An Empirical Study of Ontology-Based Multi-Document Summarization in Disaster Management

Lei Li and Tao Li

***Abstract*—Domain ontology, as a conceptual model, provides a meaningful framework for semantic representation of textual information. In this paper, we explore the feasibility of using the ontology in solving multi-document summarization problems in the domain of disaster management. We provide an empirical study of different approaches in which the ontology has been used for summarization tasks. Extensive experiments on a collection of press releases relevant to Hurricane Wilma in 2005 demonstrate that ontology-based multi-document summarization methods outperform other baselines in terms of the summary quality.**

***Index Terms*—Disaster management, multi-document summarization, ontology, query expansion.**

I. Introduction

It is well known that hurricanes, earthquakes, and other natural disasters cause immense physical destruction and loss of life and property around the world. In order to efficiently analyze the trend of the disasters and minimize the consequent loss for future situation, effective information gathering methods are important. Specifically, a myriad of news and reports that are related to the disaster may be recorded in the form of text documents. The domain experts expect to obtain condensed information about the detailed disaster event description, e.g., the evolutionary tendency of the disaster, the operational status of the public services, and the reconstruction process of the homestead. In the following, a representative scenario is provided, in which the information frequently investigated by a disaster analyst is described.

Scenario: Hurricane Wilma passed through South Florida in October, 2005. During Wilma, the power supply in Miami was extremely influenced. The domain experts want to check the status of the power supply during Wilma and after Wilma passed.

Manuscript received March 18, 2011; revised August 12, 2011, February 3, 2012, and July 7, 2012; accepted April 2, 2013. Date of publication November 13, 2013; date of current version January 13, 2014. This work was supported in part by the National Science Foundation under Grant HRD-0833093, Grant CNS-1126619, and Grant IIS-1213026, and the U.S. Department of Homeland Security under grant Award Number 2010-ST-062-000039, and Army Research Office under Grant W911NF-10-1-0366 and Grant W911NF-12-1-0431. This paper was recommended by Associate Editor Z. Zdrahal.

The authors are with the School of Computing and Information Sciences, Florida International University, Miami, FL 33199 USA (e-mail: lli003@cs.fiu.edu; taoli@cs.fiu.edu).

Digital Object Identifier 10.1109/TSMCC.2013.2258335

TABLE I
Example of Disaster Information

Power supply in Miami
Florida Power and Light reports 380 000 customers have lost power on October 21th, 2005.
Nearly one million FPL customers without power in Miami-Dade County on October 24th, 2005.
Power is beginning to be restored to FPL customers from October 25th, and it may take several weeks to be fully restored.

A list of descriptive sentences on this topic are shown in Table I. As is shown, the three sentences provide a summary on the status of power supply over a week in the district of Miami-Dade County. Such information can provide domain analysts a preliminary overview of how the power supply was influenced by the hurricane, and subsequently, domain analysts will contact the corresponding department and establish a set of measures that would be helpful once the situation happens again.

In the domain of disaster management, over thousands of hundreds of reports are often released by the local government or local emergency offices during the disaster, which cover most events relevant to the disaster and the time span will be days to months, depending on how severe the disaster is. The data will be presented in a format of newswire, containing a lot of routine reporting on multiple aspects of the disaster. In such a case, it is extremely difficult for domain experts to quickly find either the most important information overall (generic summarization) or the most relevant information to a specified query (query/topic-focused summarization). Therefore, multi-document summarization techniques can be used to extract meaningful information from multiple reports.

A domain ontology related to disaster management, describing the concepts and the corresponding relations of these concepts, is often provided by domain experts [1]. Such an ontology contains plentiful conceptual information related to the document set, which may be beneficial for users to summarize the documents. A natural question is how we can utilize the ontology to obtain high-quality summaries, i.e., representing topics with nonredundant sentences.

In this paper, we explore the feasibility of employing the ontology into multi-document summarization problems in disaster management domain. We first discuss how to represent a sentence as a vector using the domain ontology. We then delve into the problems from two directions: generic and query-focused summarization. In generic summarization, we provide comprehensive studies of the centroid-based sentence selection approaches by using different vector space models, and explore the possibility of utilizing the ontology to achieve the goal of reducing information redundancy. In query-focused summarization, we optimize the final summary results by employing ontology-based query expansion methods into the summarization. We conduct experiments on a collection of press releases related to Hurricane Wilma, and the results show that ontology-based methods can provide promising performance for summarization.

The rest of the paper is organized as follows. In Section II, we review the related work. After introducing the domain ontology applied in our work in Section III, we provide a comparative study on several ontology-based representations in Section IV. Section V presents the experimental results and analysis, and finally Section VI concludes the paper and covers the future work.

II. Related Work

A. Generic Summarization

For generic summarization, a saliency score is usually assigned to each sentence, the sentences are ranked according to the saliency score, and then the top ranked sentences are selected as the summary based on the ranking result. Recently, both unsupervised and supervised methods have been proposed to analyze the information contained in a document set, and extract highly salient sentences into the summary based on syntactic or statistical features [2]–[6]. For example, MEAD [7] is an implementation of the centroid-based method in which the sentence scores are computed based on sentence-level and inter-sentence features.

However, most existing methods ignore the conceptual information in the sentence level. In most cases, the conceptual information can provide users more readable results for summaries. Some researchers utilize the explicit concepts within sentences to address multi-document summarization [8], [9], e.g., using Wikipedia. However, such techniques cannot be directly applied to domain-specific document summarization tasks, since Wikipedia contains too many concepts not relevant to a specific domain. In our previous work [25], we explored the possibility of using domain-specific ontology for multi-document summarization; however, no detailed semantic relationship is considered.

B. Query-Focused Summarization

In query-focused summarization, the information related to a given topic or query should be incorporated into summaries, and the sentences suiting the user's declared information need should be extracted. Many methods for generic summarization can be extended to incorporate the query information. Saggion *et al.* [10] presented a robust summarization system developed within the GATE architecture that makes use of robust components for semantic tagging and co-reference resolution provided by GATE. Wei *et al.* [11] incorporated the query influence into the mutual reinforcement chain to cope with the need for query-oriented multi-document summarization. Wan *et al*. [12] used both relationships among sentences and relationships between the given query and the sentences by manifold ranking. Probability models have also been proposed with different assumptions on the generation process of the documents and the queries [13], [14].

C. Query Expansion

Query expansion is the process of augmenting the user's query with additional terms in order to improve search results. For instance, when we are ready to search "panther" by some search engine, we can expand such query by adding synonyms of "panther" to the query, such as "jaguar," "cougar," etc. Query expansion has also been explored in the field of document summarization, where the quality of the generated summary can be improved. For example, Daume and Marcu [15] propose a justified query expansion technique in the language modeling for IR framework. However, it fails to consider the semantic relatedness between the sentences and the query string.

III. Disaster Management Domain

A. Domain Description

It is well known that hurricanes, earthquakes, and other natural disasters cause immense physical destruction, loss of life and property around the world. The purpose of the disaster management program is to enhance efficient coordination and collaboration among public safety organizations by enabling the interoperable sharing of emergency alerts and incident-related data between disparate systems. One of the disaster management systems aims to analyze the news and reports related to the disaster to provide concise and recapitulative information for domain experts.

B. Domain-Specific Ontology

Generally speaking, an ontology is often provided by domain experts in disaster management domain [1]. Such an ontology provides answers for the questions concerning what entities exist in disaster management, and how such entities can be related within a hierarchy and subdivided according to similarities and differences among them. The ontology described in this paper is related to the domain of hurricane management, involving 109 concepts and 326 concept relations. This ontology is obtained from the disaster management project at Florida International University [26] (http://www.bizrecovery.org). The ontology is created for the purpose of research included in this project, and is provided by the domain experts from the State Emergency Operations Center (EOC)[1] of Florida. The ontology consists of the *Root*, a set of *concepts*, a set of *is-a relations*, a set of *equivalent-class* and a set of *individuals*. A subset of concepts in the ontology hierarchy is shown in Fig. 1.

[1]http://www.floridadisaster.org/eoc/Update/Home.asp.

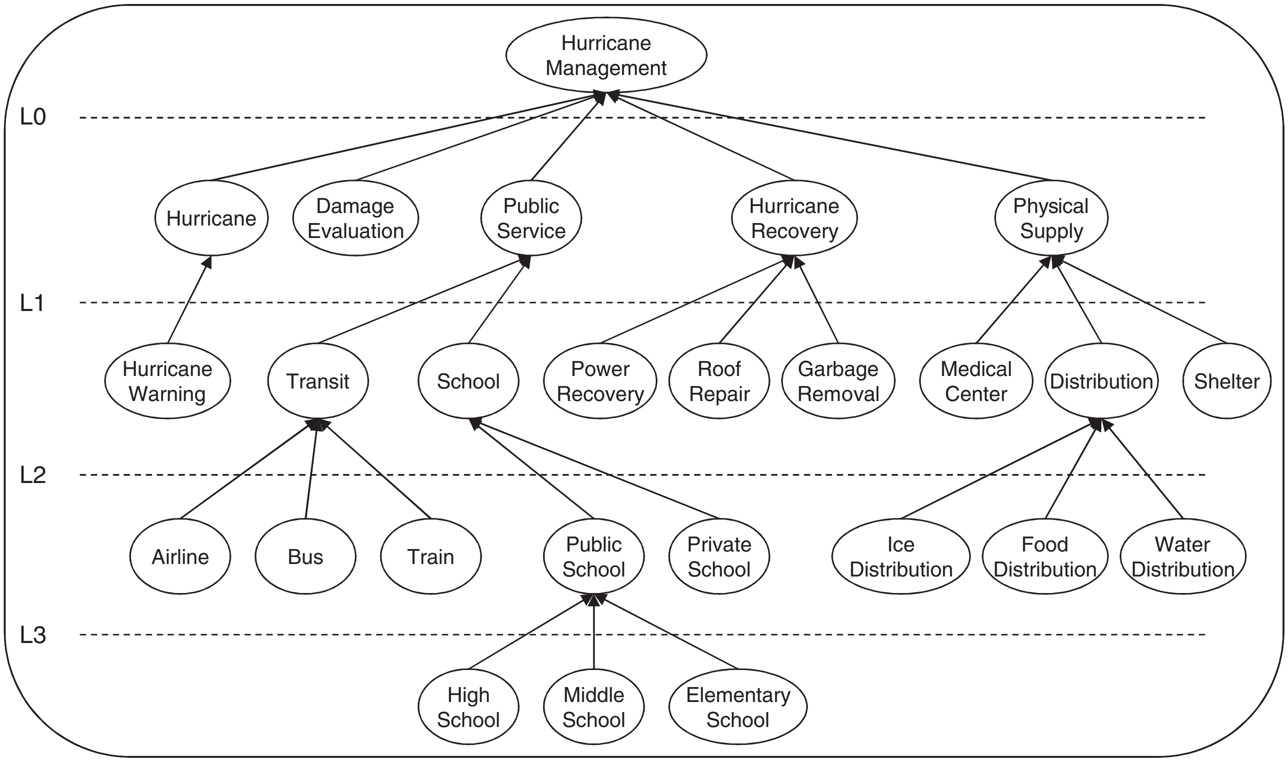


Fig. 1. "Hurricane management" ontology.

IV. Summarization Approaches

To address the summarization issues in the domain of hurricane management, we first map most sentences in the document set onto the domain ontology, and then take advantage of the intrinsic properties of the ontology to represent each sentence. In this section, we explore the effect of the ontology in multi-document summarization tasks from two directions: generic summarization and query-focused summarization.

A. Sentence Mapping

Ontology in disaster management domain provides us abundant conceptual and semantic information, which might facilitate the procedure of multi-document summarization. To utilize the ontology for better understanding the documents, we initially decompose the collection of domain-specific documents into sentences, and then map each sentence to the ontology hierarchy. For each concept of the ontology hierarchy, a group of keywords (i.e., nouns) are assigned by the experts for the sake of sentence mapping. The procedure of sentence mapping is executed based on the following criteria.

1) If the sentence is related to only one concept, map this sentence to the corresponding concept.
2) If the sentence is related to two or more concepts, map this sentence to the least common ancestor (LCA) of these concepts. If the LCA is the most general concept of the ontology, then map the sentence to the original specific concepts.

In this process, we calculate the word set overlapping between a sentence (only considering nouns in the sentence) and the keyword set assigned to each concept as the measure of relatedness, and then rank the scores to select the most related concept. Since different concepts in the ontology have different unambiguous representative noun sets assigned by domain experts, it is unlikely that the same noun will appear in more than one concept. When the condition of the second criterion holds, it means that the sentence contains different words that can map to different concepts. In order to avoid that a single sentence will be linked to multiple concepts and thus make more redundant information, we introduce the LCA of concepts and link the sentence to the LCA if it contains two or more concepts. Based on these criteria, we can guarantee that most sentences are mapped to at least one concept of the ontology because the ontological concepts are representative in a specific domain, and the mapping is reasonable since the mapped sentences can be regarded as instances of the corresponding concepts.

B. Sentence Representation

A key question in multi-document summarization using the ontology is how to represent the sentences we have mapped onto the ontology. We examine several ways to model a sentence into a vector, including term frequency (TF) model [16], term frequency-inverse sentence frequency (TFISF) model [16], term frequency-inverse concept frequency (TFICF) model, concept hierarchy (CH) model [17], [18], and the linear combinations of these models. The vector space models mentioned above provide different insights for document summarization. In the following, we delve into the details of these models for the purpose of comparison.

Term Frequency Model: In this model, each entry of a sentence vector denotes the term weight (normalized term frequency to prevent a bias toward longer sentences and to give a measure of the importance of the term t_i within the particular sentence s_j). Thus we have the term frequency defined as follows:

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \tag{1}$$

where $n_{i,j}$ is the number of occurrences of term t_i in sentence s_j, and the denominator $\sum_k n_{k,j}$ is the sum of number of occurrences of all the terms in sentence s_j.

TFISF Model: Similar to TFIDF [16], term frequency-inverse sentence frequency (TFISF) is used to evaluate how important a word is to a sentence in a corpus. The importance increases proportionally to the number of times a word appears in the sentence but is offset by the frequency of the word in the corpus. The inverse sentence frequency is a measure of the general importance of the term (obtained by dividing the total number of sentences by the number of sentences containing the term, and then taking the logarithm of that quotient), which is defined as

$$isf_i = \log \frac{|S|}{|s : t_i \in s|} \tag{2}$$

where $|S|$ is the total number of sentences in the corpus, and $|s : t_i \in s|$ is the number of sentences where the term t_i appears (that is, $n_{i,j} \neq 0$). Then TFISF is defined as

$$\text{TFISF}_{i,j} = tf_{i,j} * isf_i. \tag{3}$$

A high weight in TFISF is reached by a high term frequency (in the given sentence) and a low sentence frequency of the term in the whole document collection; the weights hence tend to filter out common terms. The TFISF value for a term will always be larger than or equal to zero.

TFICF Model: Term frequency-inverse concept frequency (TFICF) is used to evaluate how important a word is to a concept in an ontology hierarchy. Compared with TFISF, the difference is that the importance increases proportionally to the number of times a word appears under the concept but is offset by the frequency of the concept in the ontology hierarchy. The inverse concept frequency is computed as

$$icf_i = \log \frac{|C|}{|c : t_i \in c|} \tag{4}$$

where $|C|$ is the total number of concepts in the ontology hierarchy, and $|c : t_i \in c|$ is the number of concepts where the term t_i appears (that is, $n_{i,j} \neq 0$). If the term is in the corpus but not appears in any concepts, then the icf value is set to be 0. TFICF is defined as

$$TFICF_{i,j} = tf_{i,j} * icf_i. \tag{5}$$

Concept Hierarchy Model: The is-a taxonomical relations between concepts in the ontology provide apparent hierarchical information, which can be applied to representing a sentence instead of using the whole set of terms/words. In this section,

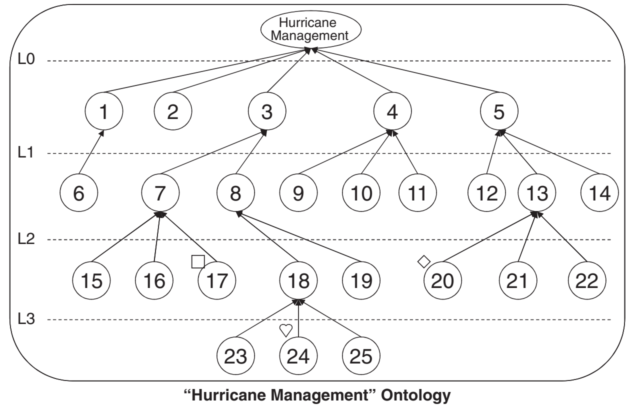


Fig. 2. "Hurricane management" ontology.

we modify the approach proposed by Yuan and Sun [17], [18]. The new approach, called concept hierarchy (CH) model, is used as an ontology-based structured representation for sentences. The procedure of sentence modeling by CH is as follows.

1) Given the ontology described in Section III, we number all of the nodes level by level from top to bottom except the root. By doing this, we can obtain an ontology with numbered concepts. Then we map the sentence set onto the ontology based on the criteria described in Section IV-A (Fig. 2).
2) We represent a sentence as a vector V_c with the size of the total number of concepts in the ontology, denoted as $\langle c_1, c_2, \cdots, c_{|C|}\rangle$. This vector is initialized with zeros. The value of each entry c_i is increased by 1 when encountering an existence of the i-th concept node in the sentence. For instance, the term vector of a sentence (related to node 24) unfolds as (0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0).
3) To impose the structure into a vector, all of the ancestor nodes of the 1-valued c_i are accounted as existence. Using the previous example of the sentence of node 24, the resulting vector becomes (0,0,1,0,0,0,0,1,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,1,0), as the ancestor nodes of 24 include node 3, 8, 18.
4) V_c is l_2-normalized so that $||V_c||_2 = 1$.

Intuitively, more general concepts in the ontology hierarchy should contribute more to the sentence clustering, since we expect that the sentences can be separated into difference clusters. Consider two sentences that describe two different fine-grained concepts and these two concepts belong to different general concepts, the CH representations of these two sentences are totally different from each other. On the other hand, the fine-grained concepts in the hierarchy would be more influential to differentiate two sentences that are in the same sentence cluster, and therefore might influence the ranking of the sentences.

Linear Combinations: The first three vector space models—TF, TFISF, TFICF—are constructed based on the term feature in the sentences, whereas CH utilizes the concepts appearing in the whole document collection to represent a sentence. In the following, we try to combine the term-based VSMs and the concept-based VSM together to verify if it is helpful to summarize multiple documents. If we define V_{tf} to be the term frequency vector, V_{tfisf} to be the TFISF vector, V_{tficf} to be the TFICF vector and V_c to be the CH vector, then we can synthesize them as follows:

1) *CH+TF*: $V_{c+tf} = \langle \lambda_1 V_c, (1-\lambda_1)V_{tf} \rangle$;
2) *CH*+TFISF: $V_{c+tfisf} = \langle \lambda_2 V_c, (1-\lambda_2)V_{tfisf} \rangle$;
3) *CH+TFICF*: $V_{c+tficf} = \langle \lambda_3 V_c, (1-\lambda_3)V_{tficf} \rangle$.

Here, λ_1, λ_2, and λ_3 denote the importance of the corresponding concept-based vectors in the combination forms, respectively. The intuition underlying these combinations is that concepts provide an alternative information channel that should be counted separately and weighted independently from any term observations.

C. Generic Summarization

For generic summarization in disaster management domain, the main task in general is to distill the most important overall information from a set of documents related to the disaster. To emphasize the diversity of topic coverage in a generic summary, we employ the standard K-Means method to cluster the sentences of a document collection into different topical groups, and then apply sentence weighting models within each topical group for sentence selection. In addition, we explore the intrinsic properties of the ontology hierarchy to reduce the information redundancy. Fig. 3 represents the generic summarization framework.

Sentence Selection: We apply the centroid-based methods to select important sentences as the summary. To do so, we run the standard k-Means on the sentence set, where the cluster number k is specified as the number of concepts in the first level of the ontology, i.e., the five nodes lying in L_0 in Fig. 1. The intuition is that the diversity of topics in the whole document collection should be restricted in the concept set of the first level of the ontology. Generally speaking, in the domain of disaster management, the analysts often perform post event analysis on different aspects of a single disaster that they might be interested in. By using the ontology, we are able to separate the sentences into different groups, and also the information not mapped directly to the ontology is filtered out by the procedure of sentence mapping, which is not important in terms of post event analysis. To compute the similarity of sentence pairs, we use the cosine similarity [19] based on the vector space models we discussed in Section IV-B. To obtain a reasonable amount of sentences for reading, we select a set of sentences with the total length of L as the candidate sentences from each cluster. Therefore, the cardinality of the candidate sentence set is $k * L$. By this way, we believe that the informative content in the original sentence set can be kept in the candidate sentence set.

Redundancy Reduction: To reduce the information redundancy, we retrieve the concept nodes that correspond to the sentences in the candidate sentence set of each sentence cluster. Then we compare the two concepts of each sentence pair. Here we are concerned with the sentences linked to the same concept in the ontology hierarchy, which is intuitive because the sentences containing the same concept are very likely to describe the same event. Note that the removal of sentences is based on the evolution of the event. Every sentence in the corpus has a timestamp to indicate the time that the event happens or is updated. If the two sentences being considered have close timestamps, for example, the same day, it is very possible that these two sentences are describing the same status of the corresponding event. In such a case, we randomly remove one sentence; otherwise, we keep both of them in the final summary. By doing this, we can maximally reduce the information redundancy of the candidate sentence set of length $k * L$.

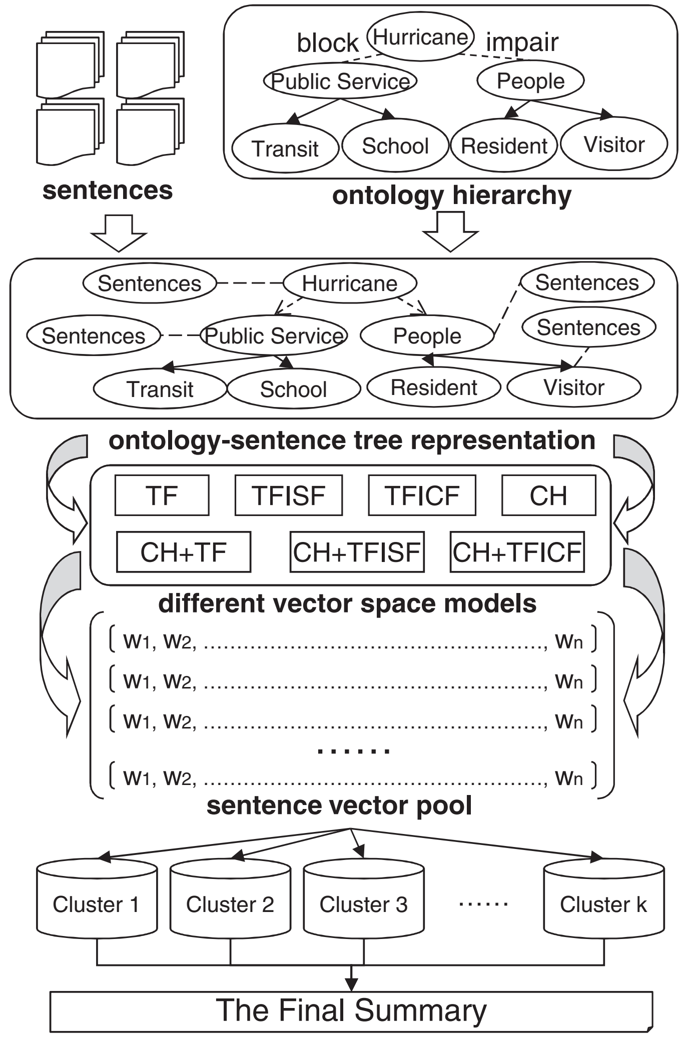


Fig. 3. Generic summarization framework.

Sentence Ranking: After the procedure of reducing information redundancy, we need to rank the remaining sentences to show the importance of each sentence, which is a crucial part of generating text summaries. To this end, we calculate the information content (IC) [20] of the concepts related to each sentence. IC measures the amount of information of a given concept from its probability of occurrence in a corpus. The larger the information content, the more important the concept. IC of a concept is defined as the negation of the logarithm of the probability $p(a)$ of encountering a concept a in a given corpus as

$$IC(a) = -\log p(a). \tag{6}$$

Resnik [20] proposed the premise that IC of the subsumer must be lower than its specializations, since if the IC associated to each ontological concept does not monotonically increase as concepts are specialized, similarity values would be negatively affected. To guarantee this property, the probability of a concept can be calculated as the sum of the individual occurrences of all the concepts which are subsumed by it as follows:

$$p(a) = \sum_{n \in \text{specializations(a)}} \frac{\text{count(n)}}{N} \tag{7}$$

where specializations(a) is the set of terms subsumed by concept a, and N is the total number of concepts observed in the corpus. To rank the sentences in the candidate sentence set, we compute the IC value of each concept correlated to the sentences in the candidate set, and then sort the sentences under the decreasing order of the IC values. The top ranked sentences are selected as the final summary until the summary length L is reached.

D. Query-Focused Summarization

Query-focused summarization aims at generating a short summary based on a given document set and a given query. The generated summary reflects the condensed information related to the given query within the specified summary length. Given a set of documents and a query, the strategy proposed in Section IV-C can also be applied to query-focused summarization tasks. The only difference is that we need to map the given query to the ontology hierarchy so that we can find the concept related to the query and summarize the sentences attached to this concept and its children. A natural property of the ontology is that many concepts in the ontology have equivalent classes with identical or similar semantic meaning to the concept. This property can serve to expanding the query terms in some cases. Given a set of documents, a domain-specific ontology, and a query, the procedure of query-focused summarization is as follows.

1) First, map all the sentences onto the ontology hierarchy under the same criteria described in Section IV-A. Here we introduce equivalent classes to facilitate the mapping procedure. The equivalent classes of a concept can be treated as implicit relevant nodes of this concept. When mapping a sentence to the ontology, we treat each equivalent class as an explicit node.
2) Second, map the given query onto the ontology. If the query matches multiple concepts on the ontology, we treat them as multiple sub-queries,[2] and combine the generated results for the final summary within a given summary length. If the concepts contained in the query have no equivalent classes, the algorithm will stop expansion.
3) Third, calculate the pairwise concept similarities (Resnik's similarity [20]) between the query classes and the corresponding equivalent class, and then select the most similar equivalent class as the expanded content of the original query.
4) Finally, extract all the sentences linked to the original class and the selected equivalent class as the candidate result set, treat such set as one single cluster, and then compute the centroid of this cluster by using vector space models discussed in Section IV-B. The final query result is formed by the top ranked sentences close to the centroid until the summary length is reached.

Intuitively, the sentences linked to the equivalent classes are also relevant to the given query, and therefore when summarizing the query-relevant information, these sentences have the opportunity to be selected as the final summary, which can help cover broader semantic information related to the given query. Hence, the final summary can be enriched by the procedure of query expansion to some extent.

[2]If the concepts are in the same subtree of the ontology and have clear hierarchical structure, we treat the result obtained from the most general one as the final summary.

TABLE II
DESCRIPTION OF THE DATASET

Topic	# of documents
Hurricane Information	105
Public Services	268
Social Events	96
Damage Evaluation	47
Hurricane Recovery	438
Physical Supplies	375
Human Management	233
Emergence Management	138

V. Experimental Results

A. Real World Data

The document set used in our experiments is a collection of press releases from Miami-Dade County Department of Emergency Management and Homeland Security during Hurricane Wilma from Oct. 19, 2005 to Nov. 4, 2005. It contains 1700 documents in total, concerning all the related events before Wilma came, during Wilma and after Wilma passed. For instance, before Wilma came, a myriad of precaution measures were reported, such as the movement of Wilma, the location of evacuation zones, the canceled social activities, etc. The data used in our experiment differ from general newswire documents in a sense that they contain a lot of routine reporting on multiple aspects of the disaster. Specifically, there are eight major topics in this document set, the description of which is listed in Table II. Note that the topics covered in the documents are summarized by domain experts from EOC, as well as the mapping between the documents and the topics. Each document corresponds to a news article, and when it is released, domain experts assign the article to the relevant topic.

In order to compare the quality of the generated summaries by different approaches, we use human generated summaries as references. For hurricane data, we hire five human labelers

TABLE III

STATISTICS OF THE SENTENCE SET

Total # of sentences	25,016
% of sentences containing one concept	61.4%
% of sentences containing two or more concepts	30.5%

(including one domain expert) to manually create five reference summaries for both generic and query-focused[3] summarization tasks. Under the instruction that the summary should epitomize the major events through the hurricane, the human labelers read all 1700 documents for the reference summaries. The summary length (100 words) is the same for all the compared summaries.

B. Evaluation Metrics

In the domain of disaster management, the domain experts are concerned with the status of different aspects relevant to the disaster. They track the tendency of the events during the disaster and after the disaster passed. Therefore, the expected summary will be several sentences describing the status of some events. In some sense, the format of the summary is similar to the one of general summarization tasks, and therefore we can adopt widely used metrics, for example, ROUGE, to evaluate the quality of the summarization result.

C. Evaluation on Sentence Mapping

Sentence mapping is an important step in our proposed ontology-based method for multi-document summarization. In this section, we provide some empirical evaluation on the efficacy of this procedure. The data used in this experiment includes all 1700 documents of our hurricane dataset. Table III shows some statistics after we decompose documents into sentences. Note that besides 91.9% of sentences containing concepts defined in the ontology, 8.1% of sentences do not contain any concept, and therefore this type of sentences would be filtered out by the procedure of sentence mapping.

To evaluate the efficacy of sentence mapping procedure, we record the percentages of different types of sentences after mapping, i.e., sentences mapped to the direct concept, sentences mapped to the LCA etc. We then randomly choose a subset of sentences (100 sentences) from each type of sentence set, and manually evaluate the quality of the mapped sentences. Table IV describes the statistics, associated with the quality measurement (accuracy[4]) by human evaluation.

Most sentences in the dataset are mapped to at least one concept in the ontology, and the quality of the mapping is acceptable under manual evaluation.

D. Generic Summarization

For the generic summarization, we evaluate different vector space representations for comparison. Note that since we have eight general categories at the top level of the domain ontology, we set the length of candidate sentence set as $8 * 100$, which means in each category, a summary with length of 100 words is generated. After removing redundancy and ranking sentences, we iteratively select the top ranked sentences as our final summary, until the length reaches to 100. In order to get better performance, we need to figure out the importance factor λ_1, λ_2, and λ_3, which may result in different experimental results as discussed in Section IV-B. To do so, we conduct experiments on a subset of documents[5] to evaluate the sensitivity of the parameters in different combinations, using ROUGE-2 value as the metric. The factor evaluation results are shown in Fig. 4.

TABLE IV

EVALUATION OF SENTENCE MAPPING

Sentence type	Percentage	Accuracy
Sentences mapped to one direct concept	60.8%	96.5%
Sentences mapped to two or more concepts	10.3%	97.9%
Sentences mapped to the LCA	20.0%	94.2%
Sentences filtered out	8.9%	98.1%
Sentences mapped to at least one concept per document on average	90.4%	–

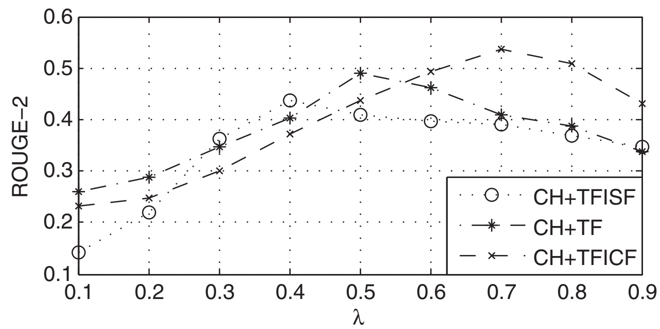


Fig. 4. ROUGE-2 for evaluating λ.

We set λ_1 as 0.5 for *CH+TF*, λ_2 as 0.4 for *CH*+TFISF, and λ_3 as 0.7 for *CH+TFICF*. We implement the following widely used or recent published methods for generic summarization to compare with the methods we discussed in Section IV-C.

1) MEAD [7]: extracts sentences based on centroid value, positional value and first sentence overlap.
2) Latent Semantic Analysis (LSA) [21]: identifies semantically important sentences by conducting latent semantic analysis.
3) Nonnegative Matrix Factorization (NMF) [4]: performs NMF on sentence-term matrix and selects the high ranked sentences.
4) LexPageRank [5]: first constructs a sentence connectivity graph based on cosine similarity and then selects important sentences based on the concept of eigenvector centrality [5].
5) BSTM [22]: a Bayesian sentence-based topic model making use of both the term-document and term-sentence associations.

As for the methods which generate nondeterministic results, we run them 10 times and calculate the average ROUGE score.

[3]The query information is described in Section V-E.

[4]The accuracy here means that the percentage of sentences that belong to the corresponding sentence type over the sentence set.

[5]It contains 400 documents related to hurricane management. We use the remaining documents for summarization.

TABLE V
RESULTS ON GENERIC SUMMARIZATION

System	ROUGE-SU4	ROUGE-L	ROUGE-S*	ROUGE-2
MEAD	0.35152	0.43113	0.12821	0.13541
LSA	0.23611	0.38356	0.09689	0.10461
NMF	0.36478	0.44720	0.13812	0.14576
LexPageRank	0.22642	0.37267	0.11500	0.12167
BSTM	0.45113	0.59259	0.21916	0.22897
TFISF	0.13115	0.28109	0.06490	0.06974
TF	0.22667	0.36842	0.08862	0.09580
TFICF	0.23030	0.37126	0.12299	0.12946
CH	0.32381	0.47170	0.19948	0.20460
CH+TFISF	*0.43166*	*0.56738*	*0.19805*	*0.20724*
CH+TF	*0.47682*	*0.58438*	*0.26355*	*0.27256*
CH+TFICF	**0.54546**	**0.60162**	**0.30827**	**0.31607**

TABLE VI
RESULTS ON QUERY-FOCUSED SUMMARIZATION

System	ROUGE-SU4	ROUGE-L	ROUGE-S*	ROUGE-2
Qs-MRF	0.35676	0.45913	0.16861	0.18059
Wiki	0.33290	0.44835	0.15378	0.17542
SNMF	0.47279	0.51314	0.19806	0.21032
CH+TFISF	0.42034	0.55607	0.18924	0.19442
CH+TF	0.46743	0.57238	0.25205	0.26857
CH+TFICF	0.53290	0.59061	0.29437	0.30549
CH+TFISF(Q)	*0.47825*	*0.51805*	*0.19746*	*0.21653*
CH+TF(Q)	*0.52360*	*0.55019*	*0.25896*	*0.27258*
CH+TFICF(Q)	**0.60058**	**0.60732**	**0.31640**	**0.32908**

Table V presents the experimental comparison of different generic summarization approaches. Bold means the result is statistically significant at the 99% level.

From the comparison results, we observe that: (1) The methods which take into consideration the concepts appearing in the document collection are better than the traditional approaches; (2) The methods using the combinations of different VSMs perform better than the ones solely using one VSM, since the information relevant to the concept hierarchy is integrated into the VSMs, which enriches the representation pattern from both macroscopic and microscopic perspectives; (3) The method with *CH+TFICF* vector representation clearly outperforms the other rivals. Upon the above analysis, we conclude that for the generic summarization in disaster management, the concepts contained in the disaster-related document collection provide substantial benefits for our summarization results.

In order to evaluate how other system designs, such as sentence mapping and redundancy removal, can impact the summarizer, we conduct the experiments as follows:

1) for sentence mapping, we compare the quality of the summaries using LCA and mapping the sentence to different concepts if it contains multiple concepts; and
2) for redundancy removal (RR), we try to evaluate the results with redundancy removal and without redundancy removal. We run the experiments ten times and average the ROUGE-2 scores. Table VII shows the comparison result.

From the comparison, we can observe the following:

1) sentence mapping using Least Common Ancestor can help improve the quality of the summary to some extent, compared with the design that maps the sentences to the corresponding concepts instead of LCA; and
2) redundancy removal can indeed provide high-quality summaries.

TABLE VII
EVALUATION ON DIFFERENT SYSTEM DESIGNS IN GENERIC SUMMARIZATION

System	No LSA	LSA	No RR	RR
CH	0.28175	0.32294	0.22907	0.32350
CH+TFISF	0.41291	0.43215	0.34729	0.43189
CH+TF	0.44706	0.47593	0.40068	0.47626
CH+TFICF	0.51137	0.54431	0.46715	0.54609

E. Query-Focused Summarization

For query-focused summarization, a list of query strings (20 queries) are manually generated for the purpose of the query action. In general, the specified queries contain at least one of the concepts in the ontology hierarchy, so that the proposed method can automatically locate the position of the concept on the ontology. To get the best summarization results, we continue to use the important factors obtained in the generic summarization. We compare the method we discussed in Section IV-D with some widely used and recently published systems.

1) Qs-MRF [11]: extends the mutual reinforcement principle between sentence and term to document-sentence-term mutual reinforcement chain, and uses query-sensitive similarity to measure the affinity between the pair of texts.
2) Wiki [8]: uses Wikipedia as external knowledge to expand query and builds the connection between the query and the sentences in documents.
3) SNMF [4]: calculates sentence-sentence similarities by sentence level semantic analysis, clusters the sentences via symmetric non-negative matrix factorization, and extracts the sentences based on the clustering result. Table VI presents the experimental comparison of different query-focused summarization approaches, Bold means the result is statistically significant at the 99% level. Table VIII lists a set of sample queries and the corresponding system summaries.

Note that "CH+TFISF(Q)," "CH+TF(Q)," and "CH+TFICF(Q)" represent the methods with different VSMs combined with the procedure of query expansion. From the comparison, we have two observations:

1) the methods with query expansion outperform other recently published systems;
2) query expansion enriched summarization methods do improve the summary quality in disaster management domain, particularly, by using the sentence representation of "CH + TFICF." It is straightforward that the procedure of query expansion serves to supplying great opportunity for more sentences relevant to the given query to be selected into the final summary. Therefore, the summarization results are enriched by query expansion in terms of summary quality.

TABLE VIII
SAMPLE QUERIES AND SUMMARIES

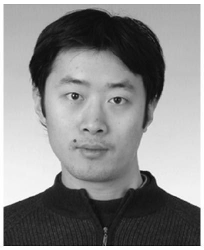
Query	Description	CH+TFICF-based Summary
1	What is the *transit* status?	**Metrobus** has experienced increased demand. **Buses** will operate along the metrorail alignment free of charge. **Parking** at **metrorail stations** will also be free...
2	Is the *airport* open these days?	Homestead **airport** remains closed. passengers should continue checking with their **airlines** regarding specific **flights**...
3	*Evacuation center* status.	Some **distribution points** have been closed. Residents should not gather at designated **distribution centers** unless they have checked for availability...
4	*Schools* at Miami-Dade county.	**Public schools** are closed Monday and Tuesday. Miami dade **public schools** will be open tomorrow : Thursday, October 20...

VI. CONCLUSION AND FUTURE WORK

In this paper, we gave an empirical study on several approaches that utilize the ontology to solve different multidocument summarization problems in disaster management domain. For generic summarization, we employed different vector space models to represent sentences in the document collection, and explored the feasibility of different combinations of the VSMs. Then the centroid-based methods were utilized to cluster the sentence set and the important sentences close to the centroids of the sentence clusters are extracted. The final summary was subsequently generated by reducing information redundancy and ranking sentences. For query-focused summarization, we delved into the effect of query expansion in summarization tasks. The ontology is rich in conceptual information related to the specific domain. We will keep working on the issue of ontology-based multidocument summarization, particulary, on some other document summarization tasks, i.e., update summarization and comparative summarization. Another interesting direction is to explore deeply how to utilize the hierarchical correlations in the ontology to further improve the quality of the summary and to perform hierarchical text categorization [24], [23]. In addition, we will try to employ information extraction techniques to further improve summarization results. We are also interested in extending our proposed method to the summarization using public ontologies, for example, WordNet and Wikipedia. The generality and scalability issues should be taken into account for further extension.

REFERENCES

[1] E. Klien, M. Lutz, and W. Kuhn, "Ontology-based discovery of geographic information services—An application in disaster management," *Comput., Environ. Urban Syst.*, vol. 30, no. 1, pp. 102–123, 2006.
[2] H. Hsu, C. Tsai, M. Chiang, and C. Yang, "Topic generation for web document summarization," in *Proc. IEEE SMC*, 2008, pp. 3702–3707.
[3] X. Yong-dong, W. Xiao-long, L. Tao, and X. Zhi-ming, "Multi-document summarization based on rhetorical structure: Sentence extraction and evaluation," in *Proc. IEEE SMC*, 2008, pp. 3034–3039.
[4] D. Wang, T. Li, S. Zhu, and C. Ding, "Multi-document summarization via sentence-level semantic analysis and symmetric matrix factorization," in *Proc. SIGIR*, 2008, pp. 307–314.
[5] G. Erkan and D. Radev, "Lexpagerank: Prestige in multi-document text summarization," in *Proc. EMNLP*, vol. 4, 2004, pp. 365–371.
[6] X. Wan and J. Yang, "Multi-document summarization using cluster-based link analysis," in *Proc. SIGIR*, 2008, pp. 299–306.
[7] D. Radev, H. Jing, M. Sty, and D. Tam, "Centroid-based summarization of multiple documents," *Inf. Process. Manage.*, vol. 40, no. 6, pp. 919–938, 2004.
[8] V. Nastase, "Topic-driven multi-document summarization with encyclopedic knowledge and spreading activation," in *Proc. EMNLP*, 2008, pp. 763–772.
[9] C. Lee, Z. Jian, and L. Huang, "A fuzzy ontology and its application to news summarization," *IEEE Trans. Syst., Man, Cybern., B Cybern.*, vol. 35, no. 5, pp. 859–880, Oct. 2005.
[10] H. Saggion, K. Bontcheva, and H. Cunningham, "Robust generic and query-based summarisation," in *Proc. ECAL*, 2003, pp. 235–238.
[11] F. Wei, W. Li, Q. Lu, and Y. He, "Query-sensitive mutual reinforcement chain and its application in query-oriented multi-document summarization," in *Proc. SIGIR*, 2008, pp. 283–290.
[12] X. Wan, J. Yang, and J. Xiao, "Manifold-ranking based topic-focused multi-document summarization," in *Proc. IJCAI*, 2007, pp. 2903–2908.
[13] J. Tang, L. Yao, and D. Chen, "Multi-topic based query-oriented summarization," in *Proc. SDM*, 2009.
[14] A. Haghighi and L. Vanderwende, "Exploring content models for multi-document summarization," in *Proc. HLT-NAACL*, 2009, pp. 362–370.
[15] H. Daumé and D. Marcu, "Bayesian query-focused summarization," in *Proc. ACL*, vol. 44, no. 1. 2006, p. 305.
[16] D. Jurafsky, J. Martin, A. Kehler, K. Vander Linden, and N. Ward, *Speech And Language Processing*. Englewood Cliffs, NJ, USA: Prentice-Hall, 2000.
[17] S.-T. Yuan and J. Sun, "Ontology-based structured cosine similarity in speech document summarization," in *Proc. WI*, 2004, pp. 508–513.
[18] S. Yuan and J. Sun, "Ontology-based structured cosine similarity in document summarization: With applications to mobile audio-based knowledge management," *IEEE Trans. Syst., Man, Cybern., B Cybern.*, vol. 35, no. 5, pp. 1028–1040, Oct. 2005.
[19] P. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*. Boston, MA, USA: Pearson Addison Wesley, 2006.
[20] P. Resnik, "Using information content to evaluate semantic similarity in a taxonomy," Arxiv preprint cmp-lg/9511007, 1995.
[21] Y. Gong and X. Liu, "Generic text summarization using relevance measure and latent semantic analysis," in *Proc. SIGIR*, 2001, pp. 19–25.
[22] D. Wang, S. Zhu, T. Li, and Y. Gong, "Multi-document summarization using sentence-based topic models," in *Proc. ACL-IJCNLP*, 2009, pp. 297–300.
[23] T. Li, S. Zhu, and M. Ogihara. "Text categorization via generalized discriminant analysis," *Inf. Process. Manage.*, vol. 44, no. 5, pp. 1684–1697, 2008.
[24] T. Li, S. Zhu, and M. Ogihara. "Hierarchical document classification using automatically generated hierarchy," *J. Intell. Inf. Syst.*, vol. 29, no. 2, pp. 211–230, 2007.
[25] L. Li, D. Wang, C. Shen, and T. Li, "Ontology-enriched multi-document summarization in disaster management," in *Proc. SIGIR*, 2011, pp. 819–820.
[26] L. Zheng, C. Shen, L. Tang, T. Li, S. Luis, S. Chen, and V. Hristidis, "Using data mining techniques to address critical information exchange needs in disaster affected public-private networks," in *Proc. SIGKDD*, 2010, pp. 125–134.



Lei Li received the M.S. degree in software engineering from Beihang University, Beijing, China, in 2008. He is currently pursuing the Ph.D. degree with Florida International University, Miami, FL, USA.

His current research interests include data mining, machine learning, information retrieval, and building recommender systems.

Tao Li received the Ph.D. degree in computer science from the Department of Computer Science, University of Rochester, Rochester, NY, USA, in 2004.

He is currently an Associate Professor with the School of Computing and Information Sciences, Florida International University, Miami, USA. His current research interests include data mining, computing system management, information retrieval, and machine learning.

Dr. Li was a recipient of the NSF CAREER Award and multiple IBM Faculty Research Awards.