic morphology has become an important part of many Arabic Information Retrieval (IR) and other natural language processing applications. Arabic words are divided into three types: noun, verb, and particle. Nouns and verbs are derived from a closed set of around 10,000 roots [19,37]. The roots are commonly three or four letters and are rarely five letters. Arabic nouns and verbs are derived from roots by applying templates to the roots to generate stems and then introducing prefixes and suffixes. For Arabic IR, several early studies suggested that indexing Arabic text using roots significantly increases retrieval effectiveness over the use of words or stems [1,4,18]. However, other studies used small test collections of only hundreds of documents and the morphology in many of the studies was done manually [5,18,24].

Since the grammar of Arabic has been standardized for centuries, Khoja in [23] decided to derive initial tag set from this grammatical tradition rather than from an Indo-European based tag set. The reason for this is that Arabic is a very different language from Indo-European languages, and should have its own tag set. Also, Arabic linguists will be basing their studies on a traditional Arabic grammar rather than an Indo-European grammar. Arabic grammarians traditionally analyze all Arabic words into three main parts-of-speech. These parts-of-speech are further sub-categorized into more detailed parts-of-speech which collectively cover the whole of the Arabic language. The three main parts of speech are: Nouns (N), Verbs (V) and Particles (P). We designed POS tag tree for Khoja tag set in Figure 4 to use it in calculating the membership function.

Readers can know the subject of many document fields of finding only some specific words without reading the whole text and give others the outline. Document fields can be decided efficiently if there are many FA words and if the frequency rate is high [15]. The concept of FA words is introduced in many papers, some of these papers introduced this concept and gave algorithms to find FA terms in a specific field [14], other papers used FA terms in document similarity [7-9] and other papers used FA terms in passage retrieval. This paper presents a strategy for building an Arabic FA fuzzy ontology and uses it in Arabic document summarization as follows:

(1) Extract efficient FA words from Arabic documents.
(2) Building fuzzy domain ontology uses FA words with its weight and POS similarity.
(3) After that, the relation between an object and an attribute in ontology is represented by a membership function between [0, 1], which is called fuzzy ontology.
(4) Concatenate words and relation using a sentence generator algorithm to perform sentences and it is denoted as follows:
    Sentence: [Concept Name, Attribute value, Operation] Relation [Concept Name, Attribute value, Operation].
(5) Eliminate the unnecessary sentence using a sentence filter algorithm.
(6) Obtain the summarization of Arabic documents.

From the experimental results, it is clear that the average of summarization accuracy achieved 69%, 68% and 68% in terms of Recall, Precision and F-measure. The results show that the Arabic document based on FA words and fuzzy ontology can effectively operate for summarization. Moreover, our new method using FA fuzzy ontology (FAFO) technique could be improved by 30-35% in terms of Recall, Precision and F-measures than traditional methods such as Latent Semantic Analysis (LSA), Local and Global Properties (LGP), Semantic Variable Matrix (SVM), and Material Requirements Planning (MRP).

This paper is organized as follows. Section 2 gave some definitions of fuzzy ontology. The fuzzy ontology construction is illustrated in Section 3. Section 4 introduced the Arabic text summarization using fuzzy ontology. The experimental result introduced in terms of Recall, Precision and F-measures and the comparison with some traditional method are done in Section 5. Section 6 focuses on conclusion and possible future work.

2. Definitions of the Fuzzy Ontology.

In this section, we will present a fuzzy ontology for Arabic document summarization. Because the fuzzy ontology is an extension of the domain ontology, a formal definition of the domain ontology for document summarization will be introduced here.

Definition 2.1. *(Ontology) A domain ontology defines a set of representational terms that we call concepts. Interrelationships among these concepts describe a target world (an example for domain ontology can be found in Figure 2).*

Ontology is 4-tulle $ONT = (C, F, R, O)$, where:

1. C is a set of concepts defined for the domain. A concept is often considered as a class in ontology. Each concept contains a concept name C_i with a term set $T = \{T_{C_i1}, T_{C_i2}, \ldots, T_{C_iq_i}\}$.

2. F is a set of fields which is collected by domain expert; $F = \{F_1, F_2, \ldots, F_p\}$ each field comprises several concepts of class level.

3. R is a set of binary semantic relations defined between concepts in C. $R_t = \{$generalization, aggregation, association$\}$ is the set of relation type, where:

- The generalization relation is the relationship between a domain and its corresponding category that represents "is-kind-of" relationship.
- The aggregation relation is the relationship between each category and its corresponding field and it represents "is-part-of" relationship.
- The association relation represents a semantic relationship between concepts in class level.

4. O is a set of operation $\{O_{C_i1}, O_{C_i2}, \ldots, O_{C_iq_i}\}$ for an application domain.

In the following, domain ontology extended to be fuzzy ontology by adding a set of membership degrees of each concept of the domain ontology and adding fuzzy relationships among the fuzzy concepts.

Definition 2.2. *(Fuzzy Concept)*

A fuzzy concept is a refined concept derived from domain ontology. It is a refinement by embedding a set of membership degrees associated with a set of the fields in the concept of the domain ontology. If a domain ontology has a field set $F = \{F_1, F_2, \ldots, F_p\}$, and a concept C_i, then we can refine the C_i into the fuzzy concept and denote the fuzzy concept as $\{C_{i;\mu_{C_iF_1},\mu_{C_iF_2},\ldots,\mu_{C_iF_p}}\}$ with a term set $T = \{T_{C_i1}, T_{C_i2}, \ldots, T_{C_iq_i}\}$ and an operation set $\{O_{C_i1}, O_{C_i2}, \ldots, O_{C_iq_i}\}$, where $\mu_{C_iF_i}$ represents the membership degree of C_i for field F_j. The term $T_{C_iq_i}$ and the operation $O_{C_iq_i}$ are denoted the q_ith term and q_ith operation of C_i, respectively.

Definition 2.3. *(fuzzy relationship)*

$R = \{r : r \subseteq C \times C \times R_t\}$ is a set of binary semantic relations defined between concepts in C. $R_t = \{one - to - one, one - to - many, many - to - many\}$ is the set of relations type. A set of basic relations is defined as {synonym of, kind of, part of, instance of, property of} $\subset R$ which have the following interpretations:

(1) c_i Synonym of c_j: c_i is equivalent to c_j. The synonym relation of natural language is modeled in an ontology using the equivalence relation. If two concepts c_i and c_j are declared equivalent in an ontology, then instances of concept c_i can also be inferred as instances of c_j and viceversa.

(2) c_j Kind of c_i: c_i is a generalization of c_j. When an ontology specifies that c_i is a generalization of c_j, then c_j inherits all property descriptors associated with c_i, and these need not be repeated for j c while specifying the ontology.

(3) c_j Part of c_i: c_i has part j c. In ontology, a concept which is defined as aggregation of other concepts is expressed using this relation.

(4) c_j instance of c_i: c_j is an instance of c_i.

(5) c_j property of c_i: c_j is a property of c_i.

Definition 2.4. *(fuzzy ontology)*

A fuzzy ontology is extended domain ontology with fuzzy concepts and fuzzy relation. Definitions for a fuzzy ontology found from the literature, e.g., [32,34,40] draw influence from both fuzzy set theory and existing ontology languages. According to [30], a fuzzy ontology is based around the idea that each concept (or term) is related to every other concept in the ontology, with a degree of membership assigned to that relationship based on fuzzy logic. The fuzzy membership value μ ranges from 0 to 1, and for each concept: $\sum_{i=1}^{n} \mu_i = 1$, where n is the number of relations a particular concept has, which is one less than the total number of concepts in the ontology. The membership value of the relation of conceptual A to concept B, μ_{AB} is not necessarily the same as it is in the relation of concept B to concept A, μ_{BA}. The definition is illustrated in [30] with an example of how the concept "Apple" can simultaneously be understood to represent a computer company, a fruit, and a tree.

Since there is no universally accepted definition of a fuzzy ontology yet, we set about defining a set of requirements for a fuzzy ontology that illustrate what kind of ontology constructs are needed in order for an ontology to be of good use in our domain.

Definition 2.5. *(Fuzzy domain ontology)*

Fuzzy domain ontology is a 4-tuple $ONT_F = (C, P_F, R_F, A_F)$, where:

1. C is a set of concepts. Differing from Definition 2.1, every concept here has some properties whose value is fuzzy concept or fuzzy set.

2. P_F is a set of properties. A property $p_F \in P_F$ is defined as a 5-tuple of the form $p_F(c, v_F, q_F, f, U)$, where $c \in C$ is an ontology concept, v_F represents property values, q_F models linguistic qualifiers, which can control or alter the strength of a property value v_F, f is the restriction facets on v_F, and U is the universe of discourse. Both v_F and q_F are the fuzzy concepts but, q_F changes the fuzzy degree of v_F. For example, "price" is a property of concept "fruit". The value of "price" may be either fuzzy concept "cheap" or fuzzy number "around 50", and the linguistic qualifiers may be "very", "little", "close to", etc. Therefore, the final value of "price" may be "very cheap" or "little expensive".

3. R_F is a set of inter-concept relations between concepts. Like fuzzy concept properties, $r_F \in R_F$ is defined as a 5-tuple of the form $r_F = (c_1, c_2, t, s_F, U)$, where $c_1, c_2 \in C$, are ontology concepts, t represents relation type, U is the universe of discourse, and s_F models relation strengths and it is fuzzy concept at U, which can represent the strength of association between concept-pairs $< c_1, c_2 >$.

4. A_F is a set of fuzzy rules. In a fuzzy system the set of fuzzy rules is used as knowledge base. The fuzzy domain ontology y is used to model domain expert knowledge. However, due to the lack of relationships between fuzzy concepts that can be the value of properties, it is difficult to integrate diverse ontology systems. For example, in ontology the set of property "price" value is {cheap, appropriate, expensive ...}, and in another ontology the same set is {high, low, middle ...}. To map these ontologies, it is necessary to define the semantic relationship between fuzzy concepts, e.g., "cheap" and "expensive" which have the relation of disjointness. Also, "low" and "high" have the same relation of disjointness, etc.

3. **Fuzzy Ontology Construction.** In the following a fuzzy inference mechanism will apply to construct a fuzzy domain ontology. The process of the fuzzy ontology construction can show in Figure 1. First, the Arabic domain ontology in Arabic and English are predefined by domain experts as shown in Figures 2 and 3. Second, the document preprocessing mechanism will generate the classified FA terms based on Algorithm 1 (extracting FA terms). In addition, the preprocessing consists of a Part of Speech tagger (POS) as in Figure 4. Furthermore, Table 1 shows some of POS defined by Khoja in [23]. Khoja defined 177 tags that it is 103 Nouns, 57 Verbs, 9 Particles, 7 Residual, and 1 Punctuation. The term filter will preserve the meaningful terms according to the POS tagging. Finally, the fuzzy inference mechanism will generate the membership degrees for each fuzzy concept of the fuzzy ontology.

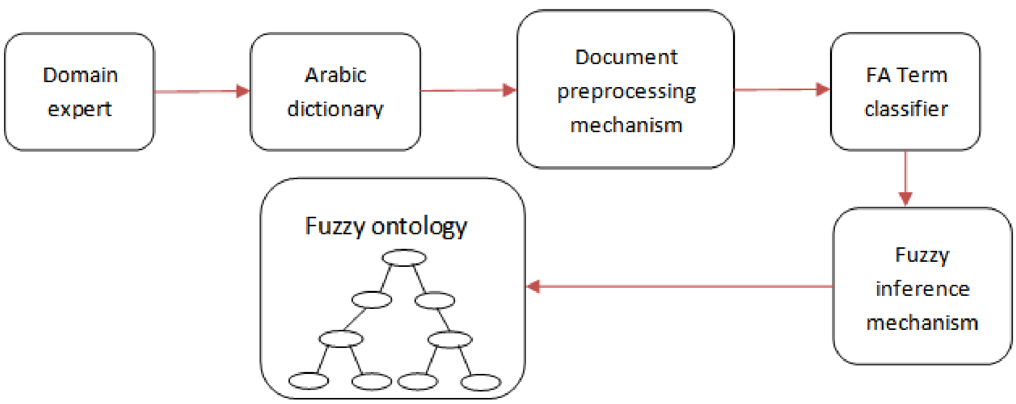

FIGURE 1. Process of the fuzzy ontology construction

TABLE 1. Examples for Part of Speech tags of arabic words derived by Khoja

Tag	Description of word category	Example (Arabic)
NCSgMNI	Singular, masculine, nominative, indefinite common noun	كتابٌ (ktaabon- which means book in English)
NCSgMAI	Singular, masculine, accusative, indefinite common noun	كتابا(ktaaba- which means book in English)
NCSgMGI	Singular, masculine, genitive, indefinite common noun	كتابٍ (ktaaben- which book in English)
NCSgMND	Singular, masculine, nominative, definite common noun	الكتابُ (al ktaabo- which means the book in English)
NCSgMAD	Singular, masculine, accusative, definite common noun	الكتابَ (al ktaaba- which means the book in English)
NCSgMGD	Singular, masculine, genitive, definite common noun	الكتابِ(al ktaabe- which means the book in English)
NCPlMNI	Plural, masculine, nominative, indefinite common noun	كتبٌ (ktobon- which means the books in English)
NCPlMAI	Plural, masculine, accusative, indefinite common noun	كتبا (kataba- which means books in English)
NCPlMGI	Plural, masculine, genitive, indefinite common noun	كتبٍ (ktoben- which means books in English)
NCPlMND	Plural, masculine, nominative, definite common noun	الكتبُ (al kotobo- which means the books in English)
NCPlMAD	Plural, masculine, accusative, definite common noun	الكتبَ (al ktabaa- which means the books in English)
NPrPSg1	First person, singular, neuter, personal pronoun	كتابي (ktabey- which means my book in English)
NPrPSg2M	Second person, singular, masculine, personal noun	كتابكَ (ktabaka- which means your book in English for masculine)
NPrSg2F	Second person, singular, feminine, personal noun	كتابكِ (ktabeke- which means your book in English for feminine in English)
NPrPSg3M	Third person, singular, masculine, personal pronoun	كتابه (katabaho- which means his book in English)
NPrPSg3F	Third person, singular, feminine, personal pronoun	كتابها (katabaha- which means her book in English)

3.1. **Document preprocessing mechanism.** The document preprocessing mechanism consists of a POS tagger and a term filter to get the meaningful term set including nouns and verbs. Table 1 shows some of POS tags that are developed by Khoja. The term filter will preserve the meaningful terms according to the POS tags [23]. In addition, the meaningful terms will be classified according to the domains of the corpus by using the FA term classifier. In the following, we describe the FA term classifier algorithm.

Techniques based on FA words [6,39,40] can recognize fields by using specific words without reading the whole document. For example, people who encounter the word "ميزان المدفوعات" (mizan al madfoat- which means balance of payments in English) can recognize the document field <الأقتصاد> (al Alaguetsad- which means the economy in English). Therefore, document fields can be decided efficiently if there are many FA words and they are frequent enough. Moreover, FA words can help to extract fields-coherent passages from the whole text [7,8].

Definition 3.1. *(FA words and their levels)*

Five ranks are defined to classify FA words to document fields [1,7,9] as follows:

(1) Perfect-FA words *(PFA) associate with one terminal field.*

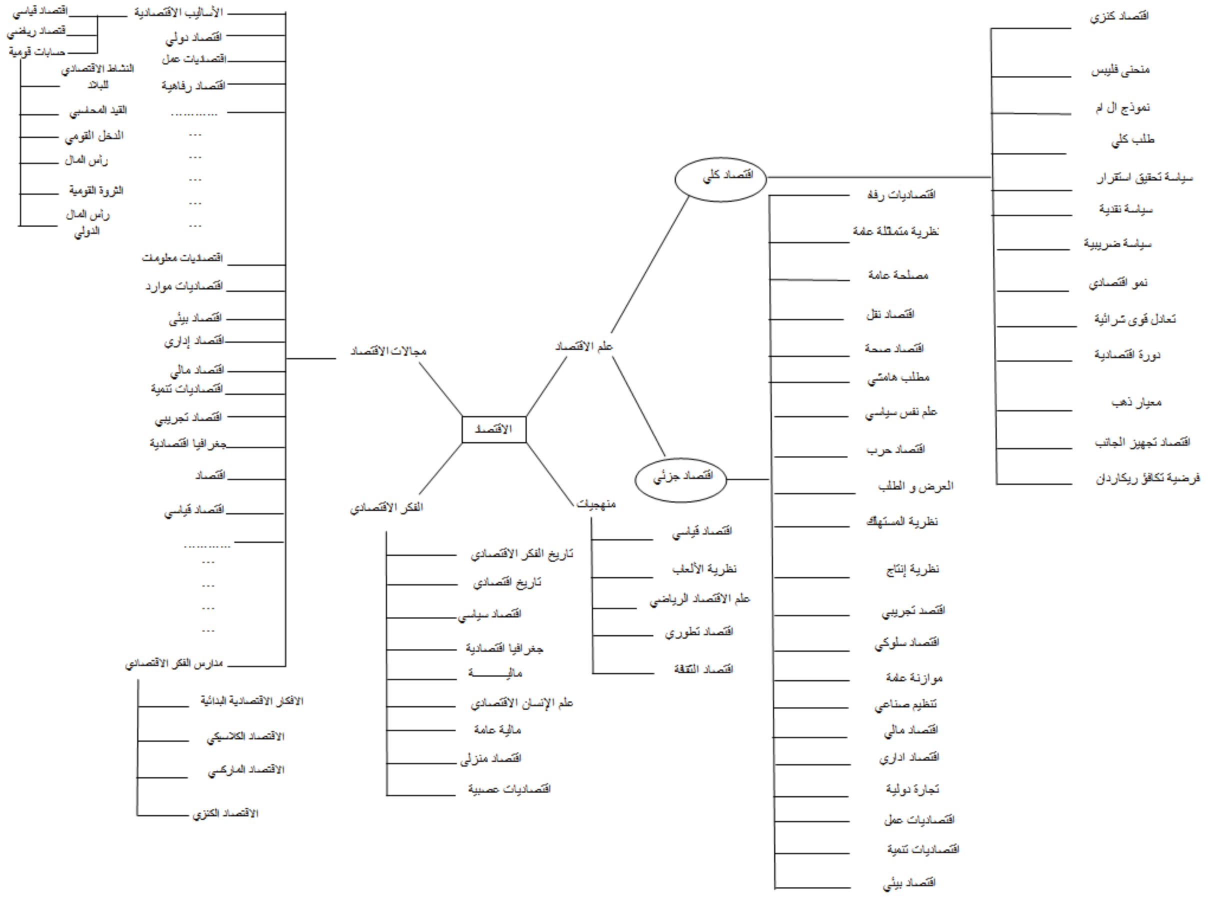

FIGURE 2. Arabic economy ontology

*(2) **Semi-perfect FA words** (SPFA) associate with more than one terminal field in one medium field.*

*(3) **Medium-FA words** (MeFA) associate with one medium field only.*

*(4) **Multiple-FA words** (MuFA) associate with more than one terminal field and more than one medium field.*

*(5) **Non-Specific FA words** (NSFA) do not specify terminal fields or medium fields.*

Non-Specific FA words include stop words (e.g., articles, prepositions, pronouns). Table 2 shows examples of Arabic FA words and their ranks.

Algorithm 1: FA Words Determination Algorithm

Input: *1- w, is a set of candidates FA*

2- Norm $(w, < S >)$ for w and $< S >$

3- $\propto$ a threshold to judge FA words ranks

Output: PFA, $SPFA$, $MeFA$, $MuFA$

Method:

1) *Set* $PFA = \{\ \}$, $SPFA = \{\ \}$, $MeFA = \{\ \}$, $MuFA = \{\ \}$.

2) *set root* $= < S >$, *set child* $< s/c >$.

3) *for the root* $< S >$ *and any child* $< S/C >$

calculate $conc(w, < S >) = ((Normalization(w, < C >))/(Normalization(w, < S >))),$

$Normalization\ (w, < T >) = ((Frequacy(w, < T >))/(Total - Frequancy(< T >)))$

if $(conc(w, < S >) \wedge conc(w, < S/c >)) \geq \propto$

$\quad$ *Then set w in class PFA*

Else

$\quad$ *if* $(conc(w, < S >) \geq \propto \wedge conc(w, < S/c >)) < \propto$

$\quad\quad$ *Then set w in class SPFA*

Else go to Step 4

4) *if* $(w \notin PFA)$ then $< s >$ is a medium field and $\exists m \geq 2$ children fields $< s/c >$ $\forall < s/c_k >$, $(1 < k < m)$

$$calculate\ Aver = \frac{1}{m}\left[\sum_{k=1}^{m} Conc(w, < c_k >)\right]$$

and $ACR(w, < c_k >) = conc(w, < c_1 >) + \ldots + conc(w, < c_m >)$.

5) if $ACR(w, < c_k >) > \propto$ *and* $< s/c_k >$ *are all terminal field, then append* $(SPFA, w)$

else

if $(ACR(w, < c_k >) \leq \propto$ *then append* $(MeFA, w)$

$\quad$ *if* $< S/c_k >$ *are not all terminal children go to next step.*

6) *extract the terminal field* $< S/c >$ *from k children*

$\quad$ *if* $< S/c_k >$ *is medium filed and it is a terminal field then append* $(MuFA, w)$

7) *return 1*

3.2. **Fuzzy inference mechanisms.** We apply a fuzzy inference mechanism to build the fuzzy ontology. The fuzzy inference mechanism will generate the membership degrees for each fuzzy concept of the fuzzy ontology. Every fuzzy concept has a set of membership

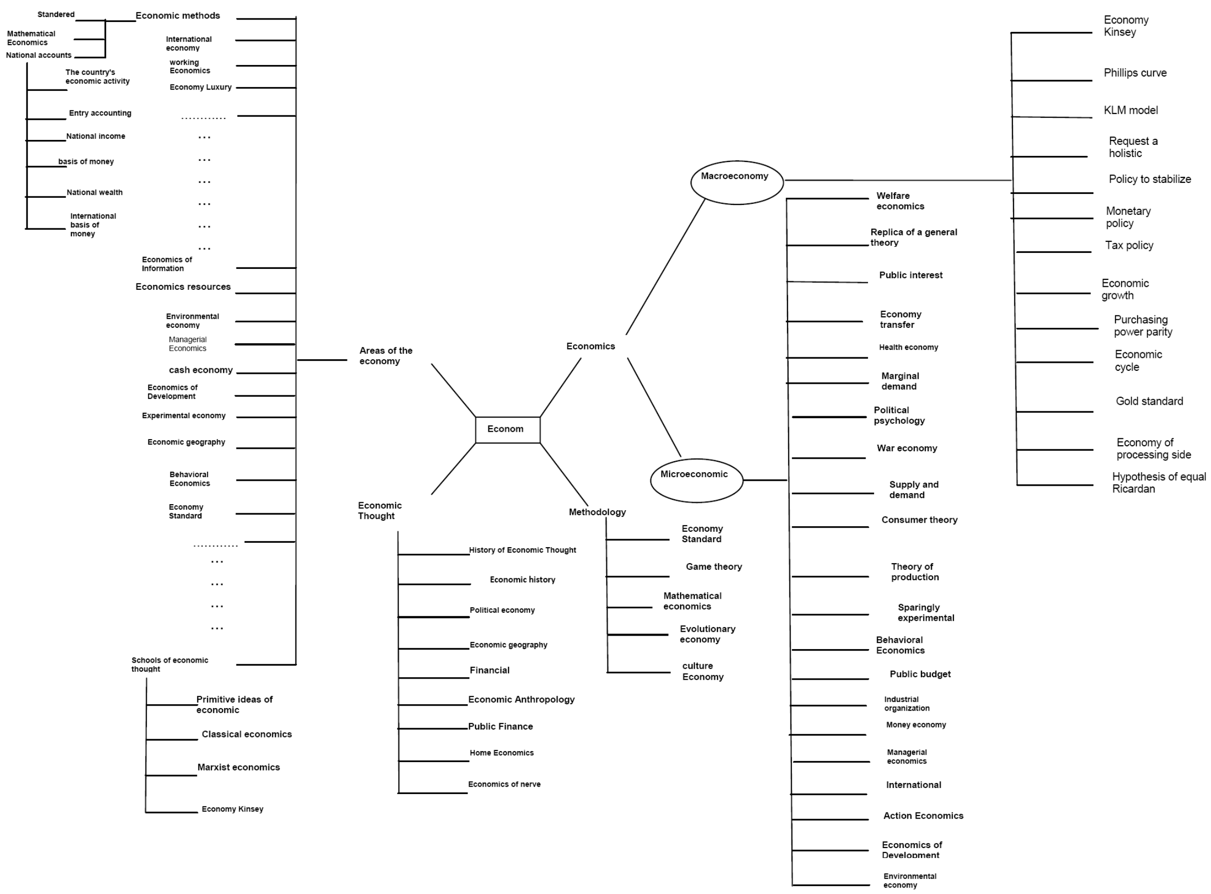

FIGURE 3. English economy ontology

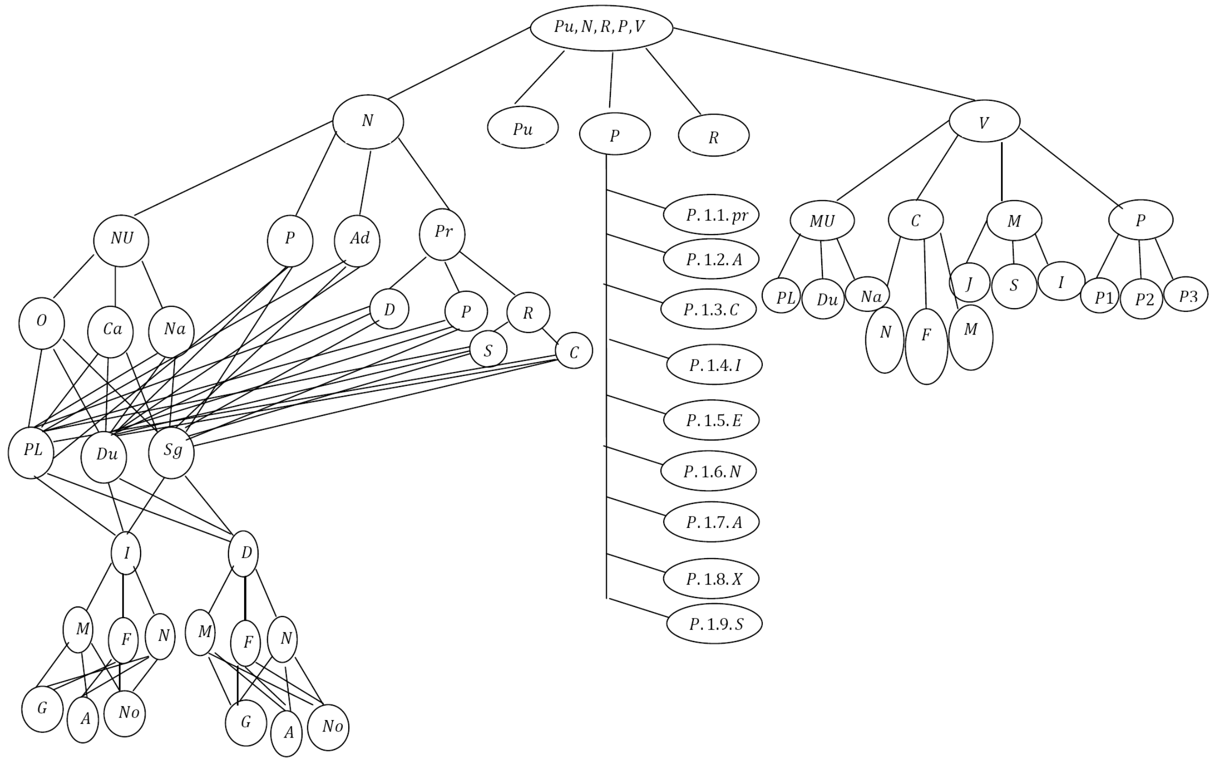

FIGURE 4. Arabic part of speech tags

degrees associated with various events of the domain ontology. In the following we describe the fuzzy inference mechanism in detail; this process consists of five steps.

Step 1: This step is called input linguistic; in this step we apply the FA term extraction algorithm. The input vectors are the term set of concepts that are retrieved from FA term algorithm and these concepts are classified into four categories PFA, SPFA, MeFA and MuFA. The nodes in the first step transmit input values to next step.

Step 2: This step performs the membership function to compute the membership degree for all FA terms derived. There are two input fuzzy variables, including Term Weight (TW) and Term Part-of-Speech (POS) similarity, considered in input term layer for each term's property. The structure of an input term node contains two parts, including fuzzy variable part and linguistic term part, to perform the membership degree computing. The fuzzy variable part will extract the input values from $(T_{FA_1 1}, T_{FA_1 2}, \ldots, T_{FA_1 n})$ for the fuzzy variable and send the extracted values to the Linguistic Term Part. In addition, the linguistic term part operates to calculate the membership degree with respect to the input values. Each fuzzy variable has three linguistic terms Low (L), Median (M) and High (H).

The first fuzzy variable proposed for computing the weight of any term is Term Weight (TW). To compute the membership degree for this term, we use the following equation.

$$\mu_{TW} = \frac{\sum_{i=0}^{k} L_i W_i}{\sum_{i=0}^{k} A_i W_i}$$

μ_{TW} is a function for each FA term occurring in a particular location in the ontology. It is calculated by summing the number of FA term discovered in each section of a document, and they are discovered k FA term multiplied by its weight (W_i). This was normalized by dividing the sum of all terms discovered multiplied by the weight (W_i) over all documents in the set k.

TABLE 2. Examples of arabic FA words and their ranks

Ranks	*FA words*	FA term weight
PFA	"الاقتصاد المصري" (al Alaguetsad al msrey- which means the Egyptian economy in English)	4
SPFA	"القطاعات الاقتصادية" (al qtaat al Alaguetsadye- which means the Economic sectors in English) "البنك المركزي" (al bank al markazy- which means the Central Bank in English) "دولار" (dollar- which means the dollar in English) "الاستثمارات" (al estthemarat- which means the investments in English)	3
MeFA	"السنة المالية" (al sanah al malyha- which means Financial year in English) "فائض كلي" (faed koly- which means Total surplus in English) "ميزان المدفوعات" (mizan al madfoat- which means Balance of Payments in English) "العائدات" (al aedat- which means the Proceeds in English)	2
MuFA	"أسعار" (asar- which means prices in English) "صرف" (al sarf- which means exchange in English) "نقد" (nqed- which means cash in English) "دخل" (dakhl- which means income in English)	1
NSFA	"من" (mn which means of in English) - "في" (fe- which means "in" in English)	0

TABLE 3. Fuzzy linguistic term values

Fuzzy variable	**Linguistic Term**	**Value**
POS	POS Low	1
	POS Median	0.5
	POS High	0
TW	TW Low	0
	TW Median	0.5
	TW High	1
TRS	TRS Very Low	0
	TRS Low	0.3
	TRS Median	0.5
	TRS High	0.7
	TRS Very High	1

A weight was introduced so that terms of PFA level weight are as 4, terms with SPFA level weight as 3, terms with MeFA level weight as 2 and terms with MuFA level weight as 1. An example of FA level and its weight can be shown in Table 2.

The second fuzzy variable proposed for computing the relation strength of any term pair is POS. In Figure 4 each node of the tagging tree represents a Part of Speech tagging. We will utilize the length of the path between every two nodes to compute the POS similarity for each term pair. The path length of every two terms defined by the tagging tree is bounded in the interval $[0, \frac{a}{n}]$, where n is the number of path length, $n = 11$, a is the distance path between two nodes. For example, let us have two terms with NIMANUCaSg and VCF POS tags; hence the distance path between them is 10. There are three linguistic terms including POS High, POS Median and POS Low defined for POS similarity. Figure 5 shows the membership function of the fuzzy sets for POS similarity. Table 3 shows the values of the fuzzy number for POS similarity.

To compute the μ_{POS} function we use the following equation:

$$\mu_{POS} = \frac{1}{1 + \alpha^{\frac{\max(R_i - R_{\min}, 0)}{\max(-R_i, R_{\min})}}}$$

where, R_i is the length similarity for each term, $R_{\min}$ is the smallest length similarity and α is a threshold function defined by domain experts. The membership function consider the root of concept "close" which refer to how close any word to another [34].

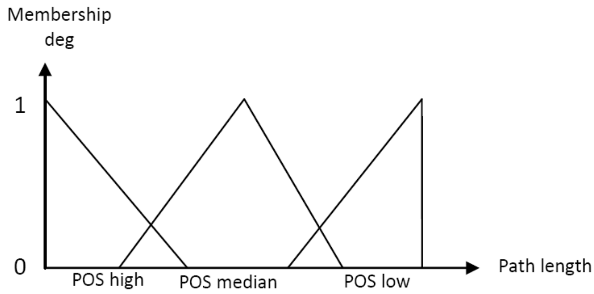

Figure 5. POS membership function

Step 3: This step is called rule step, and each node is a rule node to represent a fuzzy rule. The rule node should perform the fuzzy AND operation [42-44], and the outputs will be linked with associated linguistic node in the fourth step. We use the algebraic product operation to compute the matching degree. The rules are defined by domain expert's knowledge previously, and we show them in Table 4.

Table 4. Fuzzy inference rule for fuzzy inference mechanism

POS	TW	TRS
H	H	VH
H	M	H
H	L	H
M	H	H
M	M	M
M	L	L
L	H	H
L	M	L
L	L	VL

Step 4: This is the output step, and in this step we have two outputs terms and linguistics. The output term performs the fuzzy OR operation to integrate the fired rules that have the same consequences. The fuzzy variable defined in output term is Terms Relation Strength (TRS). There are five linguistic terms, including TRS Very Low (VL), TRS Low (L), TRS Median (M), TRS High (H), and TRS Very High (VH) in TRS. Table 3 shows the parameter values of the fuzzy number for TRS fuzzy variable defined by domain experts. On the other hand, the output linguistic performs the defuzzification process to get the TRS value.

Step 5: This is the summation and integration step, the summation process computing the TRS values of the specific concept C_j of T_{FA_1}. The integration process integrates the membership degrees of concept C_j that belongs to all T_{FA_s} of the domain ontology. Finally, the fuzzy ontology is represented as: $\{(C_1; \mu_{C_1FA_1}, \mu_{C_1FA_2}, \ldots, \mu_{C_1FA_p}),$

$(C_2; \mu_{C_2FA_1}, \mu_{C_2FA_2}, \ldots, \mu_{C_2FA_p}), \ldots, (C_m; \mu_{C_mFA_1}, \mu_{C_mFA_2}, \ldots, \mu_{C_mFA_p})\}$, where the output vector denotes the membership degree of the jth fuzzy concept.

Algorithm 2: Fuzzy ontology construction algorithm
Input: *a) T a set of terms generated by FA algorithm.*
b) C a set of concept domain ontology.
Output: *fuzzy domain ontology.*
Method: *for all (t, c) and it is corresponding TRS*
if TRS value is greater than threshold
then
the concept C will join to TRS
else
return 1
join each concept with its corresponding weight and
a domain ontology to generate fuzzy domain ontology
End

Example 3.1. *For the economic ontology in previous Figure 2, first we obtain the set of FA words using 134 candidates word. Then, apply the fuzzy inference mechanism to construct the fuzzy economic ontology. Figure 6 is an example of economic fuzzy ontology. In Figure 6, the domain name is "الاقتصاد" (al eqtasad- which means the economic in English) and it consists of several categories such as "الاقتصاد المصري" (al eqtasad al masry- which means the Egyptian economic in English), and others. In addition, there are several events defined in this domain ontology. For example, the economic events "القطاعات الاقتصادية" (al qata3t al aqtqsadyah- which means the Economic sectors in English), "البنك المركزي" (al bank al markazy- which means the Central Bank in English), "ميزان المدفوعات" (mizan al madfoat- which means Balance of Payments in English), "فائض كلي" (faed koly- which means Total surplus in English), and others are related to the categories "الاقتصاد المصري" (al eqtasad al masry- which means the Egyptian economic in English). For example, the membership degrees of the fuzzy concept "البنك المركزي" (al bank al markazy- which means the Central Bank in English) for the events "فائض كلي" (faed koly- which means Total surplus in English) is 0.9.*

4. Apply Fuzzy Ontology to Arabic Document Summarization.

Our aim is to perform a summarization for Arabic document. In Section 3 we describe how to build fuzzy domain ontology. In this section we illustrate how to obtain sentence from the fuzzy ontology using sentence path extraction. Furthermore, delete the unnecessary sentence that is called noisy sentence and create a set of summarize sentences using sentence filter technique.

4.1. Sentence path extractor and sentence generator.

The sentence generator uses the following algorithm to look sentence paths from fuzzy ontology.

Algorithm 3: Sentence Path Extractor Algorithm
Input: *All classified terms generated by FA words Algorithm and all*
fuzzy concepts of the fuzzy domain ontology.
Output: *A set of sentence paths.*
Method: *Let $FO = \varnothing$ a set of fuzzy ontology.*
Let R_{ij} be the relation between two concepts C_i, C_j
$\forall(\boldsymbol{t} \in C_i \wedge t$ *Generated by FA words Algorithm)*
If $\boldsymbol{t} \in C_i$ then
If $C_i \notin F$ then
Join C_i to FO

Join t to FO
If $t = R_{ij}$ then join R_{ij} to FO

End

The result of Sentence Path Extractor is denoted as follows:
Sentence Path P: [Concept Name] → [Concept Name] → ... → [Concept Name]

Example 4.1. *For the economic ontology in Example 3.1, the path extractor for the words "البنك المركزي" (al bank al markazy- which means the Central Bank in English) "ميزان المدفوعات" (mizan al madfoat- which means Balance of Payments in English), "فائض كلي" (faed koly-which means total surplus in English) is as follows:* البنك المركزي ← فائض كلي ← ميزان المدفوعات.

4.2. **Sentence filter.** The sentence filter contains two steps first location combination process and sentence combination process to filter the redundant sentences. In the first step, the relevant location and repeated phase are considered to filter. In addition, the second step will generate the brief sentence set based on the results of location combination process. In the sentence combination process, the sentence filter will combine the sentences, which have some common concepts.

Example 4.2. *For the economy fuzzy ontology in Example 3.1 the sentence "الاستثمارات مكنت الدولة من زيادة احتياطي النقد" (which means "Investments have enabled the government to increase cash reserve" in English) with possibility 0.99 and the sentence "الاستثمارات أدت إلى زيادة احتياطي النقد" (which means "investments led to increased currency reserves" in English) with possibility 0.95. Both sentences have the same concepts "الاستثمارات" and*

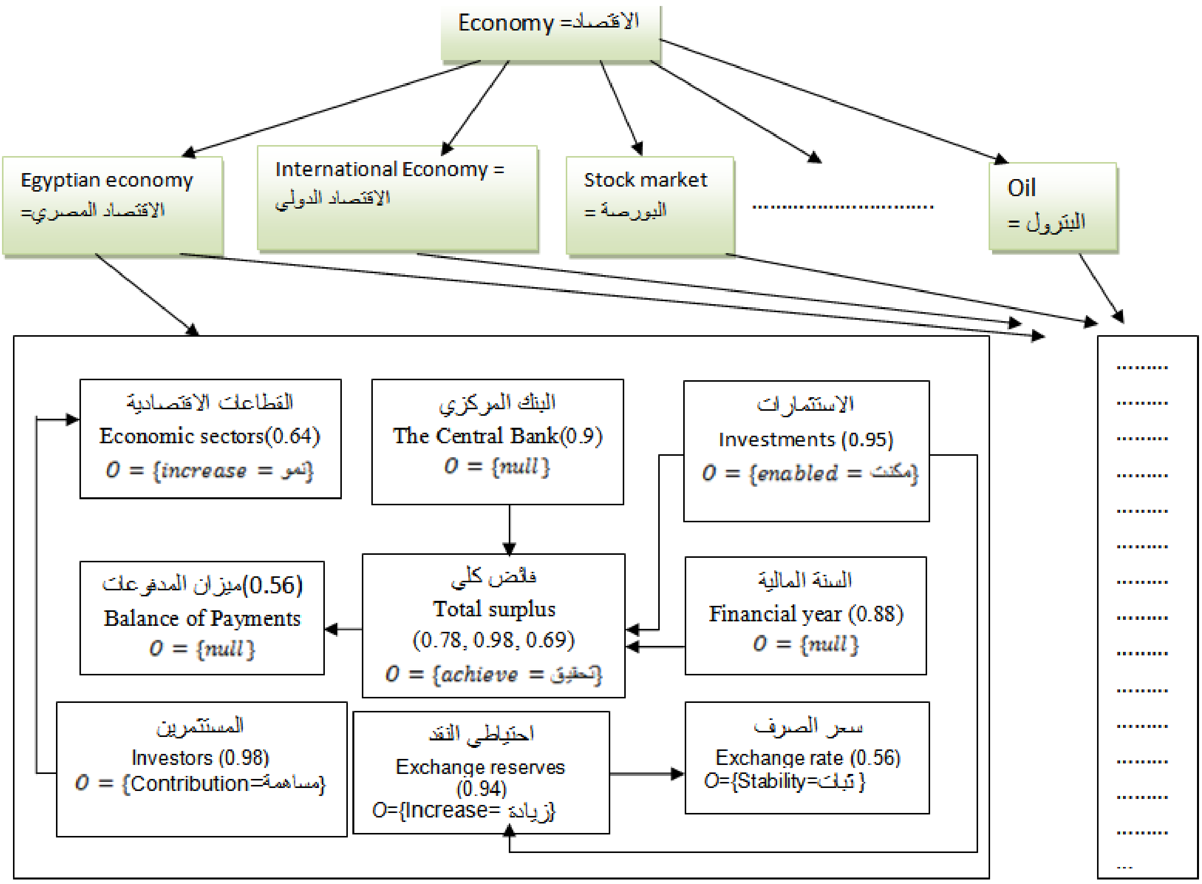

FIGURE 6. Arabic fuzzy economic ontology

“احتياطي النقد”; therefore, the sentence filter will combine these sentences into the former sentence; “الاستثمارات مكنت الدولة من زيادة احتياطي النقد” (which means “investments have enabled the government to increase cash reserve” in English) with possibility 0.99.

5. **Experimental Evaluations.** For evaluating the performance of our approach, we adopt the performance measures Precision (P) and Recall (R) in our system. We evaluated the performance of our summarization technique on a variety of document collections like economy, health, sports and others. Our experiments trained the system using Arabic documents collected from the Internet. It is mainly collected from Al-Jazeera Arabic news channel which is the largest Arabic site, Al-Ahram newspaper, Al-watan newspaper, Al Akhbar, Al Arabiya, Al hayaha and Wikipedia the free encyclopedia.

The formulas of precision and recall measures utilized in this paper are as follows

$$\begin{aligned} &Precision \\ &= P \\ &= \frac{\text{The number of relevant sentences between the standard summary and our generated summary}}{\text{The number of sentences in standard summary}} \end{aligned}$$

$$\begin{aligned} &Recall \\ &= R \\ &= \frac{\text{The number of relevant sentences between the standard summary and our generated summary}}{\text{The number of sentences in our generated summary}} \end{aligned}$$

$$F\text{-}measure = \frac{2 \times P \times R}{P + R}$$

Table 5 shows the summarization accuracy using FA fuzzy ontology technique in terms of Recall, Precision and F-measure.

Table 5. The summarization accuracy achieved by FA fuzzy ontology technique

Data	P	R	F
الاقتصاد (al aqtasad- which means economics in English)	0.714	0.83	0.76
البيئة (al byaah- which means environment in English)	1	0.83	0.9
التغذية (al taghzya- which means feeding in English)	0.8	0.57	0.66
الثقافة (al thaqafah- which means culture in English)	0.875	0.77	0.82
الرياضة (al ryadah- which means sports in English)	0.6	0.66	0.63
الصحة (al sahah- which means health in English)	0.73	0.8	0.76
الطب (al tab- which means medicine in English)	0.7	0.63	0.66
الطفل (al tafel- which means child in English)	0.57	0.66	0.61
الفن (al fan- which means art in English)	0.33	0.4	0.36
السياسة (al syasah- which means politics in English)	0.66	0.66	0.66
Average	**0.69**	**0.68**	**0.68**

From Table 5, it is clear that the average of summarization accuracy achieved 69%, 68% and 68% in terms of Recall, Precision and F-measure for various fields such as Economics, Environments, Feeding, Culture, Sports, Health, Medicine, Arts and Politics, respectively.

Figure 7 shows the comparison of the proposed technique using FA fuzzy ontology (FAFO) technique with the traditional methods Latent Semantic Analysis (LSA), Local and Global Properties (LGP), Semantic Variable Matrix (SVM), and Material Requirements Planning (MRP) proposed by [28,41].

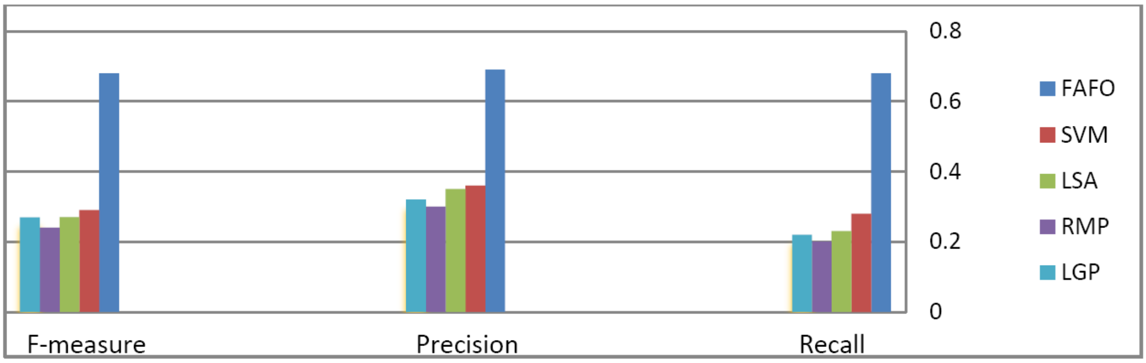

Figure 7. Comparison between our new method and traditional methods

From Figure 7, it is clear that our new method using FA fuzzy ontology (FAFO) technique could be improved by 30-35% in terms of Recall, Precision and F-measure than traditional methods.

6. **Conclusion and Future Work.** In this paper, we propose an FA fuzzy ontology for Arabic document summarization. The FA word algorithm is used to compute each concept to assist in constructing the FA fuzzy ontology. In addition, this ontology is used to extract a summarization of Arabic text. By the experimental results, we express that the proposed approach can summarize the Arabic text effectively. In the future, we will extend the FA fuzzy ontology summarization agent to deal with the Heritage Arabic documents. From the experimental results, it is clear that the average of summarization accuracy achieved 69%, 68% and 68% in terms of Recall, Precision and F-measure. The results show that the Arabic document based on FA words and fuzzy ontology can effectively operate for summarization. Moreover, our new method using FA fuzzy ontology (FAFO) technique could be improved by 30-35% in terms of Recall, Precision and F-measures than traditional methods. Future work could focus on comparing our new system for Arabic document summarization with other models developed such that, the Arabic Query-Based Text Summarization System (AQBTSS), uses standard retrieval methods to map a query against a document collection and to create a summary. And, the Arabic Concept-Based Text Summarization System (ACBTSS), creates a query-independent document summary.

REFERENCES

[1] H. Abu-Salem, M. Al-Omari and M. Evens, Stemming methodologies over individual query words for arabic information retrieval, *Journal of the American Society for Information Science and Technology*, vol.50, no.6, pp.524-529, 1999.
[2] M. Abulaish and L. Dey, Interoperability among distributed overlapping ontologies – A fuzzy ontology framework, *Proc. of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence*, 2006.
[3] H. Alani, S. Kim, D. E. Millard, M. J. Weal, W. Hall, P. H. Lewis and N. R. Shadbalt, Automatic ontology based knowledge extraction from web documents, *IEEE Intelligent System*, vol.18, no.1, pp.14-21, 2003.
[4] I. Al-Kharashi and M. Evens, Comparing words, stems, and roots as index terms in an arabic information retrieval system, *Journal of the American Society for Information Science and Technology*, vol.45, no.8, pp.548-560, 1994.
[5] E. Antworth, PC-KIMMO: A two-level processor for morphological analysis, *Occasional Publications in Academic Computing*, Summer Institute of Linguistics, Dallas, TX, 1990.
[6] E.-S. Atlam, K. Morita, M. Fuketa and J. Aoe, A new method for selecting English compound terms and its knowledge representation, *Information Processing Management Journal*, vol.38, no.6, pp.807-821, 2002.

[7] E.-S. Atlam, M. Fuketa, K. Morita and J. Aoe, Document similarity measurement using field association terms, *Information Processing & Management Journal*, vol.39, no.6, pp.809-824, 2003.
[8] E.-S. Atlam, G. Elmarhomy, M. Fuketa, K. Morita and J. Aoe, Automatic building of new field association word candidates using search engine, *Information Processing & Management Journal*, vol.42, no.4, pp.951-962, 2006.
[9] E.-S. Atlam, M. Fuketa, K. Morita and J. Aoe, A new approach for Arabic text classification using Arabic field association terms, *Journal of the American Society for Information Science and Technology*, vol.62, no.11, pp.2266-2276, 2011.
[10] T. Berners-Lee, *Semantic Web Road Map, W3C Design Issues*, http://www.w3.org/DesignIssues/Semantic.html, 1998.
[11] S. Calegari and C. Davide, Towards a fuzzy ontology definition and a fuzzy extension of an ontology editor, *Enterprise Information Systems Lecture Notes in Business Information Processing*, vol.3, pp.147-158, 2008.
[12] D. W. Embley, R. D. Campkll, D. M. Smith and S. W. Liddle, Ontology based extraction and structuring of information from data-rich unstructured documents, *Proc. of ACM Conference on Information and Knowledge Management*, USA, pp.52-59, 1998.
[13] D. Fensel, F. van Harmelen, I. Horrocks, D. L. McGuinness and P. F. Patel-Schneider, OIL: An ontology infrastructure for the semantic web, *IEEE Intelligent Systems*, vol.16, no.2, pp.38-45, 2001.
[14] L. Figueiredo, J. A. Isabel, M. Tenreiro, R. Jose and J. L. Martins de Carvalho, Towards the development of intelligent transportation systems, *Proc. of IEEE Intelligent Transportation Systems Conference*, USA, pp.1206-1211, 2001.
[15] M. Fuketa, S. Lee, T. Tsuji, M. Okada and J. I. Aoe, A document classification method by using field association words, *Information Science*, vol.126, pp.57-70, 2000.
[16] N. Guarino, C. Masolo and G. Vetere, OntoSeek content-based access to the web, *IEEE Intelligent Sysem.*, vol.14, no.3, pp.70-80, 1999.
[17] H. V. Halteren, New feature sets for summarization by sentence extraction, *IEEE Intelligent Systems*, vol.18, no.4, pp.34-42, 2003.
[18] I. Hmeidi, G. Kanaan and M. Evens, Design and implementation of automatic indexing for information retrieval with arabic documents, *Journal of the American Society for Information Science and Technology*, vol.48, no.10, pp.867-881, 1997.
[19] I. Manzour and L. Al-Arab, *http://www.muhaddith.org/*, 2006.
[20] A. Ibrahim, *Fuzzy Logic for Embedded Systems Applications*, Newnes, 2006.
[21] D. Jurafsky and J. Martin, *Speech and Language Processing*, Prentice Hall, Saddle River, NJ, 2000.
[22] D. Z. Kang, B. W. Xu, J. J. Lu and Y. H. Li, Description logics for fuzzy ontologies on semantic web, *Journal of Southeast University*, vol.22, no.3, pp.343-347, 2006.
[23] S. Khoja, *APT: An Automatic Arabic Part-of-Speech Tagger*, Computing Department, Lancaster University, Lancaster, UK, 2003.
[24] G. Kiraz, Arabic computational morphology in the west, *Proc. of the 6th International Conference and Exhibition on Multi-Lingual Computing*, Cambridge, 1998.
[25] W. Lam and K. S. Ho, FIDS: An intelligent financial web news articles digest system, *IEEE Transactions on SMC*, vol.31. no.6, pp.753-762, 2001.
[26] C. S. Lee, Y. J. Chen and Z. W. Jian, Ontology based fuzzy event extraction agent for Chinese e-news summarization, *Expert Systems with Applications*, vol.25, no.3, pp.431-447, 2003.
[27] C. S. Lee, Z. W. Jian and L. K. Huang, A fuzzy ontology and its application to news summarization, *IEEE Transactions on Systems, Man and Cybernetics (Part B)*, vol.35, no.5, pp.859-880, 2005.
[28] J. Lee, S. Park, C. Ahna and D. Kim, Automatic generic document summarization based on non-negative matrix factorization, *Information Processing and Management*, vol.45, pp.20-34, 2009.
[29] L. Mani, Recent developments in text summarization, *Proc. of the 10th International Conference on Information and Knowledge Management*, pp.529-531, 2001.
[30] R. Navigli and P. Velardi, Ontology learning and its application to automated terminology translation, *IEEE Intell. Syst.*, vol.18, no.1, pp.22-31, 2003.
[31] D. Parry, A fuzzy ontology for medical document retrieval, *Proc. of the 2nd Workshop on Australasian Information Security, Data Mining and Web Intelligence, and Software Internationalization*, Dunedin, New Zealand, pp.121-126, 2004.
[32] D. Parry, Fuzzy ontologies for information retrieval on the WWW, in *Fuzzy Logic and the Semantic Web (Capturing Intelligence)*, E. Sanchez (ed.), Elsevier, 2006.

[33] T. T. Quan, S. C. Hui, A. C. M. Fong and T. H. Cao, Automatic fuzzy ontology generation for semantic web, *IEEE Transactions on Knowledge and Data Engineering*, vol.18, no.6, pp.842-856, 2006.

[34] E. Sanchez and T. Yamanoi, Fuzzy ontologies for the semantic web, *The 7th International Conference on Flexible Query Answering Systems*, Milan, Italy, 2006.

[35] J. F. Smith and R. Rhyne, A resource manager for distributed resources: Fuzzy decision trees and genetic optimization, *Proc. of the International Conference on Artificial Intelligence*, vol.2, pp.669-675, 1999.

[36] V. W. Soo and C. Y. Lin, Ontology-based information retrieval in a multi-agent system for digital library, *Proc. of the 6th Conf. Artificial Intell. Applicat*, pp.241-246, 2001.

[37] A. Soudi, A. van den Bosch and G. Neumann, *Arabic Computational Morphology*, Springer, 2007.

[38] Q. T. Tho, S. C. Hui, A. C. Fong and T. H. Cao, Automatic fuzzy ontology generation for semantic web, *IEEE Transactions on Knowledge and Data Engineering*, vol.18, no.6, pp.842-856, 2006.

[39] T. Tsuji, H. Nigazawa, M. Okada and J. Aoe, Early field recognition by using field association words, *Proc. of the 18th International Conference on Computer Processing of Oriental Language*, vol.2, pp.301-304, 1999.

[40] T. Tsuji, M. Fuketa, K. Morita and J. Aoe, An efficient method of determining field association terms of compound words, *Journal of Natural Language Processing*, vol.7, no.2, pp.3-26, 2000.

[41] A. Ubul, E.-S. Atlam, H. Kitagawa, M. Fuketa, K. Morita and J. Aoe, An efficient method of summarizing documents using impression measurements, *Computing and Informatics*, vol.32, no.2, 2013.

[42] D. H. Widyantoro and J. Yen, A fuzzy ontology based abstract search engine and its user studies, *Proc. of the 10th IEEE International Conference on Fuzzy Systems*, Melbourne, Australia, pp.1291-1294, 2001.

[43] L. A. Zadeh and J. Kacprzyk, *Fuzzy Logic for the Management of Uncertainty*, John Wiley & Sons, Inc, 1992.

[44] J. Zhai and Y. Chen, Using ontology to achieve the semantic integration of the intelligent transport system, *Proc. of the 12th International Conference on Management Science & Engineering*, vol.3, pp.2528-2532, 2005.