A Fuzzy Ontology and Its Application to News Summarization

Chang-Shing Lee, Zhi-Wei Jian, and Lin-Kai Huang

***Abstract*—In this paper, a fuzzy ontology and its application to news summarization are presented. The fuzzy ontology with fuzzy concepts is an extension of the domain ontology with crisp concepts. It is more suitable to describe the domain knowledge than domain ontology for solving the uncertainty reasoning problems. First, the domain ontology with various events of news is predefined by domain experts. The *document preprocessing mechanism* will generate the meaningful terms based on the news corpus and the *Chinese news dictionary* defined by the domain expert. Then, the meaningful terms will be classified according to the events of the news by the *term classifier*. 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 degrees associated with various events of the domain ontology. In addition, a news agent based on the fuzzy ontology is also developed for news summarization. The news agent contains five modules, including a *retrieval agent*, a *document preprocessing mechanism*, a *sentence path extractor*, a *sentence generator*, and a *sentence filter* to perform news summarization. Furthermore, we construct an experimental website to test the proposed approach. The experimental results show that the news agent based on the fuzzy ontology can effectively operate for news summarization.**

***Index Terms*—Chinese natural language processing, fuzzy inference, news agent, news summarization, ontology.**

I. Introduction

AN ontology is a formal conceptualization of a real world, and it can share a common understanding of this real world [11]. With the support of the ontology, both user and system can communicate with each other by the shared and common understanding of a domain [16]. There are many ontological applications that have been presented in various domains. For example, Embley *et al.* [3] present a method of extracting information from unstructured documents based on an application ontology. Alani *et al.* [1] propose the Artequakt that automatically extracts knowledge about artists from the web based on an ontology. It can generate biographies that tailor to a user's interests and requirements. Navigli *et al.* [14] propose the OntoLearn with ontology learning capability to extract relevant domain terms from a corpus of text. OntoSeek [4] is a system designed for content-based information retrieval. It combines an ontology-driven content-matching mechanism with moderately expressive representation formalism. Handschuh *et al.* [6] provide a framework, known as S-CREAM, that allows for creation of metadata and is trainable for a specific domain. Ont-O-Mat is the reference implementation of the S-CREAM framework. It provides a plugin interface for extensions for further advancements, e.g., collaborative metadata creation or integrated ontology editing and evolution. Vargas-Vera *et al.* [17] present an annotation tool, called MnM, which provides both automated and semi-automated support for annotating web pages with semantic contents. MnM integrates a web browser with an ontology editor and provides open application programming interfaces (APIs) to link to ontology servers and for integrating information extraction tools.

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. McKeown *et al.* [13] develop a composite summarization system that uses different summarization strategies dependent on the type of documents in the input set. Mani [12] introduces two text summary approaches, including topic-focused and generic summary. Lam *et al.* [9] propose the Financial Information Digest System (FIDS) to summarize online financial news articles automatically. The FIDS can integrate the information from different articles by conducting automatic content-based classification and information item extraction. Halteren [5] develops a new technique to find out several feature sets through the set of style makers and extract summarization sentences from the document, according to the found feature sets for users. Lee *et al.* [8] propose an ontology-based fuzzy event extraction agent for Chinese news summarization. The summarization agent can generate a sentence set for each Chinese news, but the results need to be improved to suit for various Chinese news websites. Riloff [15] proposes the AutoSlog system that can automatically create domain-specific dictionaries for information extraction by an appropriate training corpus.

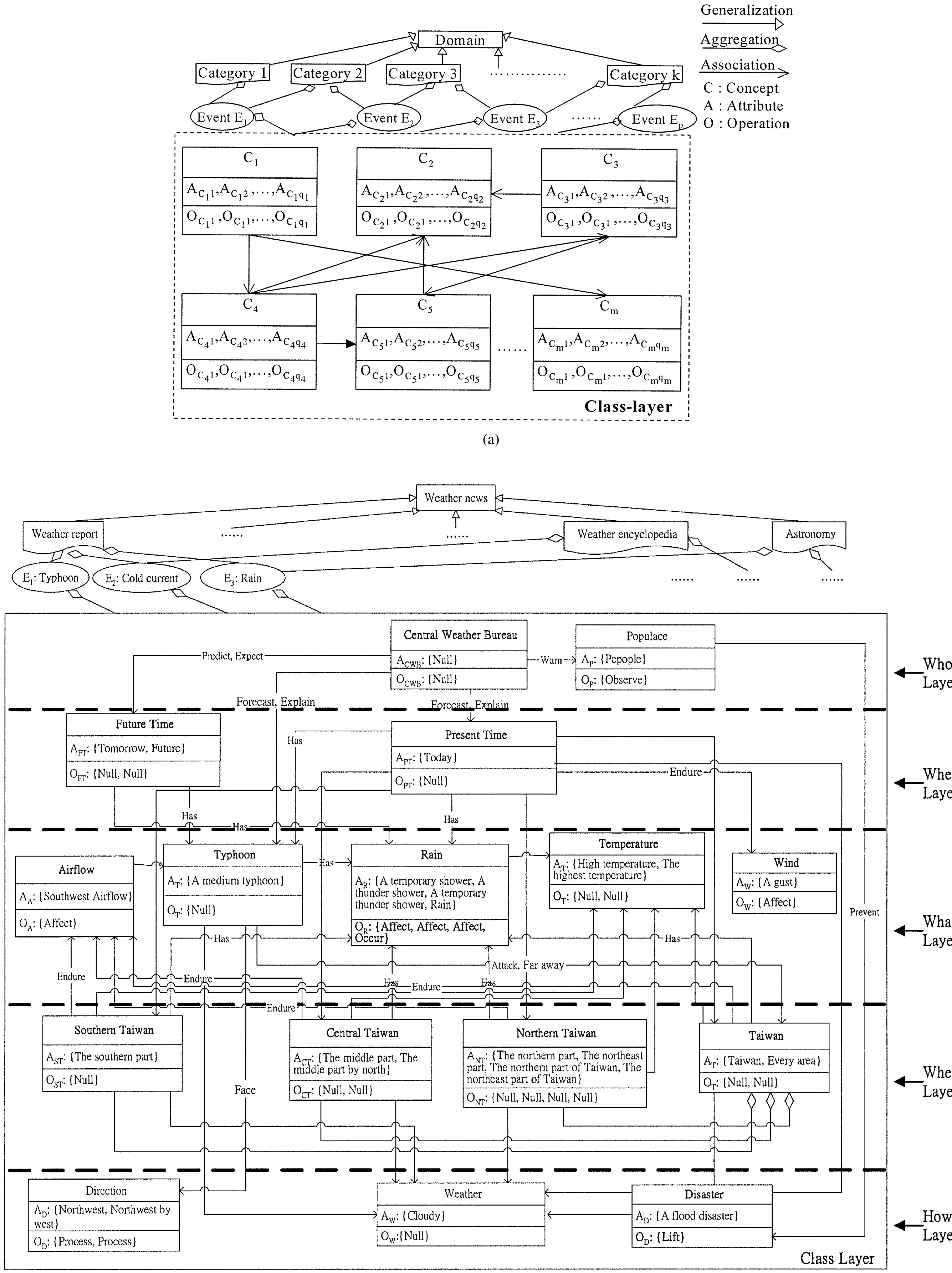
In this paper, we present a fuzzy ontology and apply it to news summarization. The fuzzy ontology is an extension of the domain ontology that is more suitable to describe the domain knowledge for solving the uncertainty reasoning problems. In addition, a news agent based on the fuzzy ontology is also developed for news summarization. The remainder of this paper is structured as follows: Section II describes the definition of the fuzzy ontology. Section III introduces a fuzzy inference mechanism for constructing the fuzzy ontology. Section IV proposes a news agent based on the fuzzy ontology for news summarization. Section V shows some experimental results for Chinese news summarization. Section VI presents our conclusions and the future work.

Manuscript received June 11, 2004; revised October 10, 2004 and January 9, 2005. This work was supported in part by the National Science Council of Taiwan (R.O.C.) under Grant NSC-93-2213-E-309-003 and by the Ministry of Economic Affairs in Taiwan under Grant 93-EC-17-A-02-S1-029. This paper was recommended by Associate Editor Hisao Ishibuchi.

C.-S. Lee and L.-K. Huang are with the Department of Information Management, Chang Jung Christian University, Tainan, Taiwan, R.O.C. (e-mail: leecs@mail.cju.edu.tw; leecs@cad.csie.ncku.edu.tw).

Z.-W. Jian is with the Department of Computer Science and Information Engineering, National Cheng Kung University, Tainan, Taiwan, R.O.C.

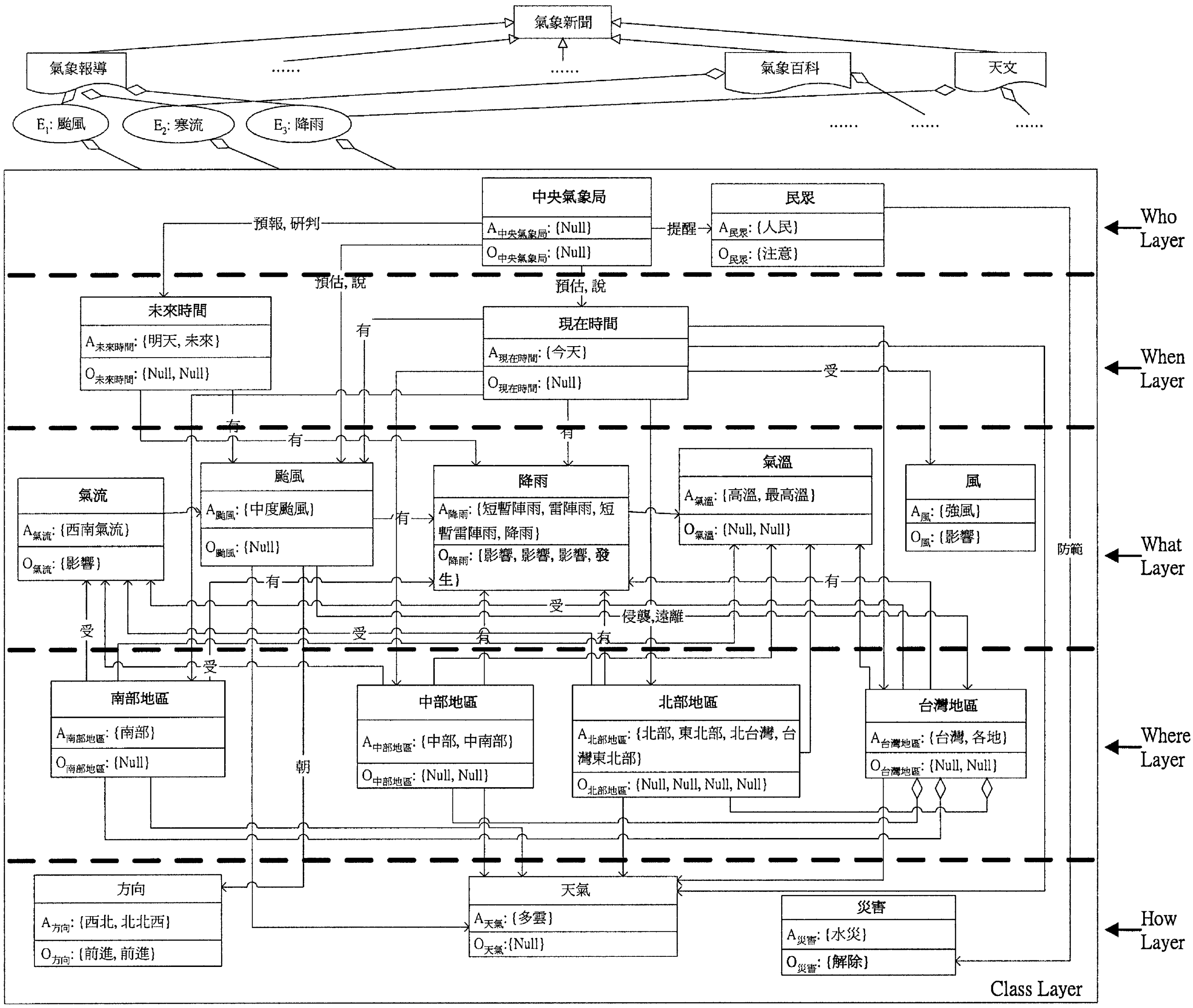
Digital Object Identifier 10.1109/TSMCB.2005.845032



(a)

(b)

Fig. 1. (a). Structure of the domain ontology. (b) Example of the domain ontology for weather news domain.



(c)

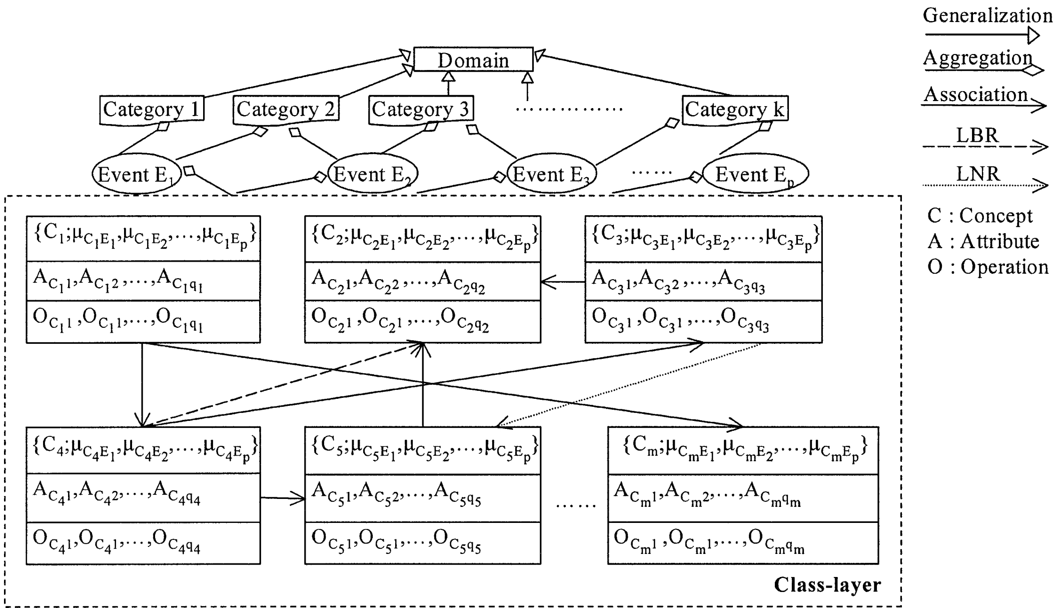
Fig. 1. (*Continued*) (c) Corresponding Chinese version of Fig. 1(b).

II. Definition of the Fuzzy Ontology

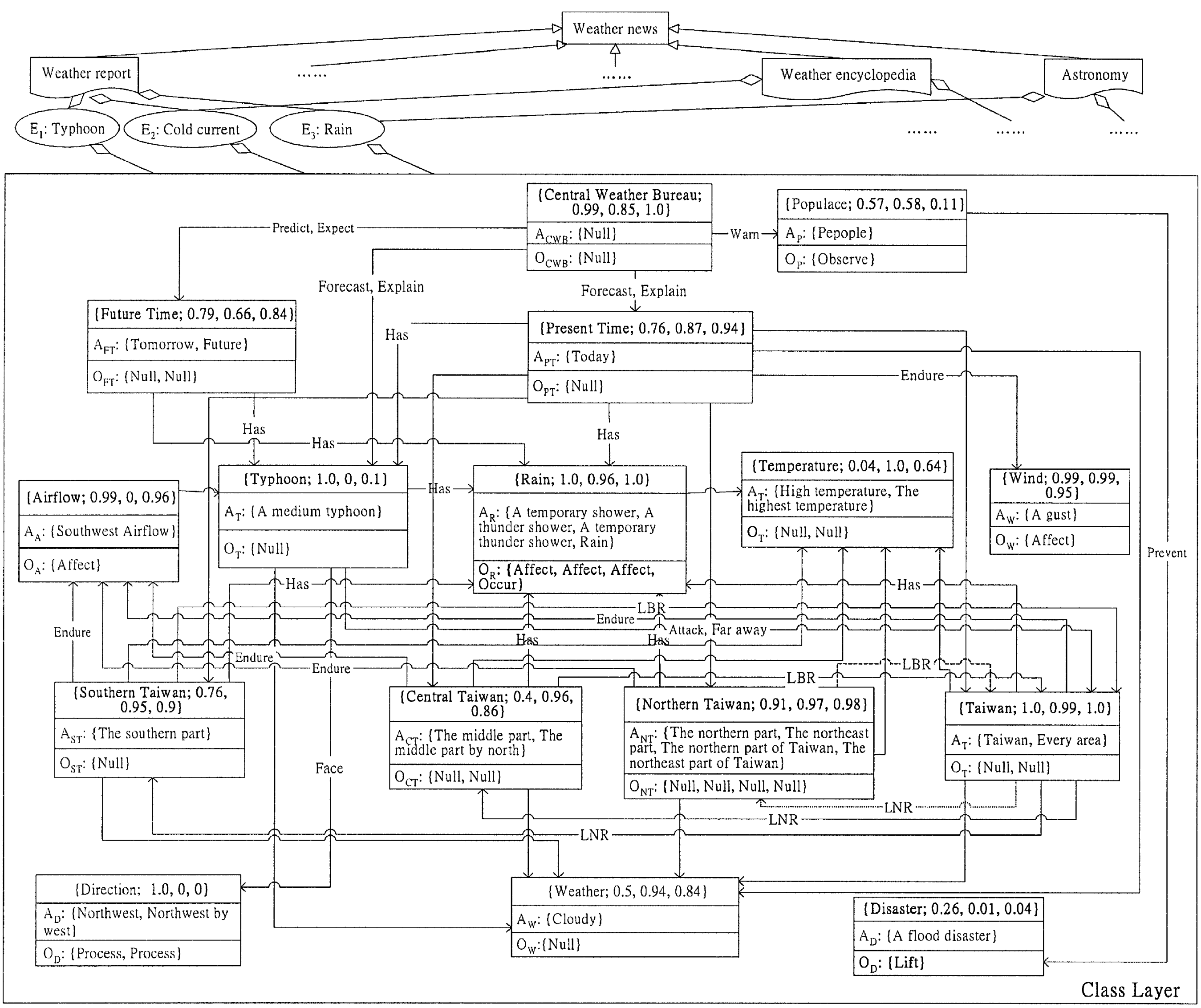
In our previous work [8], we have extended the three-layered object-oriented ontology architecture to the four-layered model for domain ontology representation. The performances of the experimental results [8] are effective, so we still adopt the four-layered model in this paper. Fig. 1(a) and (b) show the four-layered structure of the domain ontology and an example of weather news domain ontology, respectively. In Fig. 1(b), the domain name is "Weather news," and it consists of several categories such as "Weather report," "Weather encyclopedia," and "Astronomy." In addition, there are several events defined in this domain ontology. For example, the weather events "E_1: Typhoon," "E_2: Cold current," and "E_3: Rain" are related to the categories "Weather report," "Weather encyclopedia," and "Astronomy." Each event comprises several concepts, such as "Central Weather Bureau," "Present Time," "Temperature," and "Weather." Fig. 1(c) shows the corresponding Chinese version of Fig. 1(b).

In this section, we further present a fuzzy ontology for news summarization. Because the fuzzy ontology is an extension of the domain ontology, we briefly introduce the domain ontology as follows. There are many different descriptions of the domain ontology for various applications. Now, we give a formal definition of the domain ontology for news summarization.

[Definition 1] Domain Ontology: A domain ontology defines a set of representational terms that we call concepts. Inter-relationships among these concepts describe a target world [11]. There are four layers, including *domain layer*, *category layer*, *event layer*, and *class layer*, defined in the domain ontology [8]. The *domain layer* represents the domain name of an ontology and consists of various categories defined by domain experts. Each category is composed of the event set $\{E_1, E_2, \ldots, E_p\}$, which is derived from the news corpus by domain experts. Every event comprises several concepts of class layer. In *class layer*, each concept contains a concept name C_i, an attribute set $\{A_{C_i 1}, \ldots, A_{C_i q_i}\}$, and an operation set $\{O_{C_i 1}, \ldots, O_{C_i q_i}\}$ for an application domain. There are

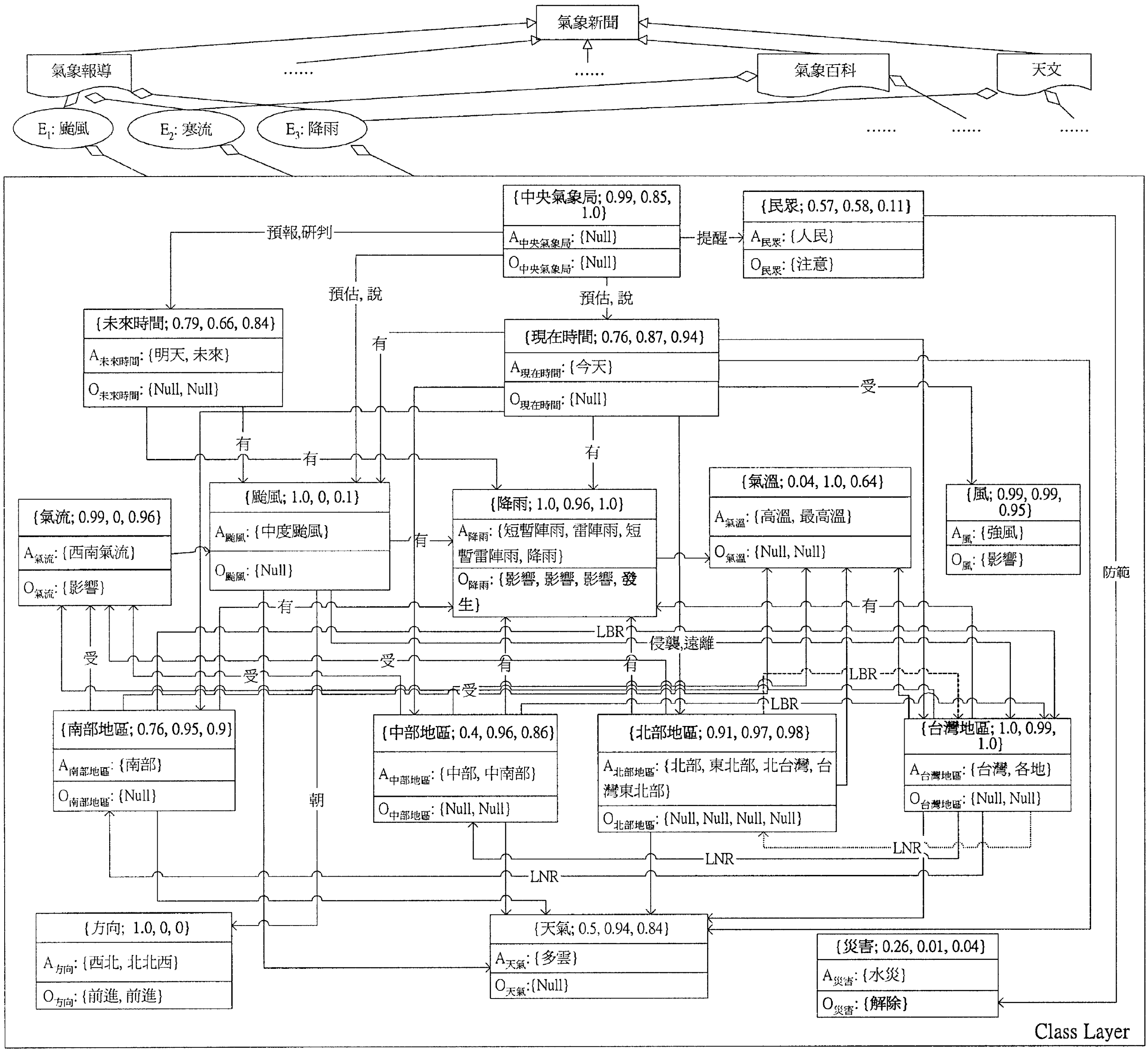


(a)



(b)

Fig. 2. (a) Structure of the fuzzy ontology. (b) Example of the fuzzy ontology for weather news domain.



(c)

Fig. 2. (*Continued*). (c) Corresponding Chinese version of Fig. 2(b).

three kinds of relationship, including *generalization*, *aggregation*, and *association*, in the domain ontology. The relationship between a domain and its corresponding category is *generalization* that represents "*is-kind-of*" relationship. The relationship between each category and its corresponding events is *aggregation*. The *aggregation* denotes "*is-part-of*" relationship. The *association* represents a semantic relationship between concepts in class layer. ■

In this paper, we extend the domain ontology to be the fuzzy ontology by embedding a set of membership degrees in each concept of the domain ontology and adding two fuzzy relationships among the fuzzy concepts. The concept with the membership degrees is called *fuzzy concept*. Now, we give the definitions of *fuzzy concept*, *fuzzy relationship*, and *fuzzy ontology* as follows.

[Definition 2] Fuzzy Concept: A *fuzzy concept* is a refined concept derived from a domain ontology. It is a refinement by embedding a set of membership degrees associated with a set of the news events in the concept of the domain ontology. If a domain ontology has an event set $\{E_1, E_2, \ldots, E_p\}$ and a concept C_i, then we can refine the C_i into the *fuzzy concept* and denote the *fuzzy concept* as $\{C_i; \mu_{C_iE_1}, \mu_{C_iE_2}, \ldots, \mu_{C_iE_p}\}$ with an attribute set $\{A_{C_i1}, A_{C_i2}, \ldots, A_{C_iq_i}\}$ and an operation set $\{O_{C_i1}, O_{C_i2}, \ldots, O_{C_iq_i}\}$, where $\mu_{C_iE_j}$ represents the membership degree of C_i for event E_j. Besides, the attribute $A_{C_iq_i}$ and the operation $O_{C_iq_i}$ denote the q_ith attribute and q_ith operation of C_i, respectively. ■

[Definition 3] Fuzzy Relationship: There are two kinds of *fuzzy relationships*, including *Location_Narrower_Relationship* (*LNR*) and *Location_Broader_Relationship* (*LBR*), defined for news summarization. If there exist two *fuzzy concepts* (location L_1 and loca- tion L_2 and $L_1 \supset L_2$), then we denote the *fuzzy relationships LNR* and *LBR* as $L_1 \xrightarrow{LNR} L_2$ and $L_1 \xleftarrow{LBR} L_2$, respectively. ■

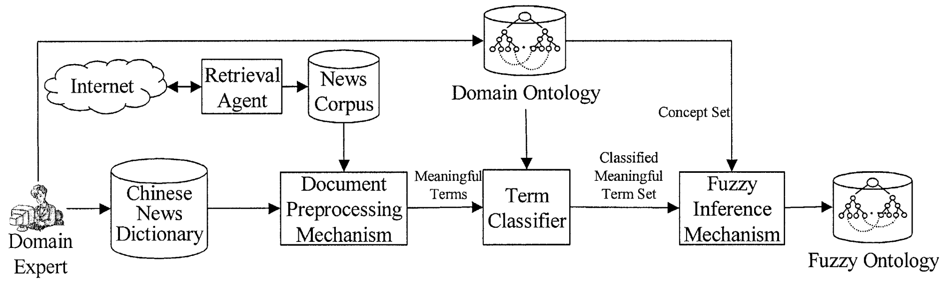


Fig. 3. Process of the fuzzy ontology construction.

[Definition 4] Fuzzy Ontology: A *fuzzy ontology* is an extended domain ontology with *fuzzy concepts* and *fuzzy relationships*. ■

Fig. 2(a) and (b) show the structure of the fuzzy ontology and an example of weather news fuzzy ontology, respectively.

In Fig. 2(b), the domain name is "Weather news," and it consists of several categories such as "Weather report," "Weather encyclopedia," and "Astronomy." In addition, there are several events defined in this domain ontology. For example, the weather events "E_1: Typhoon," "E_2: Cold current," and "E_3: Rain" are related to the categories "Weather report," "Weather encyclopedia," and "Astronomy," respectively. Each event comprises several *fuzzy concepts*, such as "{Central Weather Bureau; 0.99, 0.85, 1.0}," "{Present Time; 0.76, 0.87, 0.94}," "{Temperature; 0.04, 1.0, 0.64}," and "{Weather; 0.5, 0.94, 0.84}." For example, the membership degrees of the *fuzzy concept* "Temperature" for weather events "E_1: Typhoon," "E_2: Cold current," and "E_3: Rain" are 0.04, 1.0, and 0.64, respectively. Fig. 2(c) shows the corresponding Chinese version of Fig. 2(b).

III. Fuzzy Inference Mechanism for Fuzzy Ontology Construction

In this section, we apply a fuzzy inference mechanism to construct the fuzzy ontology. Fig. 3 shows the process of the fuzzy ontology construction.

In Fig. 3, the news domain ontology with various events is predefined by domain experts. The *document preprocessing mechanism* will generate the meaningful terms based on the news corpus produced by the *retrieval agent* and the *Chinese news dictionary* defined by domain experts. Then, the *term classifier* will classify the meaningful terms according to the events of the news. 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 degrees associated with various events of the domain ontology. Now, we briefly describe the *retrieval agent*, *document preprocessing mechanism*, and *term classifier* as follows.

A. Retrieval Agent, Document Preprocessing Mechanism, and Term Classifier

First, the *retrieval agent* periodically retrieves the news from one of the largest news Websites in Taiwan (http://www.chinatimes.com). The retrieved news will be stored into a news corpus for further processing. The *document preprocessing mechanism* consists of a part-of-speech (POS) tagger [19] provided by Chinese Knowledge Information Processing (CKIP) [19] Group in Taiwan and a term filter to get the meaningful term set including nouns and verbs. Table I shows a part of POS tags.

The term filter will preserve the meaningful terms according to the POS tags [8] selected by the domain experts based on the *Chinese news dictionary* [20] and the term frequencies of the news corpus. In addition, the meaningful terms will be classified according to the events of the news by the *term classifier*. Now, we describe the algorithm of *term classifier* as follows.

The algorithm of *Term Classifier*
Input:
A domain ontology and all meaningful terms of news generated by *document preprocessing mechanism*.
Output:
Classified meaningful term sets.
Method:
1) Parse and store all terms of various events of the domain ontology into E_p /* E_p denotes the term set of event p in the domain ontology */.
2) $T_{E_p} \leftarrow \phi$. /* T_{E_p} denotes the term set of the output event p */.
3) For $r \leftarrow 1$ to q /* q denotes the number of news */.
 3.1) For $i \leftarrow 1$ to m_r /* m_r denotes the number of all terms in the rth news */.
  3.1.1) For $p \leftarrow 1$ to n /* n denotes the number of events in the domain ontology */.
   3.1.1.1) If $t_{ri} \in E_p$ then /* t_{ri} denotes the ith term of input terms in the rth news */.
    $N_{E_p} \leftarrow N_{E_p} + 1$ /* N_{E_p} denotes the number of terms in the event p */.
 3.2) For $p \leftarrow 1$ to n
  3.2.1) If N_{E_p} is the largest, then
   3.2.1.1) For $i \leftarrow 1$ to m_r
    $T_{E_p} \leftarrow t_{ri}$.
4) End.

B. Fuzzy Inference Mechanism

In our previous work [8], we have presented a five-layer Fuzzy Inference Agent (FIA) [7], [10] for news event ontology

TABLE I
PART-OF-SPEECH TAGS OF CHINESE TERMS DERIVED FROM *CKIP*

POS Tag	Meaning	Example
Na	名詞(Noun)	颱風(Typhoon)
Naa	物質名詞(Mass noun)	水(Water)
Nab	個體名詞(Individual noun)	人(People)
Nac	可數抽象名詞(Countable abstract noun)	原因(Reason)
Nad	抽象名詞(Abstract noun)	智慧(Wisdom)
Nae	集合名詞 I(Collective noun I)	黨(Clique)
Naea	集合名詞 II(Collective noun II)	四肢(Limbs)
Naeb	集合名詞 III(Collective noun III)	車輛(Car)
Nb	專有名詞(Proper noun)	黃海(Huanghai Sea)
Nba	正式專有名詞(Formal Proper noun)	中國時報(Chinatimes)
Nbc	姓氏(Last name)	黃(Huang)
Nc	地方名詞(Place noun)	台南(Tainan)
Nca	專有地方名詞(Proper place noun)	台灣銀行(Taiwan bank)
Ncb	普通地方名詞(Common place noun)	學校(School)
Ncc	名方式地方名詞(Nominal-localizer place noun)	海外(Abroad)
Ncd	位置詞(Location noun)	東(East)
Ncda	單音節位置詞(Monosyllable location noun)	前(Front)
Ncdb	雙音節位置詞(Disyllabic location noun,)	中間(Middle)
Nce	定名式地方名詞(Determiner-nominal place noun)	本地(Local)
VA	動作不及物述詞 I(Active intransitive verb I)	入侵(Invade)
VA11	移動述詞(Motion predicate)	游泳(Swim)
VA12	姿態述詞(Posture predicate)	坐(Sit)
VA13	動作不及物述詞 II(Active intransitive verb II)	出國(Go abroad)
VA2	動作不及物述詞 III (Active intransitive verb, Theme III)	旋轉(Rotate)
VA3	氣象述詞(Climate predicate)	下雨(Rain)
VB	動作類單賓述詞 I(Active pseudo-transitive verb I)	延期(Postpone)
VB11	動作類單賓述詞 II(Active pseudo-transitive verb, Agent goal II)	幫(Help)
VB12	動作類單賓述詞 III(Active pseudo-transitive verb, Agent goal III)	朝(Face)
VB2	動作類單賓述詞 IV(Active pseudo-transitive verb, Agent theme IV)	給(Give)
VC	動作單賓述詞 I(Active transitive verb I)	檢查(Check)
VC1	動作單賓述詞 II(Active transitive verb, Theme goal II)	進入(Enter)
VC2	動作單賓述詞 III(Active transitive verb, Agent goal III)	破壞(Destroy)
VC31	動作單賓述詞 IV(Active transitive verb, Agent theme IV)	生產(Produce)
VC32	動作單賓述詞 V(Active transitive verb Agent theme V)	走私(Smuggle)
VC33	動作單賓述詞 VII(Active transitive verb, Agent theme VII)	儲存(Store)

extraction. In this paper, we modify the FIA and extend it to a seven-layer *Fuzzy Inference Mechanism* (*FIM*) for creating the fuzzy ontology. Layers 6 and 7 of *FIM* are utilized to compute and integrate the fuzzy membership degrees and then generate the fuzzy concepts for each event of the news. Fig. 4 shows the proposed *FIM*. There are seven layers, including *input linguistic layer*, *input term layer*, *rule layer*, *output term layer*, *output linguistic layer*, *summation layer*, and *integration layer* in this architecture.

Now, we describe each layer of *FIM* for event E_1 in detail.

1) Layer 1 (Input Linguistic Layer): The input linguistic layer contains two kinds of sets, including term set $T_{E_1 i}$ of events provided by *term classifier* and concept set C_j of domain ontology. The nodes in first layer just directly transmit input values to next layer. The input vectors are the term set and the concept set retrieved from *term classifier* and the domain ontology, respectively. The term set derived from *term classifier* is $(T_{E_1 1}, T_{E_1 2}, \ldots, T_{E_1 n})$, and the concept set of the domain ontology is $(C_1, C_2, \ldots, C_m)$.

2) Layer 2 (Input Term Layer): The *input term layer* performs the membership functions to compute the membership degrees for all terms derived from the retrieved news and the domain ontology. In this paper, the fuzzy number of L-R type [18] is adopted for the fuzzy variables and formulated by the following equation:

$$u(x) = \begin{cases} L\left(\frac{m-x}{\alpha}\right), & \text{for } x \le m \\ R\left(\frac{x-m}{\beta}\right), & \text{for } x \ge m \end{cases} \tag{1}$$

where $L(y) = R(y) = \max(0, 1-y)$, and $u(x)$ can be represented as a triplet $[m, \alpha, \beta]$.

There are three input fuzzy variables, including *Term Part-of-Speech* (*POS*), *Term Word* (*TW*) *similarity*, and *Semantic Distance* (*SD*) *similarity*, considered in *input term layer* for each term's property. The structure of 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 $(T_{E_1 1}, T_{E_1 2}, \ldots, T_{E_1 n})$ and $(C_1, C_2, \ldots, C_m)$ 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.

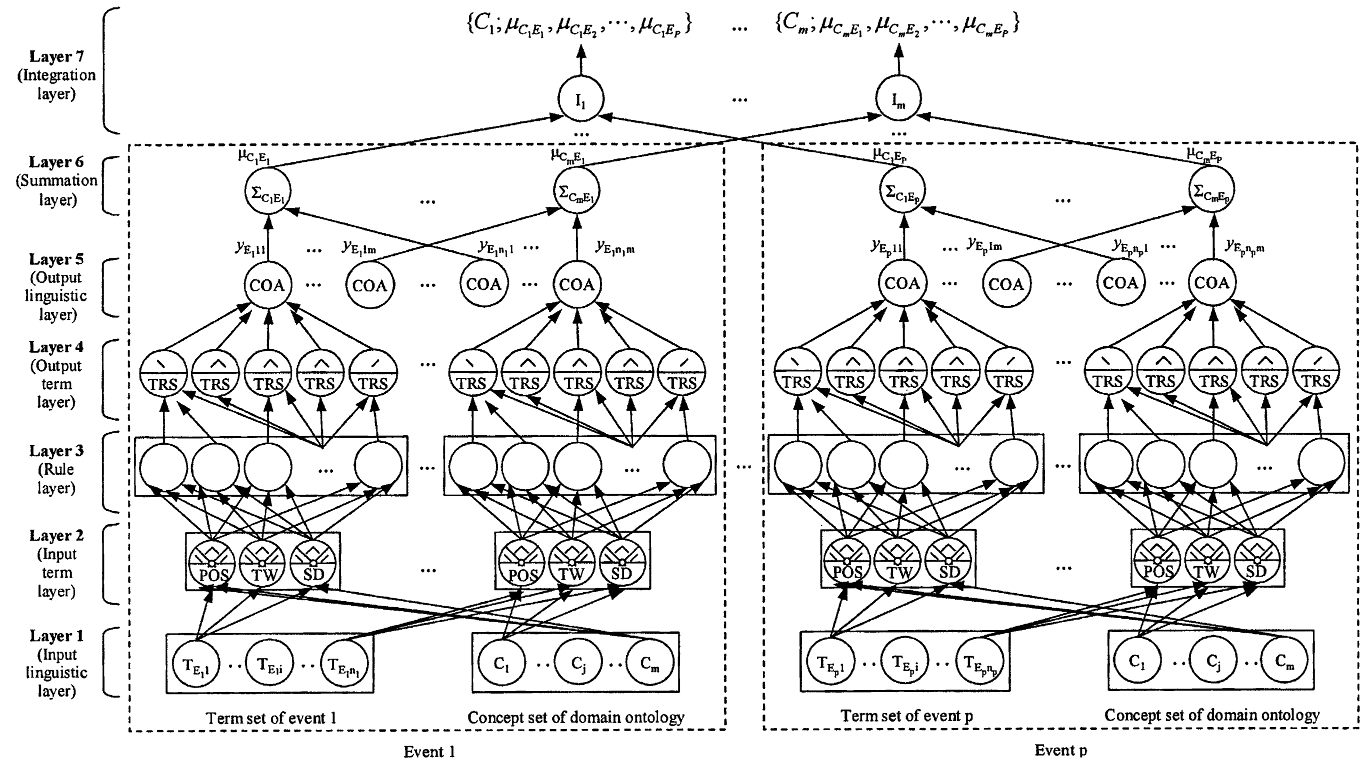


Fig. 4. Architecture of fuzzy inference mechanism.

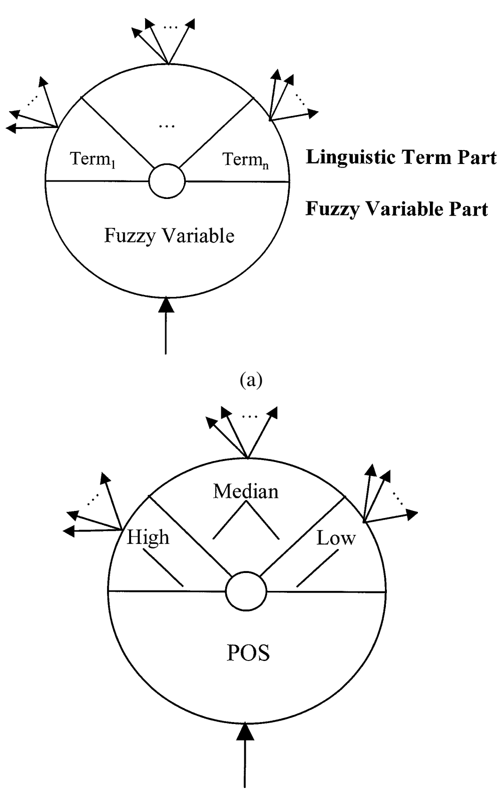


Fig. 5. (a). Structure of input term node. (b) Example of input term node for *Term POS* fuzzy variable.

Fig. 5(a) and (b) show the structure of the input term node and an example of the input term node for *Term POS* fuzzy variable with three linguistic terms, including *High*, *Median*, and *Low*, respectively.

In Fig. 6, each node of the tagging tree represents a part-of-speech of a term enumerated in the Table I. 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, 7]. For example, the two terms "颱風 (Typhoon, Na)" and "台灣 (Taiwan, Nc)" with their POS are "Na" and "Nc," respectively; hence, the path length of the two terms is 2 (Na $\rightarrow$ N $\rightarrow$ Nc). There are three linguistic terms, including *POS_High*, *POS_Median*, and *POS_Low* defined for *POS similarity*. Fig. 7 shows the membership functions of the fuzzy sets for *POS similarity*. The membership degree is usually a value in the range [0, 1], where "1" denotes a full membership, and "0" denotes no membership. Notice that domain experts predefine the membership functions of *POS_High*, *POS_Median*, and *POS_Low*. In this paper, the fuzzy numbers for *POS_High*, *POS_Median*, and *POS_Low* are $[m_{POS\_High}, \alpha_{POS\_High}, \beta_{POS\_High}]$, $[m_{POS\_Median}, \alpha_{POS\_Median}, \beta_{POS\_Median}]$, and $[m_{POS\_Low}, \alpha_{POS\_Low}, \beta_{POS\_Low}]$, respectively. Table II shows the parameter values of the fuzzy numbers for *POS similarity* defined by domain experts.

The second fuzzy variable proposed for computing the relation strength of any term pair is *Term Word* (*TW*) *similarity*. In this section, we will compute the number of the identical words that each term pair has to get *TW similarity*. In addition, because CKIP defines that the maximum number of words for any terms is 6, the bound of the number of the same word for any term pair is [0, 6]. For example,

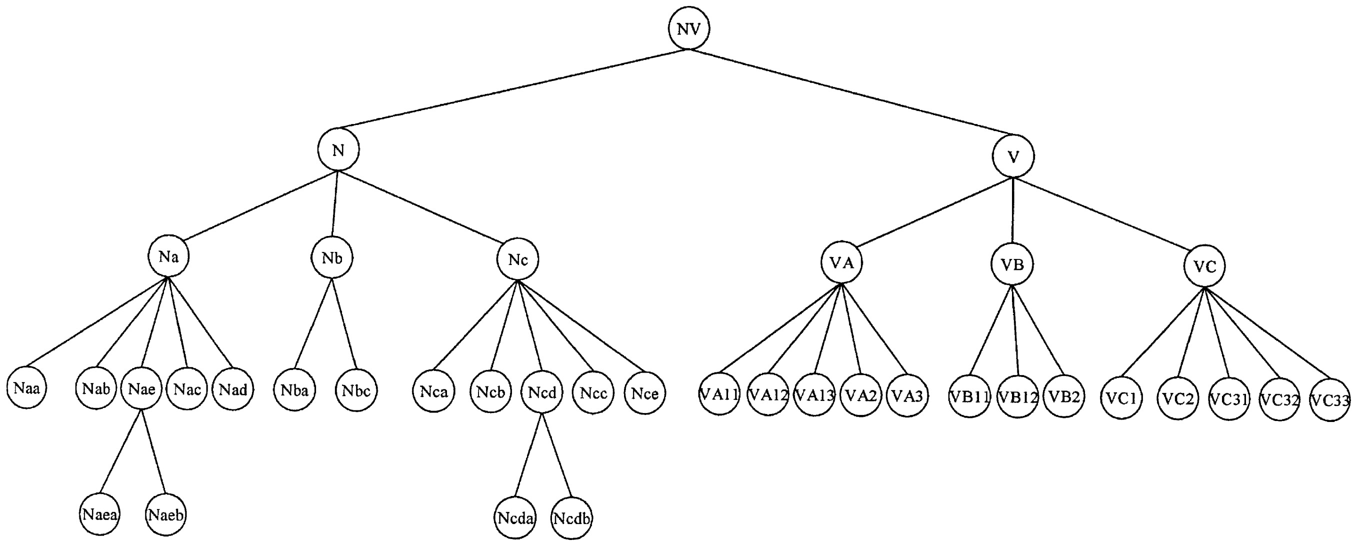


Fig. 6. Tagging tree derived from CKIP.

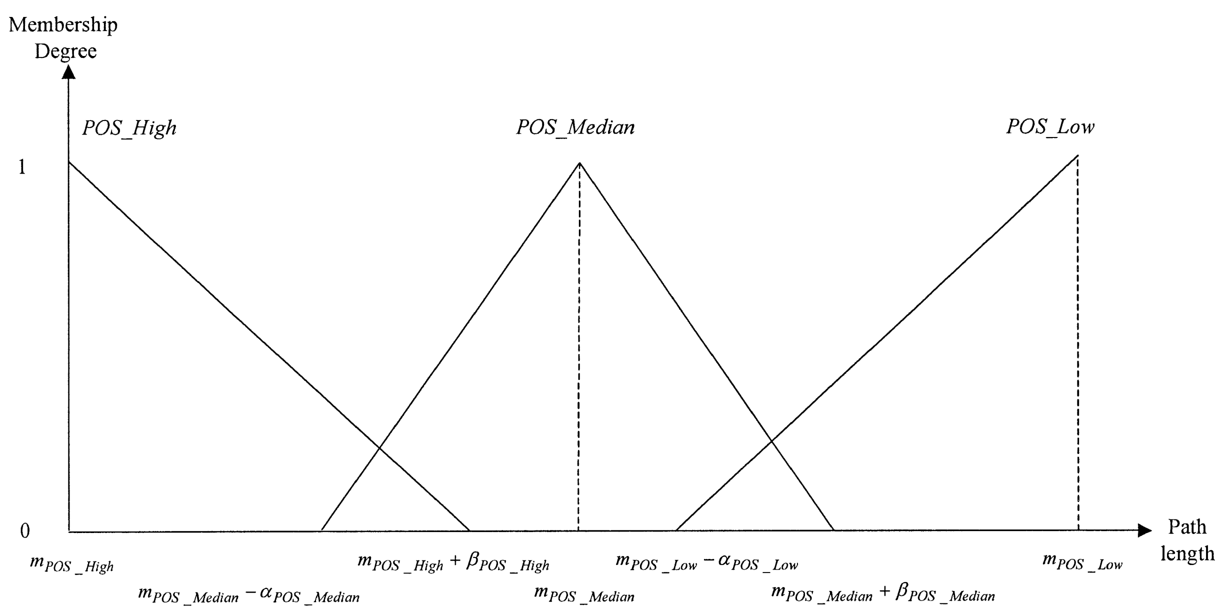


Fig. 7. Fuzzy sets of *POS similarity* for each term pair.

the term "行政院新聞局 (Government Information Office)" with six Chinese words "行," "政," "院," "新," "聞," and "局" is a term with the maximum number of words. Next, the two terms "中度颱風 (A medium typhoon)" and "輕度颱風 (A light typhoon)" have three of the same Chinese words "度," "颱," and "風" and hence, the number of the same words is 3. Table II shows the parameter values of the fuzzy numbers TW_Low, TW_Median, and TW_High for *TW similarity* fuzzy variable defined by domain experts.

The last fuzzy variable considered for input term pair is *Semantic Distance* (*SD*) *similarity*. In the *SD similarity* computation, the domain ontology will be defined as five layers, including the *who* layer, the *when* layer, the *what* layer, the *where* layer, and the *how* layer, by domain experts. For example, the two terms "氣流 (Airflow)" and "氣溫 (Temperature)" in Fig. 1(b) are both located in the same *what layer*; hence, the semantic distance is 2, shown at the bottom of the page. An example of the typhoon event ontology with the five layers is shown in Fig. 1(b). In this paper, we compute the semantic distance in the same layer of domain ontology for any term pair. Table II shows the parameter values of the fuzzy numbers for *SD similarity* fuzzy variable defined by domain experts.

Each fuzzy variable of the *input term layer* appearing in the premise part is represented with a condition node. Each of the outputs of the condition node is connected to rule nodes in the third layer to constitute a condition specified in some rules. This

("氣流 (Airflow)" → "颱風 (Typhoon)" → "降雨 (Rain)" → "氣溫(Temperature)")

TABLE II
PARAMETER VALUES OF THE FUZZY NUMBERS

Fuzzy Variable	Linguistic Term	m	α	β
POS	*POS_High*	0	0	2
	POS_Mediam	3	2	2
	POS_Low	7	2	0
TW	*TW_Low*	0	0	2
	TW_Mediam	3	2	2
	TW_High	6	2	0
SD	*SD_Low*	0	0	0.3
	SD_Median	0.5	0.3	0.3
	SD_High	1	0.3	0
TRS	*TRS_Very_Low*	0	0	0.3
	TRS_Low	0.2	0.3	0.5
	TRS_Median	0.3	0.5	0.7
	TRS_High	0.5	0.7	0.8
	TRS_Very High	0.7	1	0

TABLE III
FUZZY INFERENCE RULES FOR *FIM*

Rules \ Fuzzy variables	*POS Similarity*	*TW Similarity*	*SD Similarity*	*TRS*
1	L	L	L	VL
2	L	L	M	L
3	L	L	H	L
4	L	M	L	L
5	L	M	M	L
6	L	M	H	M
7	L	H	L	L
8	L	H	M	M
9	L	H	H	H
10	M	L	L	L
11	M	L	M	L
12	M	L	H	M
13	M	M	L	L
14	M	M	M	M
15	M	M	H	L
16	M	H	L	M
17	M	H	M	H
18	M	H	H	H
19	H	L	L	L
20	H	L	M	M
21	H	L	H	H
22	H	M	L	M
23	H	M	M	H
24	H	M	H	H
25	H	H	L	H
26	H	H	M	H
27	H	H	H	VH

VL:Very Low L:Low M:Median H:High VH:Very High

layer performs the first inference step to compute matching degrees. If the input vector of this layer is

$$[((T_{E_1 1-POS}, T_{E_1 1-TW}, T_{E_1 1-SD}) \ldots, (T_{E_1 n-POS}, T_{E_1 n-TW}, T_{E_1 n-SD})) ((C_{1-POS}, C_{1-TW}, C_{1-SD}) \ldots, (C_{m-POS}, C_{m-TW}, C_{m-SD}))]$$

then it will be transferred as follows:

$$((X_{11-POS}, X_{11-TW}, X_{11-SD}) (X_{12-POS}, X_{12-TW}, X_{12-SD}) \ldots, (X_{nm-POS}, X_{nm-TW}, X_{nm-SD})) \quad (2)$$

where X_{ij} represents the term pair $X_{ij} = (T_{E_1 i}, C_j)$, i is the ith term of the news for event 1, j is the jth concept of the domain ontology, X_{ij-POS} is the part-of-speech path length of each term pair, X_{ij-TW} represents the same word of each term pair, and X_{ij-SD} is the semantic distance of each term pair. Then, the output vector can be written as follows:

$$\begin{aligned}\mu^2 = &((u^2_{11-POS\_High}, u^2_{11-POS\_Median}, u^2_{11-POS\_Low}) \\ &(u^2_{11-TW\_Low}, u^2_{11-TW\_Median}, u^2_{11-TW\_High}) \\ &(u^2_{11-SD\_Low}, u^2_{11-SD\_Median}, u^2_{11-SD\_High}) \\ \ldots &(u^2_{n_1 m-POS\_High}, u^2_{n_1 m-POS\_Median}, u^2_{n_1 m-POS\_Low}) \\ &(u^2_{n_1 m-TW\_Low}, u^2_{n_1 m-TW\_Median}, u^2_{n_1 m-TW\_High}) \\ &(u^2_{n_1 m-SD\_Low}, u^2_{n_1 m-SD\_Median}, u^2_{n_1 m-SD\_High})) \quad (3)\end{aligned}$$

where $u^2_{ij-POS_k}$ is the membership degree of the kth linguistic term in the fuzzy variable of *POS*, $u^2_{ij-TW_k}$ is the membership degree of the kth linguistic term in the fuzzy variable of *TW*, and $u^2_{ij-SD_k}$ is the membership degree of the kth linguistic term in the fuzzy variable of *SD*.

3) Layer 3 (Rule Layer): The third layer is called the rule layer, where each node is a rule node to represent a fuzzy rule. The links in this layer are used to perform precondition matching of fuzzy logic rules. Hence, the rule nodes should perform the fuzzy AND operation [10], and the outputs will be linked with associated linguistic node in the fourth layer. In this paper, we use the algebraic product operation to compute the matching degree. In our model, the rules are defined by domain expert's knowledge previously, and we show them in Table III. Equation (4) shows the precondition matching degree of *rule node 1* as follows:

$$\mu^3_1 = \min(u^2_{11-POS\_Low}, u^2_{11-TW\_Low}, u^2_{11-SD\_Low}). \quad (4)$$

4) Layer 4 (Output Term Layer): The *output term layer* performs the fuzzy OR operation to integrate the fired rules that have the same consequences. The fuzzy variable defined in output layer is *Terms Relation Strength* (*TRS*). There are five linguistic terms, including *TRS_Very_Low*, *TRS_Low*, *TRS_Median*, *TRS_High*, and *TRS_Very_High* in *TRS*. Table II shows the parameter values of the fuzzy number for *TRS* fuzzy variable defined by domain experts.

For example, if there are r rules inferring the same consequences with the kth output linguistic term *Low*, and the corresponding gravity of linguistic term *Low* is V_{Low}, then the output of this layer for term pair X_{ij} is as follows:

$$\mu^4_{Low\_kij} = \frac{\sum_{p=1}^{r} \mu^3_{p\_kij} V_{Low_P\_kij}}{\sum_{p=1}^{r} \mu^3_{p\_kij}}. \quad (5)$$

5) Layer 5 (Output Linguistic Layer): The *output linguistic layer* performs the defuzzification process to get the *TRS* of the term pair. In this paper, the *center of area* (*COA*) method [10] is adopted to carry out the defuzzified process. Finally, the *TRS* value for term pair X_{ij} of event E_1 is as follows:

$$y_{E_1 ij} = \frac{\sum_{p=1}^{r} \sum_{q=1}^{c} \mu^4_{pq\_kij} V_{pq\_kij}}{\sum_{p=1}^{r} \sum_{q=1}^{c} \mu^4_{pq\_kij}} \quad (6)$$

where r is the number of rule nodes, c is the number of linguistic terms of output fuzzy variable, and V_{pq} is the gravity of qth output linguistic term associated with the pth rule node. Therefore, in our case, the values of r and c are 27 and 5, respectively.

6) Layer 6 (Summation Layer): The summation layer performs the summation process for computing the *TRS* values of the specific concept C_j of event 1. The input vector of this layer is as follows:

$$\{(y_{E_1 11}, \ldots, y_{E_1 1m}), (y_{E_1 21}, \ldots, y_{E_1 2m}), \ldots, (y_{E_1 n_1 1}, \ldots, y_{E_1 n_1 m})\} \quad (7)$$

where $y_{E_1 ij}$ is the *TRS* value of ith term $T_{E_1 i}$ and jth concept C_j for event 1. The summation process is as follows:

$$\mu_{C_j E_1} = \frac{\sum_{i=1}^{n_1} y_{E_1 ij}}{n_1}, \quad \text{for all } j = 1, \ldots, m. \quad (8)$$

7) Layer 7 (Integration Layer): Finally, the integration layer will integrate the membership degrees into the fuzzy ontology. The integration node of this layer will integrate the membership degrees of the concept C_j that belongs to all events of the domain ontology, that is, $(C_j; \mu_{C_j E_1}, \mu_{C_j E_2}, \ldots, \mu_{C_j E_p})$. Finally, the fuzzy ontology will be represented as follows:

$$\begin{aligned} \{&(C_1; \mu_{C_1 E_1}, \mu_{C_1 E_2}, \ldots, \mu_{C_1 E_p}) \\ &(C_2; \mu_{C_2 E_1}, \mu_{C_2 E_2}, \ldots, \mu_{C_2 E_p}) \\ &\vdots \\ &(C_m; \mu_{C_m E_1}, \mu_{C_m E_2}, \ldots, \mu_{C_m E_p})\} \end{aligned} \quad (9)$$

where the output vector $(C_j; \mu_{C_j E_1}, \mu_{C_j E_2}, \ldots, \mu_{C_j E_p})$ denotes the membership degree set of jth *fuzzy concept*.

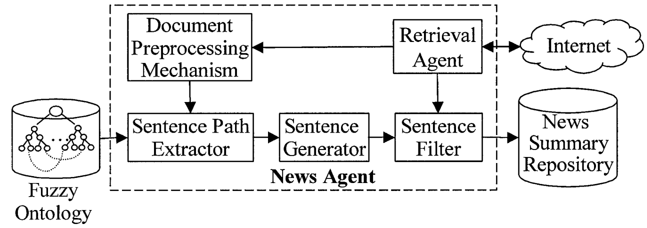


Fig. 8. Process of news agent for news summarization.

IV. Application of Fuzzy Ontology to Chinese News Summarization

In this section, we propose a news agent based on the fuzzy ontology for news summarization. In particular, we focus on Chinese news processing in this paper.

Fig. 8 shows the process of news agent for news summarization. The *retrieval agent* and *document preprocessing mechanism* perform the same tasks as we described in Section III. Moreover, the *Sentence Path Extractor* (*SPE*) will extract a set of possible sentence paths from the *fuzzy ontology* and send the template set to *Sentence Generator* (*SG*). The *SG* will produce a set of sentences from the class layer of the *fuzzy ontology*. Finally, the *Sentence Filter* (*SF*) will filter noisy sentences and create a set of summarized sentences.

A. Sentence Path Extractor

The *Sentence Path Extractor* uses the *Depth-First-Search* algorithm [2] to look for possible sentence paths from the fuzzy ontology. Now, we describe the algorithm as follows.

The algorithm of *Sentence Path Extractor*
Input:
All classified meaningful terms generated by *document preprocessing mechanism* terms of the news and all *fuzzy concepts* of the fuzzy ontology.
Output:
A set P of sentence paths.
Method:
1) $TFO \leftarrow \phi$. /* TFO represents a temporary fuzzy ontology */.
2) For all terms t generated by *document preprocessing mechanism*, and $(t \in C_i)$ or $(t = R_{C_iC_j})$ /* $R_{C_iC_j}$ is a relation between C_i and C_j^*/
 2.1) If $t \in C_i$ then
  2.1.1) If $C_i \notin TFO$ then
   Join C_i to TFO.
  2.1.2) Join t to TFO.
 2.2) If $t = R_{C_iC_j}$ then
  Join $R_{C_iC_j}$ to TFO.
3) For all concepts C_i of TFO
 3.1) For $a \leftarrow 0$ to m /* The m denotes the maximum indegree of the fuzzy concept in TFO*/

3.1.1) For $i \leftarrow 1$ to n /* The n denotes the number of the fuzzy concept in TFO*/

3.1.1.1) If (indegree of $C_i = a$) and (outdegree of $C_i > 0$) then

Join C_i into $C_{IN}[a]$ /* $C_{IN}[a]$ denotes a set of the fuzzy concept with indegree a*/.

4) $P \leftarrow \phi$. /* P denotes a set of the sentence path */.

5) For $a \leftarrow 0$ to m

5.1) For $k \leftarrow 1$ to n_a /* n_a denotes the element number of the concept set $C_{IN}[a]$*/

5.1.1) $C_x \leftarrow C_{IN\_k}[a]$.

5.1.2) While C_x isn't the leaf node of TFO

5.1.2.1) $P \leftarrow P \cup C_x$.

5.1.2.2) $C_x \leftarrow C_{IN\_k}[a]$.

5.1.3) For $e \leftarrow 1$ to N_{C_x} /* N_{C_x} denotes the number of the child nodes of C_x^*/

5.1.3.1) $C_x \leftarrow C_e$.

5.1.3.2) Go to While statement.

6) End.

The result of *Sentence Path Extractor* is denoted as follows:

Sentence Path P: [*Concept Name*] → [*Concept Name*] → ... → [*Concept Name*]

Fig. 9 shows an example of a temporary fuzzy ontology *TFO* extracted from Fig. 2(b). There are two possible sentence paths extracted by *SPE* shown as follows:

Sentence Path 1: "[*中央氣象局*]→[*未來時間*]→[*颱風*]→[*台灣地區*]"

Corresponding English Version:

"[*Central weather bureau*] → [*Future time*] → [*Typhoon*] → [*Taiwan*] "

and

Sentence Path 2: "[*中央氣象局*]→[*未來時間*]→[*降雨*]"

Corresponding English Version:

"[*Central weather bureau*] → [*Future time*] → [*Rain*]".

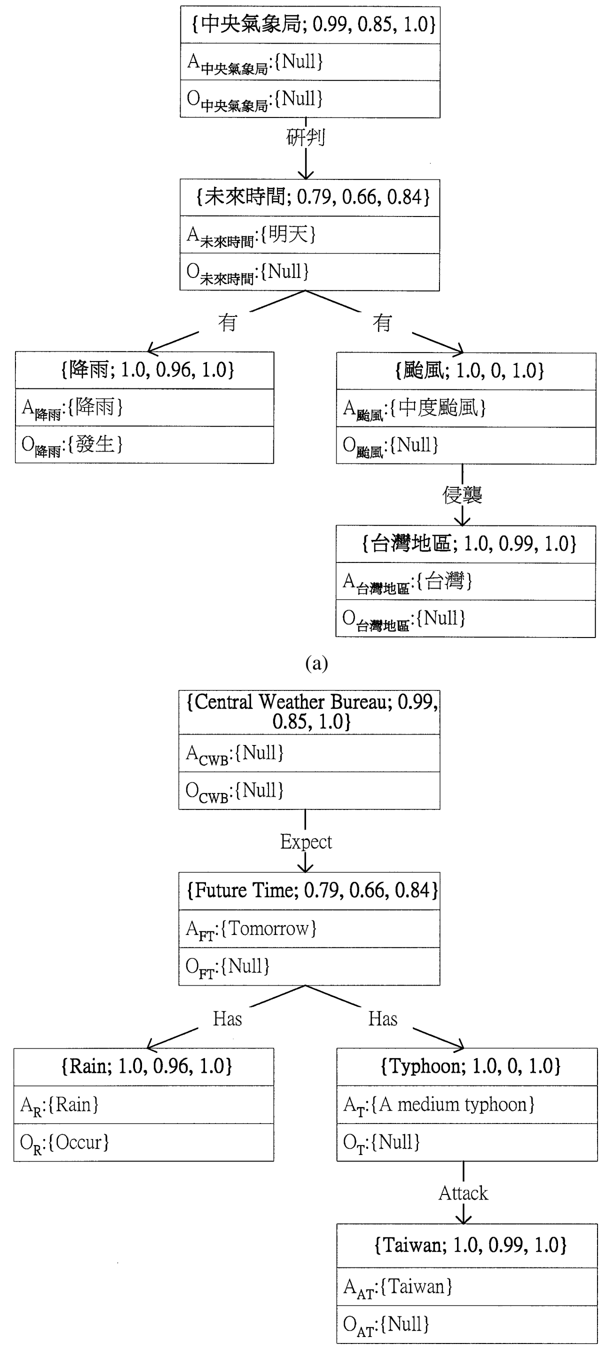


Fig. 9. (a) Temporary fuzzy ontology *TFO*. (b) Corresponding English version of Fig. 9(a).

B. Sentence Generator

The *Sentence Generator* generates a set of sentences based on the temporary fuzzy ontology. The algorithm of *Sentence Generator* is as follows:

The algorithm of *Sentence Generator*

Input:

A temporary fuzzy ontology TFO and a set of sentence paths.

Output:

A set of sentences.

Method:

1) For $i \leftarrow 1$ to m /* m denotes the number of sentence paths in the sentence path set P^*/.

1.1) Compute the number n_i of concepts in the sentence path P_i.

/* P_i denotes the ith sentence path of P^*/.

1.2) For $j \leftarrow 1$ to n_i

1.2.1) $w_i \leftarrow 1$. /* w_i denotes the possibility of the sentence S_i^*/.

1.2.2) If $j = 1$ then /* S_i denotes the ith sentence */

1.2.2.1) Get all information of concept C_j from the temporary fuzzy ontology.

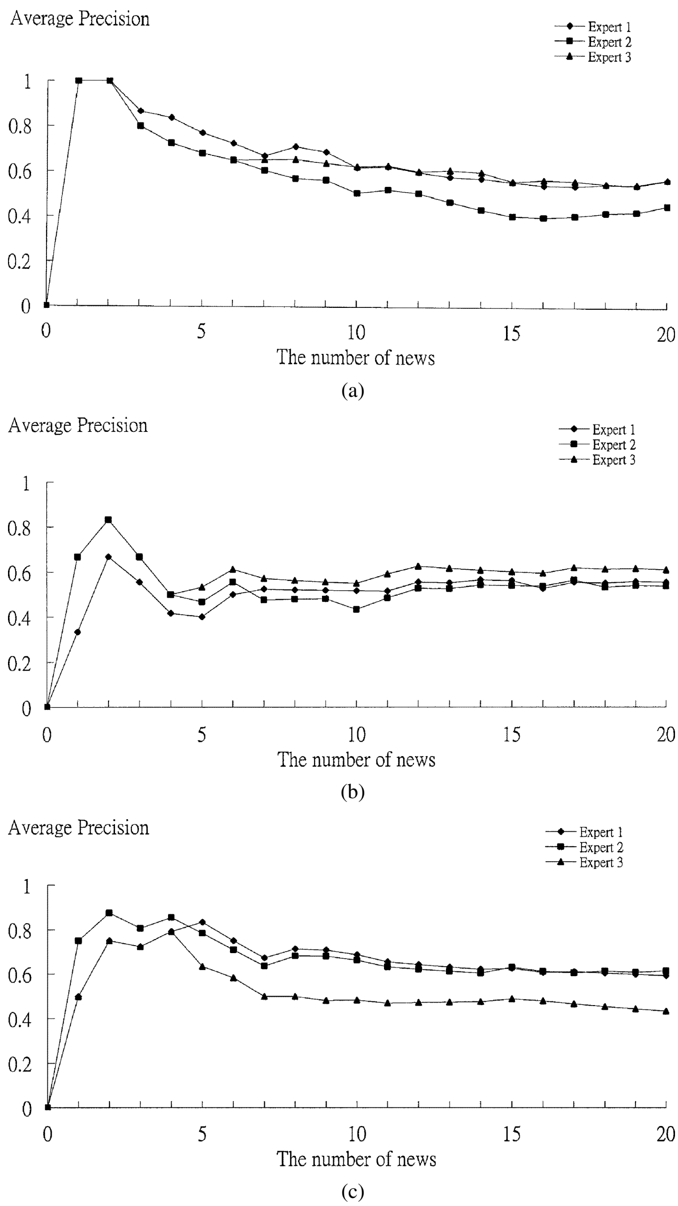


Fig. 10. Average precision of typhoon event evaluated by three domain experts. (a) Results of 2002 typhoon news. (b) Results of 2003 typhoon news. (c) Results of 2004 typhoon news.

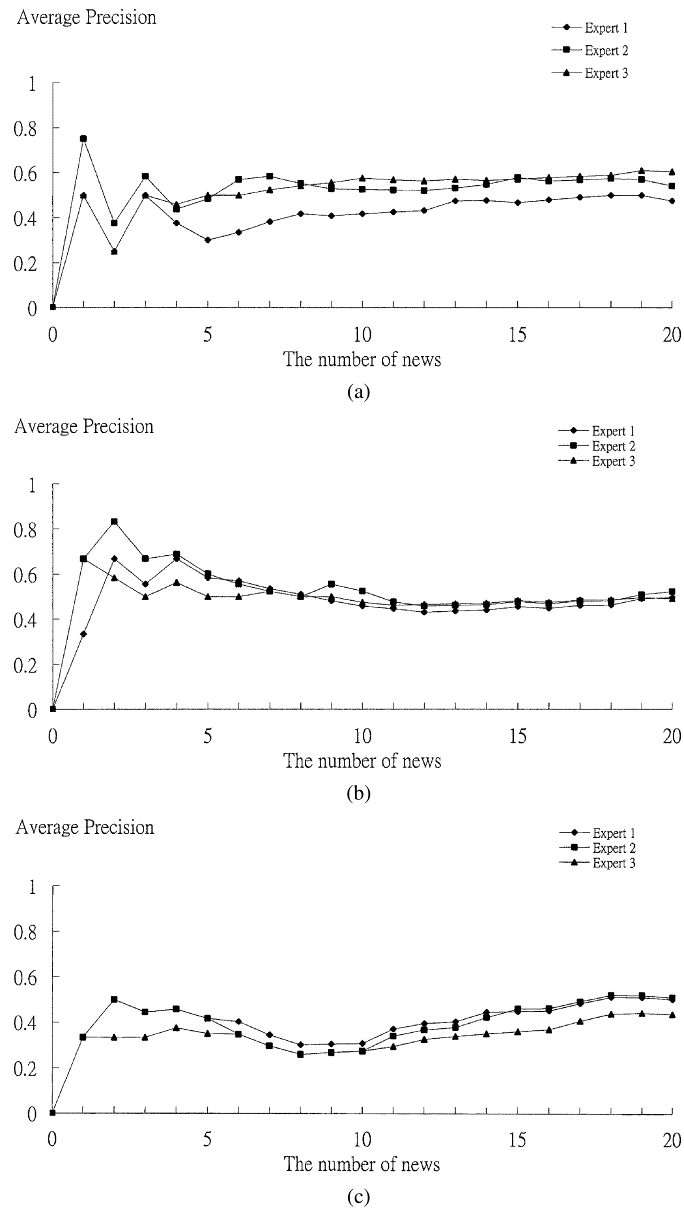


Fig. 11. Average precision of cold current event evaluated by three domain experts. (a) Results of 2002 cold current news. (b) Results of 2003 cold current news. (c) Results of 2004 cold current news.

1.2.2.2) $w_i \leftarrow w_i \times \mu_{C_jE_k}$. /* $\mu_{C_jE_k}$ denotes the membership degree of jth concept of P_i in kth event */
1.2.3) Else
1.2.3.1) Get all information of concept C_j from the temporary fuzzy ontology.
1.2.3.2) Get a relation $R_{C_{j-1}C_j}$ between C_{j-1} and C_j from the temporary fuzzy ontology.
1.2.3.3) Use the relation $R_{C_{j-1}C_j}$ to connect C_{j-1} with C_j and store them into S_i.
1.2.3.4) $w_i \leftarrow w_i \times \mu_{C_jE_k}$.
1.3) If w_i of the sentence $S_i >$ threshold.
Combine the sentence S_i with w_i and store them into the sentence set.
2) End.

The result of *Sentence Generator* is denoted as follows:

Sentence $\boldsymbol{S_i}$**:** [*Concept Name*, *Attribute_value*, *Operation*] →*Relation* → [*Concept Name*, *Attribute_value*, *Operation*] → *Relation* → … → [*Concept Name*, *Attribute_value*, *Operation*] (Possibility w_i)

By the algorithm of *Sentence Generator*, the news agent will generate the following sentences:

Sentence 1: This is the sentence shown at the bottom of the page.

"[*中央氣象局*, *Null*, *Null*]→*研判*→[*未來時間*, *明天*, *Null*]→*有*→[*颱風*, *中度颱風*, *Null*]→*侵襲*→[*台灣地區*, *台灣*, *Null*] (possibility 0.78)"

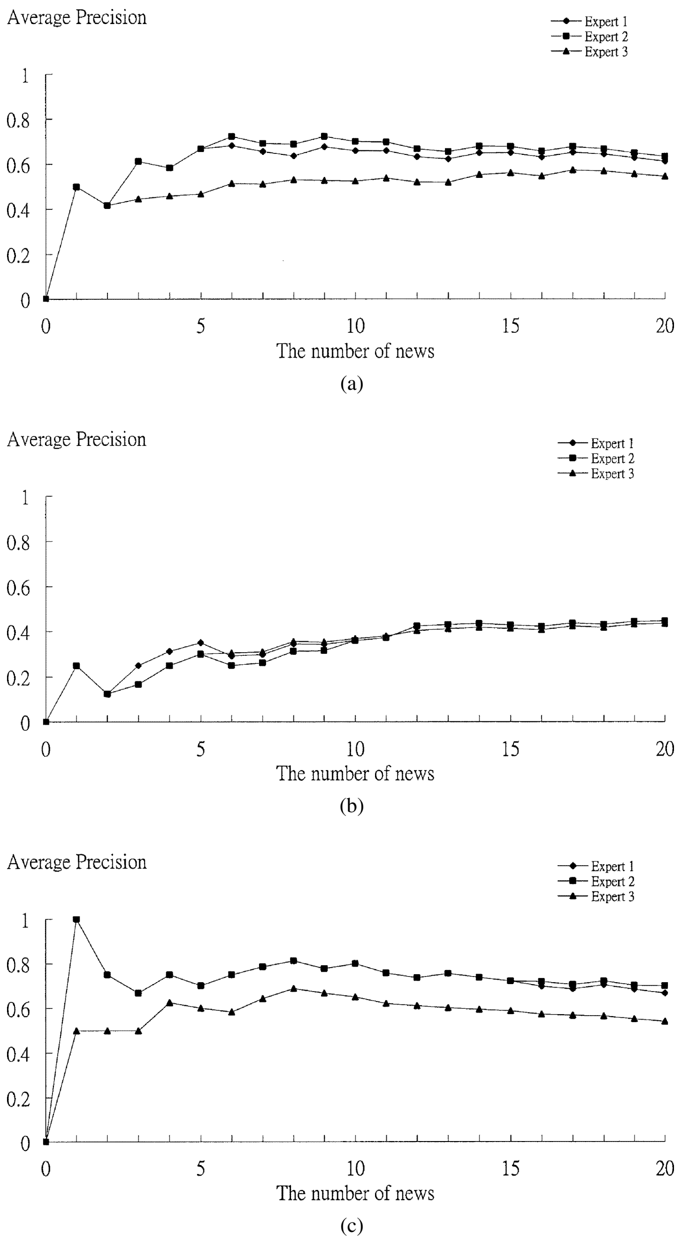


Fig. 12. Average precision of rain event evaluated by three domain experts. (a) Results of 2002 rain news. (b) Results of 2003 rain news. (c) Results of 2004 rain news.

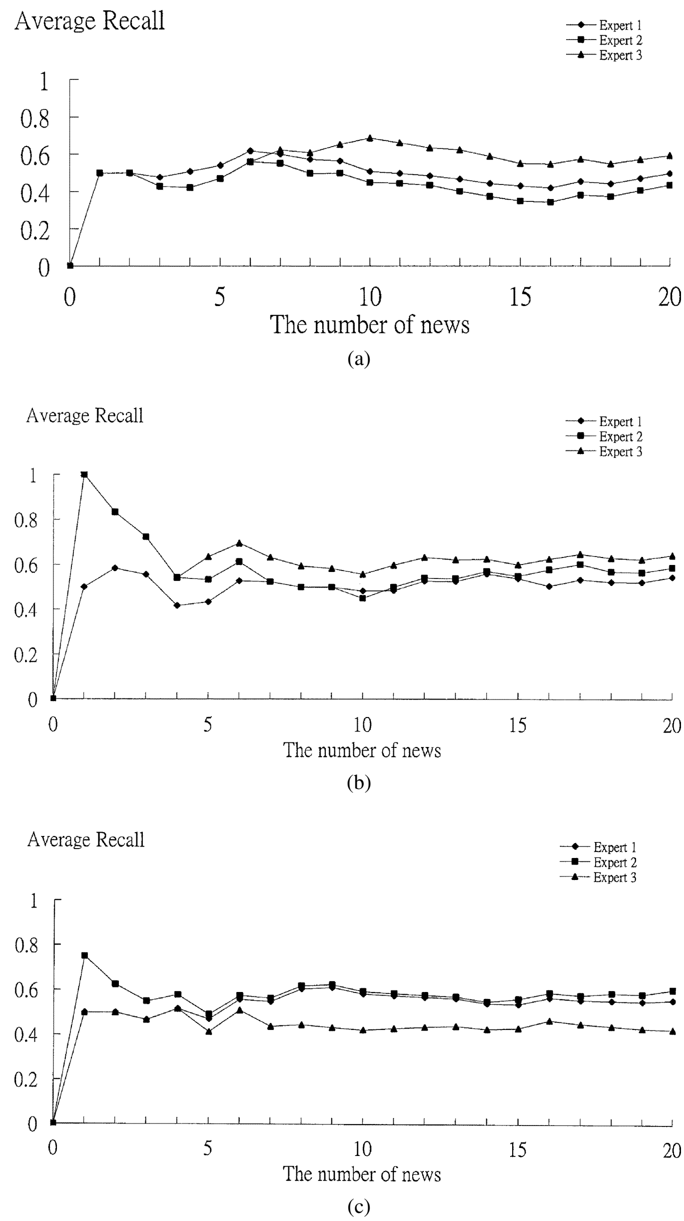


Fig. 13. Average recall of typhoon event evaluated by three domain experts. (a) Results of 2002 typhoon news. (b) Results of 2003 typhoon news. (c) Results of 2004 typhoon news.

Corresponding English Version:

"[*Central weather bureau*, *Null*, *Null*] → *Expect* → [*Future time*, *Tomorrow*, *Null*] → *Has* → [*Typhoon*, *A medium typhoon*, *Null*] → *Attack* → [*Taiwan, Taiwan*, *Null*] (Possibility 0.78)"

English Meaning:

The central weather bureau expects that the typhoon will attack Taiwan tomorrow. (Possibility 0.78).

Sentence 2: This is the sentence shown at the bottom of the page.

Corresponding English Version:

"[*Central weather bureau*, *Null*, *Null*] → *Expect* → [*Future time*, *Tomorrow*, *Null*] → *Will* → [*Rain*, *Null*, *Occur*] (Possibility 0.84)"

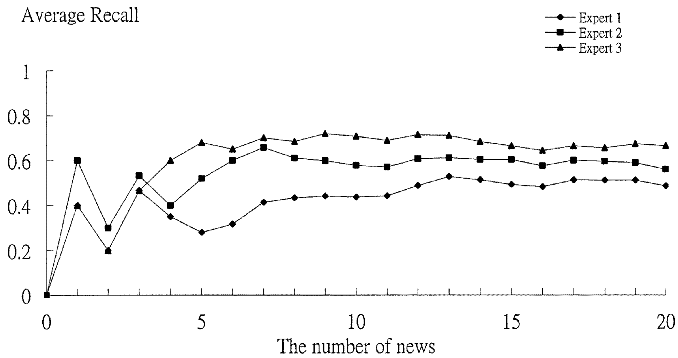
English Meaning:

The central weather bureau expects that it will rain in Taiwan tomorrow. (Possibility 0.84).

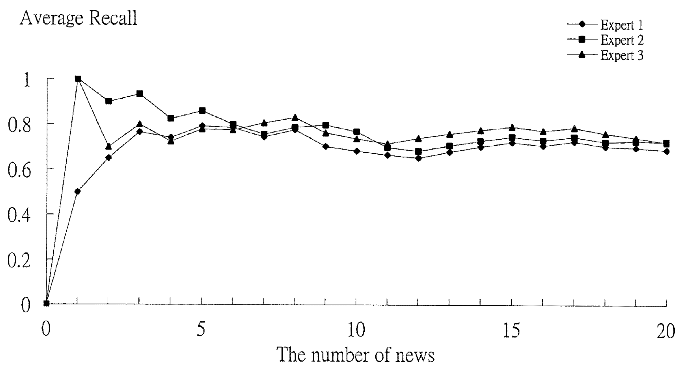
C. Sentence Filter

The *Sentence Filter* is used to filter the redundant sentences. It will generate a brief sentence set. The *Sentence Filter* will

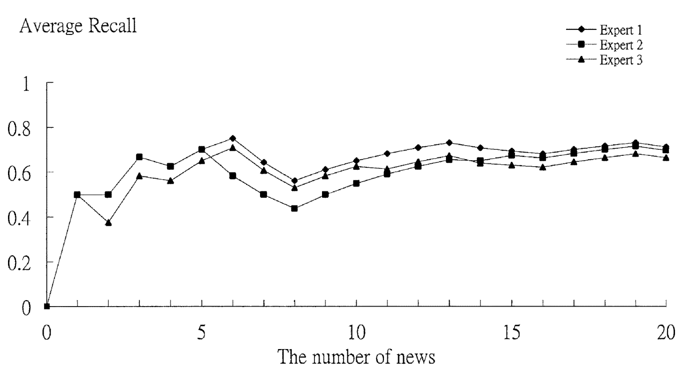
"[*中央氣象局*, *Null*, *Null*]→*研判*→[*未來時間*, *明天*, *Null*]→*有*→[*降雨*, *Null*, *發生*] (Possibility 0.84)"



(a)

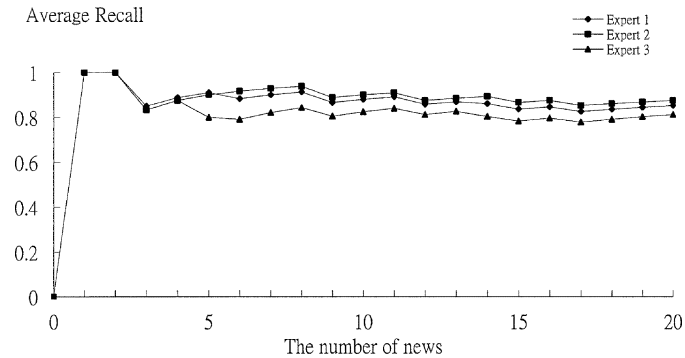


(b)

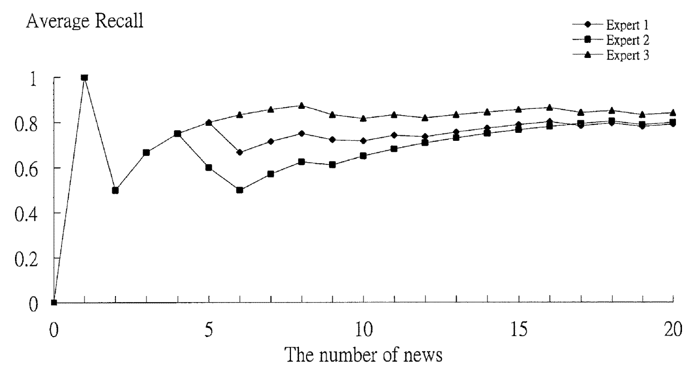


(c)

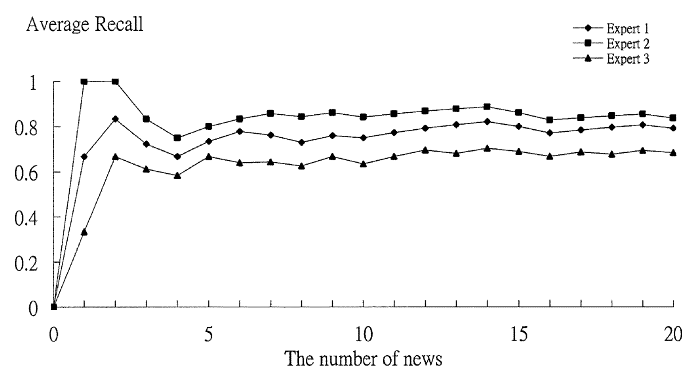
Fig. 14. Average recall of cold current event evaluated by three domain experts. (a) Results of 2002 cold current news. (b) Results of 2003 cold current news. (c) Results of 2004 cold current news.



(a)



(b)



(c)

Fig. 15. Average recall of rain event evaluated by three domain experts. (a) Results of 2002 rain news. (b) Results of 2003 rain news. (c) Results of 2004 rain news.

combine the common sentences and transfer them into semantic sentences. The algorithm of *Sentence Filter* is as follows:

The algorithm of *Sentence Filter*

Input:

A sentence set generated by *Sentence Generator* and the news retrieved by the retrieval agent.

Output:

The summary results of the news.

Method:

1) Sort the length of each sentence of the sentence set descending in order.

2) For $i \leftarrow 1$ to q /* q denotes the number of the sentences in the sentence set S^*/.

2.1) For $j \leftarrow i+1$ to q

2.1.1) $A_i \leftarrow \phi$.

2.1.2) $A_j \leftarrow \phi$.

2.1.3) If $S_i = Null$ or $S_j = Null$.

2.1.3.1) Divide S_i into A_i. /* A_i denotes a set with the concepts of the sentence S_i^*/.

2.1.3.2) Divide S_j into A_j. /* A_j denotes a set with the concepts of the sentence S_j^*/.

2.1.3.3) If $A_j \subseteq A_i$ then

Delete S_j from S.

3) Add the date of the retrieved news to FS. /* FS denotes the summary results of the news */.

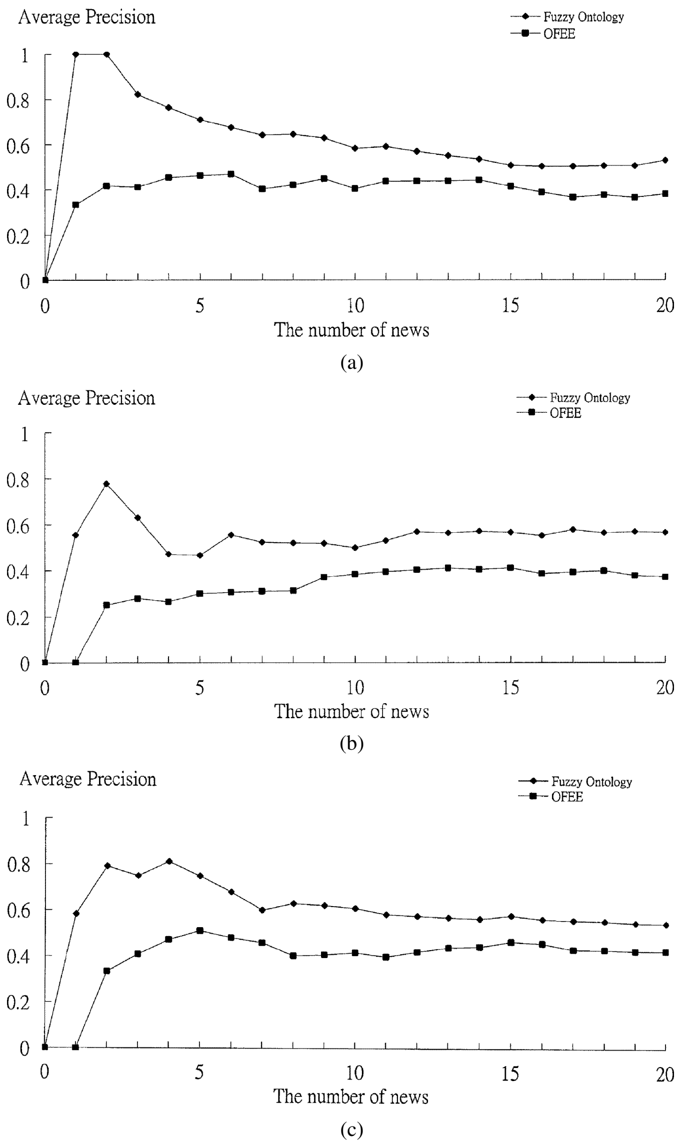


Fig. 16. Average precision curves of fuzzy ontology-based agent and OFEE agent for typhoon event. (a) Results of 2002 typhoon news. (b) Results of 2003 typhoon news. (c) Results of 2004 typhoon news.

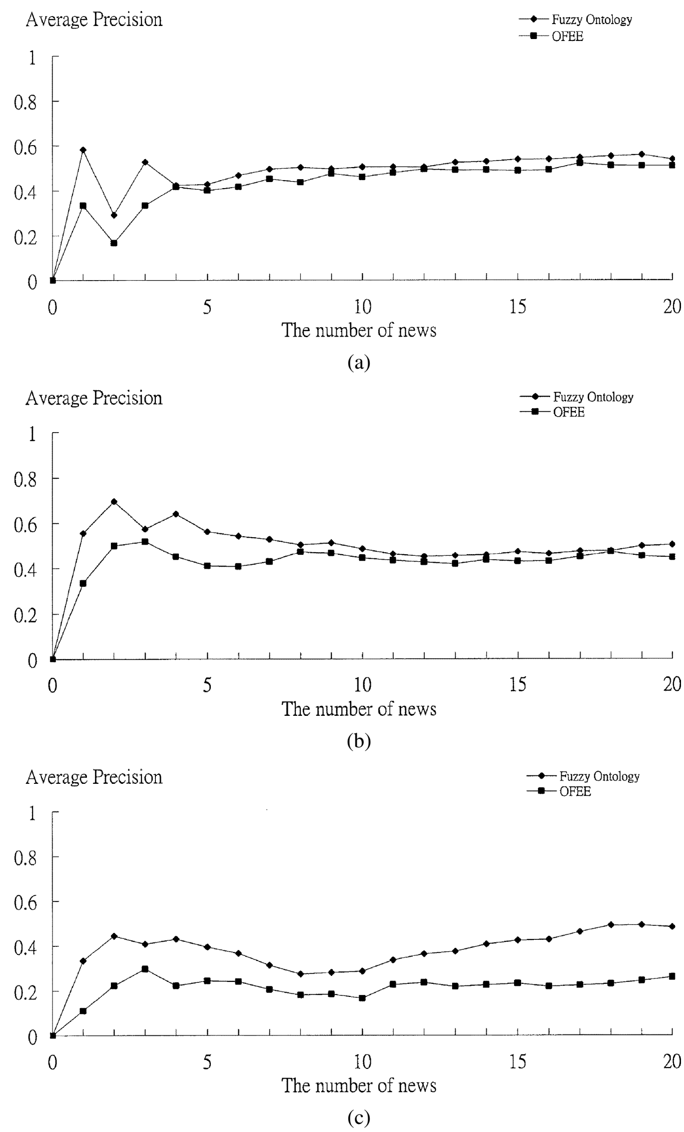


Fig. 17. Average precision curves of fuzzy ontology-based agent and OFEE agent for cold current event. (a) Results of 2002 cold current news. (b) Results of 2003 cold current news. (c) Results of 2004 cold current news.

4) Add the event title of the retrieved news to FS.
5) For $k \leftarrow 1$ to q
 5.1) $SS \leftarrow \phi$. /* SS denotes the semantic sentence */.
 5.2) Divide S_k into A_k.
/* A_k denotes a set with the result of the sentence S_k in the sentence set S^*/.
 5.3) For $m \leftarrow 1$ to n_{A_k} /* n_{A_k} denotes the number of A_k^*/.
  $SS \leftarrow SS \cup I_m$. /* I_m denotes the mth concepts or relations of A_k^*/
 5.4) Add SS to FS.
6) Compute the compression rate based on S and the news, and add it to FS.

$$\text{Compression rate} = \frac{\text{The words of } SS}{\text{The words the the retrieved news } N}.$$

7) Store FS into the summary repository.
8) End.

For example, the following sentences S_1 and S_2 will be combined and transferred to Semantic Sentence *SS* by *Sentence Filter* as follows:

> ***Sentence:*** This sentence is shown at the bottom of the page.

S_1: “[*中央氣象局*, *Null*, *Null*]→*研判*→[*未來時間*, *明天*, *Null*]→*有*→[*颱風*, *Null*, *Null*]→*侵襲*→[*台灣地區*, *Null*, *Null*] (Possibility 0.78)”
S_2: “[*颱風*, *Null*, *Null*]→*侵襲*→[*台灣地區*, *Null*, *Null*] (Possibility 1.0)”

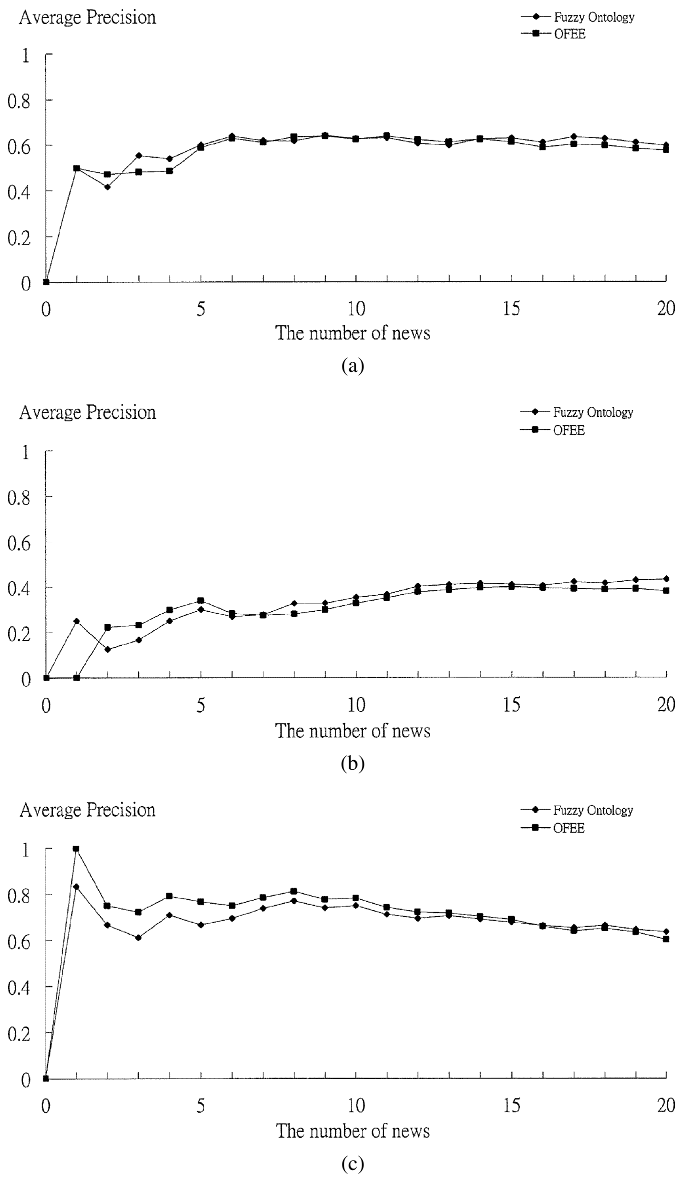


Fig. 18. Average precision curves of fuzzy ontology-based agent and OFEE agent for rain event. (a) Results of 2002 rain news. (b) Results of 2003 rain news. (c) Results of 2004 rain news.

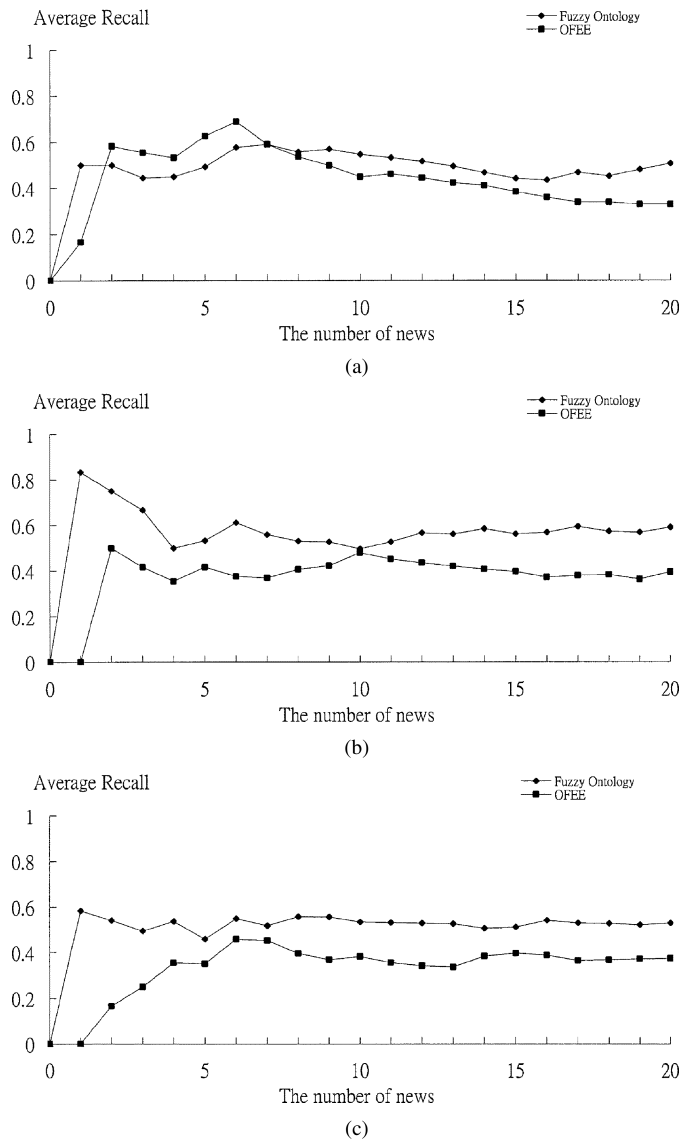


Fig. 19. Average recall curves of fuzzy ontology-based agent and OFEE agent for typhoon event. (a) Results of 2002 typhoon news. (b) Results of 2003 typhoon news. (c) Results of 2004 typhoon news.

Corresponding English Version:

S_1: "[*Central weather bureau*, *Null*, *Null*] → *Expect* → [*Future time*, *Tomorrow*, *Null*] → *Has* → [*Typhoon*, *Null*, *Null*] → *Attack* → [*Taiwan*, *Null*, *Null*] (Possibility 0.78)"

S_2: "[*Typhoon*, *Null*, *Null*] → *Attack* → [*Taiwan*, *Null*, *Null*] (Possibility 1.0)"

English Meaning:

S_1: The central weather bureau expects that the typhoon will attack Taiwan tomorrow (Possibility 0.78).

S_2: The typhoon will attack Taiwan (Possibility 1.0).

Result of the semantic sentence:

SS: 中央氣象局研判明天有颱風侵襲台灣地區 (Possibility 0.78)

Corresponding English Version:

SS: The central weather bureau expects that the typhoon will attack Taiwan tomorrow (Possibility 0.78).

V. Experimental Results

We have constructed an experimental website (http://210.70.151.160/Experimental_Results/) to test the performance of the proposed approach. The website is also located at "Decision Support and Artificial Intelligent Lab." of Chang Jung University in Taiwan, and implemented with JAVA on Linux operating system. The experimental Chinese news is retrieved from one of the largest news websites in Taiwan (http://www.chinatimes.com). There are some evaluation models for document summarization. For example, the DUC multidocument summarization evaluation involved 30 document sets [13]. For each test data set and each target summary size, one automatically generated summary was submitted from each participating site, and one gold-standard summary was created by humans. Comparisons for each data set and target summary size involved the human-created summary versus the summaries automatically produced by competing systems [13].

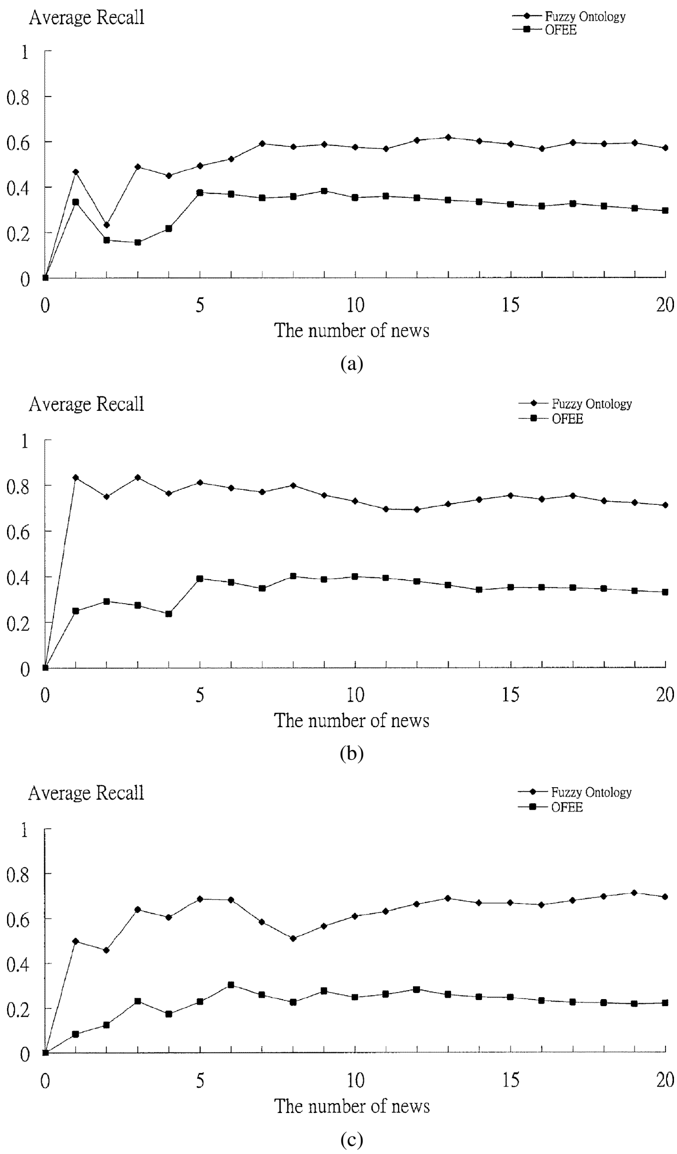


Fig. 20. Average recall curves of fuzzy ontology-based agent and OFEE agent for cold current event. (a) Results of 2002 cold current news. (b) Results of 2003 cold current news. (c) Results of 2004 cold current news.

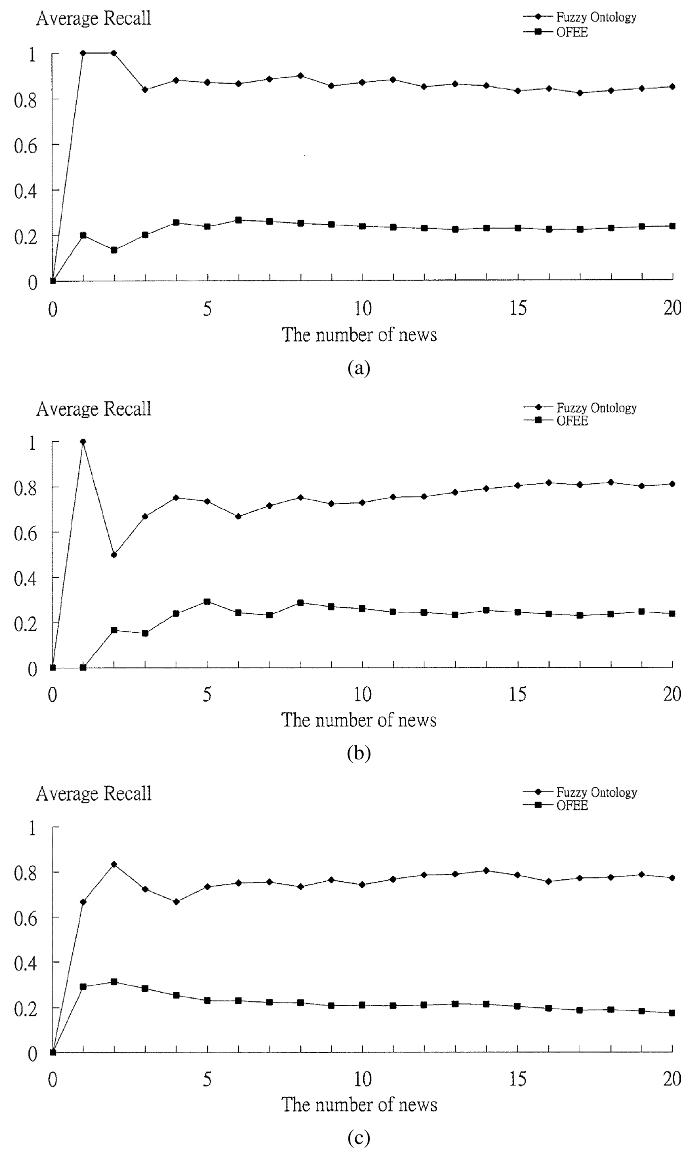


Fig. 21. Average recall curves of fuzzy ontology-based agent and OFEE agent for rain event. (a) Results of 2002 rain news. (b) Results of 2003 rain news. (c) Results of 2004 rain news.

In this paper, we collected the Chinese news of three weather events, including "Typhoon event," "Cold current event," and "Rain event," from Chinatimes website in 2002, 2003, and 2004. Twenty pieces of news are adopted for each event to test the performance of our approach. Hence, the total news number of three events is 180. There are three domain experts participating in the experiment for evaluating the performance of our approach. We adopt the performance measures *Precision* and *Recall* in this experiment. The formulas of *Precision* and *Recall* measures utilized in this paper are as in (10) and (11), shown at the bottom of the page.

Figs. 10–12 show the average precision of "Typhoon event," "Cold current event," and "Rain event" evaluated by the three domain experts, respectively. Figs. 13–15 show the average recall of "Typhoon event," "Cold current event," and "Rain

$$\text{Precision} = \frac{\begin{array}{l}\text{The number of relevant sentences between gold-standard summary and}\\ \text{automatically generated summary}\end{array}}{\text{The number of sentences in gold-standard summary}} \tag{10}$$

$$\text{Recall} = \frac{\begin{array}{l}\text{The number of relevant sentences between gold-standard summary and}\\ \text{automatically generated summary}\end{array}}{\text{The number of sentences in automatically generated summary}}. \tag{11}$$

```xml
<?xml version="1.0" encoding="Big5"?>
<News>
<Title>西南氣流影響南部有雷陣雨查特安颱風增強</Title>
<Date>2002/07/05</Date>
<Content>
雖然雷馬遜颱風已遠離台灣，不過，受西南氣流影響，中央氣象局預測，今天南部會有雷陣雨，中部以北及東北部
午後也可能有短暫陣雨，其他地區是多雲到晴的天氣，另外，關島附近的六號颱風查特安今天上午增強為中度颱風。
氣象局說，雷馬遜颱風已進入黃海，正朝韓國南部而去，今天台灣北部海面風力也逐漸減弱，不過，上午平均風力
仍有七級到八級，最大陣風有十級，台灣附近各海面平均風力也有六級到八級，最大陣風十級。今天各地氣溫仍偏
高，氣象局預測北部、南部都有攝氏三十五度以上，中部約三十四度。另在關島附近的第六號颱風查特安今天上午
八時增強為中度颱風，氣象局說，查特安目前的位置大約在鵝鑾鼻東南東方約二千七百公里海面上，以每小時二十
一公里速度朝西北方向行進，七級風暴風半徑二百公里，目前因離台灣很遠，還要再觀察幾天才能確定台灣是否會
受到影響。
</Content>
</News>
```

(a)

```xml
<?xml version="1.0" encoding="Big5"?>
<News>
<Title>Southwesterly airstream influences southern Taiwan to cause thundershower and Typhoon Zhatan strengthens<Title>
<Date>2002/07/05</Date>
<Content>
Though Typhoon Leimashi had moved away, Taiwan is expected to be influnced by southwesterly airstream of Leimashi,
said the Central Weather Bureau (CWB). According to the CWB, today is a thundershower in southern Taiwan, a short
period of showers after noon in central and northeastern Taiwan and other areas in Taiwan is sunshine. Moreover, the sixth
typhoon near Guam, Zhatan, was this morning upgraded to a medium-strength typhoon by the CWB. And, Leimashi was
moving across the Yellow Sea and toward the south of Korea. This morning's wind power had weaken significantly but the
averaged wind power in the northern Taiwan's sea surface was from 7 to 8 grades and with gusts of 10 grades, as well as in
other sea surface near Taiwan also had 6 to 8 grades and with gusts of 10 grades, said the CWB. The CWB forecasted that
today's weather is expected to be hotter all around Taiwan and it's about above 35 degrees Celsius in northern and southern
Taiwan and about 34 degrees Celsius in central Taiwan. Typhoon Zhatan near Guam had strengthened to be a medium
typhoon this morning. With a 200-km radius and packing winds gusting to 7 grades, Zhatan was estimated to be about 2700
km east-southeast of Oluanpi the southern most tip of Taiwan and was expected to move west-northwest at about 21 km per
hour. Because Zhatan is far from Taiwan, the bureau will keep observing for a few days to know if Zhatan will affect
Taiwan, said the CWB.
</Content>
</News>
```

(b)

Fig. 22. (a). News of typhoon event retrieved from ChinaTimes website. (b) The corresponding English version of Fig. 22(a).

event" evaluated by the three domain experts, respectively. These results show that the average precision and the average recall are stable when increasing the number of documents.

Next, we test the performance of the news agent based on fuzzy ontology by comparing with our previous work on the Ontology-based Fuzzy Event Extraction (OFEE) agent [8]. Figs. 16–19 show the compared curves for the average precision measure, respectively. Figs. 20–22 show the compared results for the average recall measure, respectively. By the experimental results, we observe that the news agent with fuzzy ontology can achieve better summarized results than the OFEE agent.

Finally, we illustrate three examples for Chinese weather news summarization. Figs. 22(a) and 23(a) show the news of typhoon event that retrieved from the ChinaTimes website and the summarization result, respectively. Figs. 22(b) and 23(b) shows the corresponding English version of Figs. 22(a) and 23(a), respectively.

Notice that the summarization sentences for typhoon event including "S1 颱風遠離台灣地區 (Typhoon has moved away from Taiwan)," "S2 台灣地區有降雨 (Taiwan has showers)," and "S3 中央氣象區預測今天有颱風 (Central Weather Bureau forecasts that Taiwan has typhoon today)." The results for sentence S1 and sentence S3 seem confused because the readers cannot read the summary without wondering why the typhoon is both going away and approaching. This is one of the common problems associated with many current summary generation systems; the summary sometimes contains sentences that might not be strongly related. To solve this problem for our approach, we can set the possibility for 0.8, and then, the sentence S3 will be filtered out by the *Sentence Generator*. Another approach for solving this problem is to enhance the structures of the domain ontology and fuzzy ontology. In this case, if we extract the typhoon's name of sentence S1 "雷馬遜颱風 (Typhoon Leimashi)" and sentence S3 "查安特颱風 (Typhoon Zhatan)" from the news, and collect them to the typhoon concept of the domain ontology, then the news agent will generate the sentence with the names. In this way, it may solve this confused problem. Figs. 24(a) and 25(a) show the news of rain event that retrieved from the ChinaTimes website and the summarization result,

Date: 2002/07/05
Event Title: 颱風
Summarization Sentences:
颱風遠離台灣地區 (Possibility 1.0)
台灣地區有降雨 (Possibility 0.99)
中央氣象局預測今天有颱風 (Possibility 0.76)
Compression rate: 9%

(a)

Date: 2002/07/05
Event Title: Typhoon
Summarization Sentences:
Typhoon has moved away from Taiwan. (Possibility 1.0)
Taiwan has showers. (Possibility 0.99)
Central Weather Bureau forecasts that Taiwan has typhoon today. (Possibility 0.76)
Compression rate: 9%

(b)

Fig. 23. (a). Summarization result of Chinese weather news for typhoon event. (b) Corresponding English version of Fig. 23(a).

```
<?xml version="1.0" encoding="Big5"?>
<News>
<Title>鋒面接近 明天盼好雨</Title>
<Date>2002/05/13</Date>
<Content>
明(14)日梅雨鋒面接近，台灣中部以北、東北部、東部地區及馬祖、金門、澎湖將轉爲有陣雨或雷雨的天氣，可望帶來明顯雨勢。今(13)日則除台灣東北部地區有短暫陣雨，山區午後有局部陣雨外，台灣其他各地、澎湖、金門及馬祖爲多雲到晴的天氣，全台白天最高溫將達 32、33 度左右。
</Content>
</News>
```

(a)

```
<?xml version="1.0" encoding="Big5"?>
<News>
<Title> Fronts are approaching and expected to be rainy</Title>
<Date>2002/05/13</Date>
<Content>
Tomorrow the fronts of plum rains will approach to Taiwan and the weather will break in showers or thundershowers in northern, northeastern and eastern Taiwan as well as other offshore islands Matsu, Kinmen and Penghu and it is expected to around brings more rains to the above areas, said the Central Weather Bureau (CWB). The weather forecaster reported that today all Taiwan and offshore islands are sunshine except that the region of the eastern Taiwan has a short period of showers and the mountainous region in this region will have thunder showers after noon. And the highest day-time temperature in Taiwan can reach up to 32 or 33 degrees Celsius.
</Content>
</News>
```

(b)

Fig. 24. (a) News of rain event retrieved from ChinaTimes website. (b) Corresponding English version of Fig. 24(a).

respectively. Figs. 24(b) and 25(b) show the corresponding English version of Figs. 24(a) and 25(a), respectively.

Figs. 26(a) and 27(a) show the news of cold current event that retrieved from the ChinaTimes website and the summarization result, respectively. Figs. 26(b) and 27(b) shows the corresponding English version of Figs. 26(a) and 27(a), respectively.

Date: 2002/05/13
Event Title: 降雨
Summarization Sentences:
台灣地區有降雨 (Possibility 1.0)
鋒面接近台灣地區 (Possibility 0.99)
Compression rate: 12%

(a)

Date: 2002/05/13
Event Title: Rain
Summarization Sentences:
Taiwan has showers. (Possibility 1.0)
Fronts approach Taiwan. (Possibility 0.99)
Compression rate: 12%

(b)

Fig. 25. (a) Summarization result of Chinese weather news for rain event. (b) Corresponding English version of Fig. 25(a).

VI. Conclusions and Future Work

In this paper, a fuzzy ontology and its application to news summarization are presented. The fuzzy ontology with fuzzy concepts is an extension of the domain ontology with crisp concepts that is more suitable to describe the domain knowledge for solving the uncertainty reasoning problems. In our previous work [8], we have presented a five-layer Fuzzy Inference Agent (FIA) [7], [10] for news event ontology extraction. In this paper, we modify the FIA and extend it to a seven-layer *Fuzzy Inference Mechanism* (*FIM*) to create the fuzzy ontology. Layers 6 and 7 of *FIM* are utilized to compute and integrate the fuzzy membership degrees and then generate the fuzzy concepts for each event of the news. In addition, a news agent based on the fuzzy ontology is also developed for news summarization.

```
<?xml version="1.0" encoding="Big5"?>
<News>
<Title>大陸冷氣團 下週一抵台灣 最低只有十五度</Title>
<Date>2002/11/23</Date>
<Content>
深秋來襲的豪雨特報昨日剛解除，隨即又有大陸冷氣團來勢洶洶！中央氣象局表示，今年入秋以來的第一波大陸冷
氣團將在下週一抵達台灣，目前預估中部以北最低溫降到十五、六度。連著兩天爲東北部帶來豪大雨的東北季風強
度昨日進一步減弱，氣象局隨即於下午四時三十分解除東北部的豪雨特報。
</Content>
</News>
```

(a)

```
<?xml version="1.0" encoding="Big5"?>
<News>
<Title>Continental polar air masses will arrive in Taiwan next Monday and the lowest temperature is only 15 degrees
Celsius</Title>
<Date>2002/11/23</Date>
<Content>
After torrential rain warnings were lifted yesterday, immediately the continental polar air masses broke in full fury. The first
cold air masses of this autumn will arrive in Taiwan next Monday and the lowest temperature in northern Taiwan is expected
to drop down to 15 or 16 degrees Celsius, said the Central Weather Bureau (CWB). The northeasterly seasonal winds brought
heavy rain to the northeastern Taiwan for two continuous days had died down, so the CWB lifted the torrential warnings in
this area at 4:30 p.m.
</Content>
</News>
```

(b)

Fig. 26. (a) News of cold current event retrieved from ChinaTimes website. (b) Corresponding English version of Fig. 26(a).

Date: 2002/11/23
Event Title: 寒流
Summarization Sentences:
台灣地區有降雨 (Possibility 0.95)
Compression rate: 6%

(a)

Date: 2002/11/23
Event Title: Cold current
Summarization Sentences:
Taiwan has showers. (Possibility 0.95)
Compression rate: 6%

(b)

Fig. 27. (a) Summarization result of Chinese weather news for cold current event. (b) Corresponding English version of Fig. 27(a).

The news agent contains five modules, including a *retrieval agent*, a *document preprocessing mechanism*, a *sentence path extractor*, a *sentence generator*, and a *sentence filter*, to perform news summarization. Furthermore, we construct an experimental website to test the proposed approach. The experimental results show that the news agent based on the fuzzy ontology can effectively operate for news summarization.

Besides, there are still some problems needed to be further studied in the future. For example, the summary sometimes contains sentences that might not be strongly related. This problem is also one of the common bottlenecks associated with many current summary generation systems. The further studies about how to define a suitable threshold to generate the sentence set is among our future work. In addition, although this paper presents a new approach to solve the Chinese summarization problem, the measured results for precision and recall still could be further improved. Moreover, the structure of the *fuzzy ontology*, including the *fuzzy concepts* and *fuzzy relationships*, will be further refined and extended to various application domains. Finally, in the future, we will extend the news agent to process the English news by utilizing another part-of-speech tagger and English dictionary like WordNet.

ACKNOWLEDGMENT

The authors would like to thank the anonymous referees for their constructive and useful comments.

REFERENCES

[1] H. Alani, S. Kim, D. E. Millard, M. J. Weal, W. Hall, P. H. Lewis, and N. R. Shadbolt, "Automatic ontology-based knowledge extraction from web documents," *IEEE Intell. Syst.*, vol. 18, no. 1, pp. 14–21, Jan./Feb. 2003.

[2] T. H. Cormen, C. E. Leiserson, and R. L. Rivest, *Introduction to Algorithms*. New York: McGraw-Hill, 1999.

[3] D. W. Embley, D. M. Campbell, R. D. Smith, and S. W. Liddle, "Ontology-based extraction and structuring of information from data-rich unstructured documents," in *Proc. ACM Conf. Inf. Knowledge Manage.*, 1998, pp. 52–59.

[4] N. Guarino, C. Masolo, and G. Vetere, "OntoSeek: Content-based access to the web," *IEEE Intell. Syst.*, vol. 14, no. 3, pp. 70–80, May/Jun. 1999.

[5] H. V. Halteren, "New feature sets for summarization by sentence extraction," *IEEE Intell. Syst.*, vol. 18, no. 4, pp. 34–42, Jul./Aug. 2003.

[6] S. Handschuh, S. Staab, and F. Ciravegna, "S-CREAM-semi-automatic CREAtion of metadata," in *Proc. 13th Int. Conf. Knowledge Eng. Manage.*, 2002, pp. 358–372.

[7] Y. H. Kuo, J. P. Hsu, and C. H. Wang, "A parallel fuzzy inference model with distributed prediction scheme for reinforcement learning," *IEEE Trans. Syst., Man, Cybern. B*, vol. 28, no. 2, pp. 160–172, Apr. 1998.

[8] 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.

[9] W. Lam and K. S. Ho, "FIDS: An intelligent financial web news articles digest system," *IEEE Trans. Syst., Man, Cybern. A*, vol. 31, no. 6, pp. 753–762, Nov. 2001.

[10] C. T. Lin and C. S. G. Lee, "Neural-network-based fuzzy logic control and decision system," *IEEE Trans. Comput.*, vol. 40, pp. 1320–1336, Dec. 1991.

[11] N. Lammari and E. Metais, "Building and maintaining ontologies: A set of algorithm," *Data Knowledge Eng.*, vol. 48, pp. 155–176, 2004.

[12] I. Mani, "Recent developments in text summarization," in *Proc. CIKM*, 2001, pp. 529–531.

[13] K. R. McKeown, R. Barzilay, D. Evans, V. Hatzivassiloglou, M. Y. Kan, B. Schiffman, and S. Teufel, "Columbia multi-document summarization: Approach and evaluation," in *Proc. DUC Conf. Text Summarization*, 2001.

[14] R. Navigli and P. Velardi, "Ontology learning and its application to automated terminology translation," *IEEE Intell. Syst.*, vol. 18, no. 1, pp. 22–31, Jan./Feb. 2003.

[15] E. Riloff, "An empirical study of automated dictionary construction for information extraction in three domains," *Artificial Intell.*, vol. 85, no. 1–2, pp. 101–134, Aug. 1996.

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

[17] M. Vargas-Vera, E. Motta, J. Domingue, M. Lanzoni, A. Stutt, and F. Ciravegna, "MnM: Ontology driven semi-automatic and automatic support for semantic markup," in *Proc. 13th Int. Conf. Knowledge Eng. Manage.*, 2002, pp. 379–391.

[18] H. J. Zimmermann, *Fuzzy Set Theory and its Applications*. Boston, MA: Kluwer, 1991.

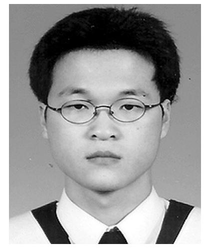
[19] (2003) Chinese Knowledge Information Processing Group. Academic Sinica, Taipei, Taiwan, R.O.C. [Online]. Available: http://godel.iis.sinica.edu.tw/CKIP/

[20] "Chinese Electronic Dictionary, Technical Report," Academia Sinica, Taipei, Taiwan, R.O.C., no. 93–05, 1993.

Chang-Shing Lee received the B.S. degree in information and computer engineering from the Chung Yuan Christian University, Chung-Li, Taiwan, R.O.C., in 1992, the M.S. degree in computer science and information engineering from the National Chung Cheng University, Chia-Yi, Taiwan, in 1994, and the Ph.D. degree in computer science and information engineering from the National Cheng Kung University, Tainan, Taiwan, in 1998.

He is currently an Associate Professor with the Department of Information Management, Chang Jung Christian University, Tainan. His research interests include intelligent agent, ontology engineering, knowledge management, Web services, semantic Web, and soft computing systems. He holds several patents on ontology engineering, document classification, and image filtering.

Dr. Lee received the MOE's Campus Software Award in 2002, the CJCU's Outstanding Research Achievement Award in 2003, and the Outstanding Teacher Award from Chang Jung Christian University in 2004. He has guest edited a special issue for the *Journal of Internet Technology*. He is a Member of TAAI.



Zhi-Wei Jian was born in Ilan, Taiwan, R.O.C., in 1980. He received the B.S. degree in information management from the Chang Jung Christian University, Tainan, Taiwan, in 2002. He is currently working toward the M.S. degree in computer science and information engineering at National Cheng Kung University, Tainan.

His research interests include document summarization, ontology application, natural language processing, and knowledge acquisition.

Lin-Kai Huang was born in Changhua, Taiwan, R.O.C., in 1981. He received the B.S. degree in information management from the Southern Taiwan University of Technology, Tainan, Taiwan, in 2003. He is currently working toward the M.S. degree in information management at Chang Jung Christian University, Tainan.

His research interests include ontology applications, quality management, and CMMI.