Geo-temporal Information Retrieval Based on Semantic Role Labeling and Rank Aggregation

Yoonjae Jeong, Gwan Jang, Kyung-min Kim and Sung-Hyon Myaeng
Korea Advanced Institute of Science and Technology
291 Daehak-ro (373-1 Guseong-dong), Yuseong-gu, Daejeon 305-701,
Republic of Korea
{hybris, gjang, kimdarwin, myaeng}@kaist.ac.kr

ABSTRACT
The purpose of this study is to improve effectiveness of geo-temporal information retrieval with semantic role labeling (SRL) for sentences in topics and documents, especially focusing on locational and temporal facets. We propose a combination of four language models (LM) representing different semantic roles and scopes of models for documents and a rank aggregation method. The rationale is based on observation that sentence-based language models using SRL retrieved relevant documents that are not ranked high by a general LM approach. Although we did not get the comparison result between the general model and our proposed method from NTCIR-9 minutely, we obtained meaningful improvement with the NTCIR-8 GeoTime corpus. Given that the current result is based on our initial effort under the time limitation, we believe that further exploration along the idea of using SRL would give a significant improvement in the geo-temporal information retrieval.

Categories and Subject Descriptors
H.3.3 [**Information Systems**]: Information Search and Retrieval—retrieval models, search process

General Terms
Experimentation, Performance, Measurement

Keywords
Geo-temporal Search, Geographic Information Retrieval, IR evaluation, Rank Aggregation, Semantic Role Labeling

1. INTRODUCTION
We participated in the NTCIR-9 GeoTime task, which is about geographic and temporal search in news articles. Although the task has both Japanese and English sub-tracks, we participated in the English sub-track only.

In the task, a topic asks for geographic- and temporal-based information, such as specification of the location and time of an event. For example a topic asks for information on where and when a particular event occurred or what event happened at a specific time and location. Therefore, the topics may be provided in the form of "where, when, and what did <*entities*> <*action*>?"

It turns out that the elements in the form correspond to semantic roles of linguistic constituents in a sentence. A semantic role is the underlying relationship that a participant (linguistic constituent) has with the main verb in a clause. In the form, where and when are reflected in AM-LOC and AM-TMP roles, which refer to location and temporal information, respectively. The <*entities*> are included in the numbered argument such as Agent and Patient, and <*action*> can become verbs.

Our basic idea is to add locational and temporal aspects to terms in a document using Semantic Role Labeling [18]. By classifying terms in a document according to their semantic roles, we might be able to increase the performance of geo-temporal information retrieval. In other words, our assumption is that it would be helpful to arrange documents based on the degree to which they match the semantic structure of a topic.

In this paper, we propose a new scheme using the language modeling approach based on semantic roles and rank aggregation. We expect that the proposed method improves effectiveness of general information retrieval approaches. For this, we propose four language models: Basic Document Language Model (BDLM), Role-based Document Language Model (RDLM), Basic Sentence Language Model (BSLM), and Role-based Sentence Language Model (RSLM). BDLM is used for baseline performance whereas RDLM attempts to utilize semantic roles of terms in documents. BSLM is similar to BDLM except that language models are constructed not for documents but for sentences. RDLM and RSLM are alike, but RSLM is based on sentences in the way BSLM is compared to BDLM.

In our test using the NTCIR-8 corpus, the four language approaches show different characteristics. RSLM has the advantage of finding documents having the sentences whose semantic structures expressed in terms of semantic roles are analogous to those of the topic. On the contrary, BDLM is much related to the term frequencies of a topic. The former reveals the property of information extraction while the latter shows the nature of general information retrieval. RDLM and BSLM are about a half way between RSLM and BDLM.

With rank aggregation, we combine the ranked lists from the four language modeling approaches. The reason is because the four models retrieve different relevant documents. However, only some of top ranking documents are relevant in RDLM, BSLM, and RSLM. It means that the constraint, only to use the top ranks, is required when the result of BDLM is combined with the other models. Thus we applied threshold values to the rank aggregation.

The rest of the paper is organized as follows. The discussion of related work is presented in Section 2. Section 3 is followed by the proposed method that we devised for geo-temporal information retrieval. Then we discuss our submitted runs and analyze those results from NTCIR-9 in Section 4. Finally, Section 5 presents our conclusions.

2. RELATED WORK
Geographic Information Retrieval (GIR) is a research area where it finds documents related to specified areas through not only keywords but also geographic constraints. It is concerned with the retrieval of thematically and geographically relevant information

resources in response to a query of the form {<*theme* or *topic*>, <*spatial relationship*>, <*location*>” [12]. GIR systems are based on a collection of geo-referenced information resources, and search those resources with a geographical location as the key. Documents are indexed according to their geographic regions, where their specific locations are encoded either directly as spatial coordinates, or indirectly by place name [9].

Adding to the locational aspect, the temporal aspects are also considered in this work. Three recent workshops, NTCIR GeoTime [7], GeoCLEF [15], and GIR [21] addressed the combination of geographic and temporal search. In the most recent NTCIR-8 GeoTime workshop, INESC group from Lisbon, Portugal [14] shows the best performance among the participants. They used geographic resources and TIMEXTAG system [23] to extract geographic expressions from topics and documents.

Semantic Role Labeling (SRL) is a task in natural language processing, whose goal is to detect the semantic roles of the arguments associated with the predicate or verb of a sentence. It has been a popular task since the availability of the PropBank and FrameNet annotated corpora [19]. The seminal work of Gildea and Jurafsky [8] and CoNLL evaluation campaigns [2] accelerated research along this line. In this area, statistical machine learning methods, ranging from joint probabilistic models to support vector machines, have been successfully adopted to provide very accurate semantic labeling.

Rank aggregation is to combine ranking results of documents from multiple ranking functions in order to generate a better one. Rank aggregation can be classified into two categories [1]. The first one is the score-based aggregation [6] where documents in individual ranked lists are assigned scores, which are used by a rank aggregation function. The second one is order-based aggregation where only the orders of the entities in individual ranked lists are used by the aggregation function. The order-based aggregation usually takes an unsupervised learning approach. Popular aggregation functions include Borda Count [1][4], median rank aggregation [5], genetic algorithm [22], Markov Chain based rank aggregation [4]. However, some approaches such as Borda Fuse [1] and Liu, et al.’s approach [13] make use of training data. The latter is a new supervised rank aggregation method based on Markov Chain.

3. THE PROPOSED METHOD

The main goal of our research is to devise an effective method for the geo-temporal information retrieval using both SRL and rank aggregation. Figure 1 illustrates an overview of the proposed method. Both documents and topics are processed for SRL and represented with terms and their semantic roles. Topics are processed further to identify the question types. For the purpose of matching topic and document representations, we propose four variations of language modeling.

Basic Document Language Model (BDLM) is an ordinary term-based language modeling approach that serves as the baseline. To emphasize the locational and temporal aspects, we propose Role-based Document Language Model (RDLM), which enforces IR model to focus more on events (verbs) and locational and temporal facets acquired by SRL.

For a geo-spatially focused topic, relevant information is sometimes found in a single sentence within a document as in question answering. In this case, the rest of the document may simply serve as noise for the topic. Basic Sentence Language Model (BSLM) is devised to deal with this situation by reducing the granularity of the text for which language models are constructed. Role-based Sentence Language Model (RSLM) adds semantic roles to BSLM like RBLM.

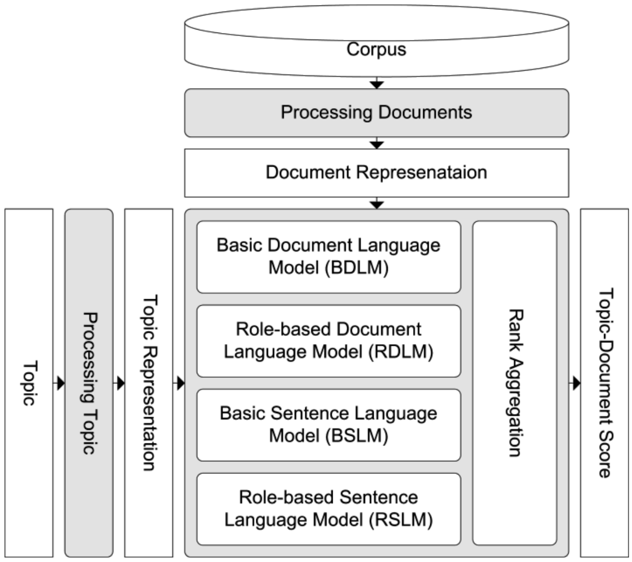

Figure 1. Overview of the proposed geo-temporal information retrieval

The four retrieval models (BDLM, RDLM, BSLM, and RSLM) have different advantages and disadvantages. As in Table 1, which shows a case (topic GeoTime-0025 for the NTCIR-8 corpus) of some relevant documents and their ranks, the four models are complementary. For example, the first two documents that are highly ranked by RSLM 4th and 3rd), are ranked very low by BDLM. On the other hand, the five documents ranked within top 10 by BDLM are ranked very low by RSLM. The parentheses in the table indicate their normalized scores of relevance.

Table 1. Relevant documents and their ranks and normalized scores of each model for Topic GeoTime-0025 in NTCIR-8 corpus

Relevant Document	Rank & Normalized Score (Z-score [10])			
	BDLM	**RDLM**	**BSLM**	**RSLM**
NYT_ENG_2 0041226.0096	24 (8.95E-01)	201 (1.54E+00)	251 (3.72E-02)	**4** **(2.82E+02)**
NYT_ENG_2 0041229.0208	167 (1.07E-04)	322 (1.07E+00)	133 (9.18E-02)	**3** **(2.86E+02)**
NYT_ENG_2 0041230.0186	**3** **(2.20E+02)**	18 (4.46E+01)	298 (3.09E-03)	102 (7.74E-01)
NYT_ENG_2 0041230.0204	**8** **(8.47E+01)**	26 (3.26E+01)	302 (3.09E-03)	98 (7.75E-01)
NYT_ENG_2 0041230.0245	**4** **(1.21E+02)**	17 (4.46E+01)	299 (3.09E-03)	101 (7.74E-01)
NYT_ENG_2 0041230.0256	**6** **(1.20E+02)**	25 (3.49E+01)	303 (3.09E-03)	100 (7.74E-01)
NYT_ENG_2 0041231.0009	**2** **(2.20E+02)**	19 (4.46E+01)	300 (3.09E-03)	99 (7.75E-01)
NYT_ENG_2 0050328.0205	36 (4.30E-03)	349 (1.07E+00)	88 (9.18E-02)	162 (7.72E-01)

Based on the observation, we realized that it is critical to devise a method for combining the ranked lists of retrieved documents from the four models and adopted a rank aggregation method. The rank aggregation module in the system assigns the final scores to the retrieved documents.

3.1 Document Representation

Documents are represented as sets of words for semantic roles. The semantic roles are assigned to verbal arguments in a sentence. They provide rich information in that they specify “Who did what to whom, how, when and where?” for a verb. Figure 2 shows an example of the SRL task, where A1 (who), AM-TMP (when), and AM-LOC (where) parts for the verb “die” are identified. We adopted SENNA [3] to label semantic roles of terms. While it shows the similar effectiveness with the state of the art approaches, it has good efficiency because it is written in C language.

```
…
[A1 Astrid Lindgren, the Swedish writer
whose rollicking, anarchic books about
Pippi Longstocking horrified a generation
of parents and captivated millions of
children around the globe], died in her
sleep [AM-TMP Monday] [AM-LOC at her home
in Stockholm, Sweden.]”
…
```

Figure 2. An example of SRL for a document

The proposed document representation consists of four attributes as summarized in Table 2. The attribute T_V contains all verbs in a document, and T_A means a set of terms in numbered argument roles (A0-5), which are argument elements for a verb. For example, in a sentence “John broke the window” *John* and *the windows* are the arguments for a verb *broke*. It contains Agent and Patient roles. $T_{AM\text{-}LOC}$ and $T_{AM\text{-}TMP}$ include terms with locational (AM-LOC) and temporal (AM-TMP) roles, respectively. The detailed information of semantic roles is defined in the CoNLL-2005 SRL task [2].

Table 2. Document representation

Attribute	Description
T_V	A set of verb in document e.g., *die*
T_A	A set of terms with numbered argument roles (A0-5) in document e.g., *Astrid*, *Lindgren*, … , *children*, *globe*
$T_{AM\text{-}LOC}$	A set of terms with location (AM-LOC) roles in document e.g., *home*, *Stockholm*, *Sweden*
$T_{AM\text{-}TMP}$	A set of terms with *temporal* role (AM-TMP) in document e.g., *Monday*

3.2 Topic Representation

A topic is represented with its question types (*Q-LOC*, *Q-TMP*, *Q-AGT*, or *Q-MSC*) and a set of vocabularies and their semantic roles among the four (V, A0-5, AM-LOC, and AM-TMP). The question types indicate what entity a topic asks for, and the sets of vocabularies include the lemmas of terms for each of the roles. Table 3 shows the attributes and their descriptions and examples for topic representation.

Table 3. Topic representation

Attribute		Description
Question types	*Q-LOC*	Whether a question is about location or not? e.g., *When and where did Astrid Lindgren die?*
	Q-TMP	Whether a question is about time or not? e.g., *When and where did Hurricane Katrina make landfall in the United States?*
	Q-AGT	Whether is a question about agent or not? e.g., *What Portuguese colony was transferred to China and when?*
	Q-MSC	The others e.g., *How old was Max Schmeling when he died, and where did he die?*
Set of terms	V_V	A set of vocabularies in verb role in topic
	V_A	A set of vocabularies in numbered argument (A0-5) roles in topic
	$V_{AM\text{-}LOC}$	A set of vocabularies in locational role (AM-LOC) in topic
	$V_{AM\text{-}TMP}$	A set of vocabularies in temporal role (AM-TMP) in topic

To determine the question types, we devised some heuristic rules based on syntactic parser results. For example, we first found SBARQ tree from the parsing result of a topic, and then examined what interrogative forms exists in its WHADVP sub-tree. Figure 3 shows an example that illustrates how a question type can be found. We used the Stanford parser [11] and SENNA [3] to parse topics and assign semantic roles, respectively.

```
Topic: When and where did Astrid Lindgren die?
Parsing Tree:
(ROOT
  (SBARQ
    (WHADVP (WRB When)
      (CC and)
      (WRB where))
    (SQ (VBD did)
      (NP (NNP Astrid) (NNP Lindgren))
      (VP (VB die)))
    (. ?)))
```

Figure 3. A parse-tree example for question type identification

3.3 Information Retrieval Models

3.3.1 Basic Document Language Model (BDLM)

We first consider a basic language model to guarantee the baseline performance. According to Ponte and Croft [20], language modeling for information retrieval tasks provides a probabilistic way of modeling the retrieval process and achieves good performance. Equation (1) shows the Basic Document Language Model (BDLM). $P_{BDLM}(q|d)$ is the probability for a document d given a topic q where t is a term in q.

$$P_{BDLM}\left(d|q\right) = P\left(q|d\right) \times \frac{P(d)}{P(q)} \approx \prod_{t \in q} P\left(t|d\right) \quad (1)$$

We also use the Dirichlet smoothing method [25], which adjusts the maximum likelihood estimator so as to correct the inaccuracy due to data sparseness. In Equation (2), $tf_{t,d}$ is the term frequency of term t in document d, and D means the set of all documents in the corpus. We set the smoothing parameter μ to 2500 because it is the default value of INDRI [17], which is a well-known search engine based on language modeling.

$$P(t|d) = \frac{tf_{t,d} + \mu \times P(t|D)}{|d| + \mu} \quad (2)$$

3.3.2 Role-based Document Language Model (RDLM)

Based on the document representation, we built the Role-based Document Language Model (RBLM) as in Equation (3) where R represents the semantic roles in q. q_r and d_r are the sets of terms given the role r in q and d, respectively.

$$P_{RDLM}(d|q) = \prod_{r \in R} \left(P(q_r|d_r) + \alpha \right)$$
$$\begin{cases} \alpha = 1, \text{ if Q-LOC is } \textit{true}, r\text{=AM-LOC and } |d_r| > 0. \\ \alpha = 1, \text{ if Q-TMP is } \textit{true}, r\text{=AM-TMP and } |d_r| > 0. \\ \alpha = 0, \text{ otherwise} \end{cases} \quad (3)$$

It is possible there is no term with AM-LOC or AM-TMP role in a topic because an interrogative (e.g., where, when) is in its place. It is therefore necessary to handle this case with a weight as in the condition part of Equation (3). When q asks for a location or time (i.e. Q-LOC or Q-TMP are true) and d contains terms in the related roles (i.e. AM-LOC or AM-TMP), we apply the weight α (=1).

On the contrary to this, the value of $P(q_r|d_r)+\alpha$ may become larger than 1 when there are another location or temporal terms in a topic (e.g., GeoTime-35 "When and where did a pipeline explosion occur in Africa ..."). In this case, the terms become the detailed information of location or time for the topic. When a document includes the detailed information, the value over 1 is the weighted result.

3.3.3 Basic Sentence Language Model (BSLM)

Sometimes the relevant information related to a topic is fully contained in one sentence in document. The Basic Sentence Language Model (BSLM) is devised to handle this case.

$$P_{BSLM}(d|q) = \max_{s \in S} P(s|q) = \max_{s \in S} \prod_{t \in q} P(t|s) \quad (4)$$

Equation (4) shows the probability of document d given topic q as in the BSDM model. Instead of building a language model for a document, however, we attempt to compute the probability of a sentence given a query $P(s|q)$ and take the maximum among those computed for individual sentences in the document, under the assumption that the sentence is likely to contain relevant information. Here, S is a set of sentence in d, and s is a sentence in S.

3.3.4 Role-based Sentence Language Model (RSLM)

The Role-based Sentence Language Model (RSLM) adds semantic roles to BSLM in the same way RBLM was constructed out of BDLM. Based on the sentence representation, RSML is expressed in Equation (5) where R is the semantic roles in q. q_r and s_r are the sets of terms given the role r in q and s, respectively.

$$P_{RSLM}(d|q) = \max_{s \in S} \prod_{r \in R} \left(P(q_r|s_r) + \alpha \right)$$
$$\begin{cases} \alpha = 1, \text{ if Q-LOC is } \textit{true}, r\text{=AM-LOC and } |s_r| > 0. \\ \alpha = 1, \text{ if Q-TMP is } \textit{true}, r\text{=AM-TMP and } |s_r| > 0. \\ \alpha = 0, \text{ otherwise} \end{cases} \quad (5)$$

3.4 Rank Aggregation

As we mentioned in Section 1, the proposed information retrieval models have difference characteristics, which can be combined to handle various retrieval cases by devising a rank aggregation method that combines ranked lists obtained from four different retrieval models. In Dwork, et al.'s work [4], the Markov Chain based approaches showed the best performances in their experiment. They proposed four heuristic Markov Chains (MC1, MC2, MC3, and MC4). We adopted MC2 because it is arguably the most representative of minority viewpoints of sufficient statistical significance; it protects specialist views [4]. The transitions in Markov Chain are defined as follows.

If the current state is i, then we first select a ranking list τ_k from the ranking lists $\tau_1, \ldots \tau_l$ that contain state i, then select state j randomly from the set of states that are ranked not lower than state i in τ_k, and define j as the next state. T_k, the transition matrix produced by ranking list τ_k, is denoted in Equation (6). $t_{ij}^{(k)}$, the element of T_k, the conditional probability of state j given state i in ranking list $j >_{\tau_k} i$ means that document j is ranked higher than document i in ranking list τ_k.

$$T_k \triangleq \left(t_{ij}^{(k)} \right)_{n \times n}$$
$$t_{ij}^{(k)} = \begin{cases} \dfrac{1}{\#\{j \mid j >_{\tau_k} i \text{ or } i = i\}}, & j >_{\tau_k} i \text{ or } j = i \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

The final transition matrix T is calculated by the average of the individual transition matrix T_k as in Equation (7). l is the number of ranked lists.

$$T = \frac{1}{l} \sum_{k=1}^{l} T_k \quad (7)$$

However, we found that the effective ranks for aggregations of RDLM, BSLM, and RSLM are a small number of top ones. We applied the threshold to the elements of transition matrix. Equation (8) shows the final element of transition matrix for those models.

$$t_{ij}^{(k)} = \begin{cases} \dfrac{1}{\#\{j \mid j >_{\tau_k} i \text{ or } i = i\}}, \\ \quad j >_{\tau_k} i \text{ or } j = i \text{ and z-score}(i) \geq \theta \\ 0, \quad \text{otherwise} \end{cases} \quad (8)$$

To normalize the scores, we use z-score (or standard score) [10]. In statistics, a z-score indicates how many standard deviations an observation or datum is above or below the mean. It allows for comparison of observations from different normal distributions. Based on the normalized score, we choose the only element of transition matrix of which the score is above the given threshold θ.

Finally, the score vector x can then be computed by Equation (9). x_0 is the initial score. We set the values as $1/|D|$, and $|D|$ is the number of all documents.

$$x = T^T x_0 \quad (9)$$

4. EVALUATION

4.1 Tests in NTCIR-8 Corpus

We first evaluated the proposed method using the NTCIR-8 GeoTime corpus by comparing BDLM and the rank aggregation methods with eight cases having different threshold values of θ. In this experiment, we used only *description* fields of topics.

Figure 4 gives a full detail about the comparisons among the eight different cases. The meaningful improvements were observed across all aggregations regardless of the threshold value. The changes in terms of nDCG@100 are from 0.4087 to 0.4925. In this experiment, the best performance is when θ is 200 or 150, but the differences are very small.

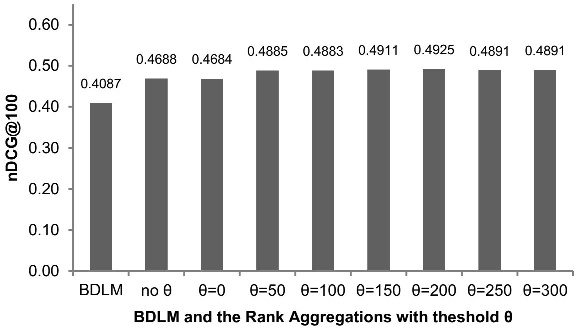

Figure 4. Test results using the NTCIR-8 GeoTime corpus

4.2 Description of Runs in NTCIR-9

For formal evaluation, we submitted four runs as in Table 4. IRNLP-EN-EN-1-D and IRNLP-EN-EN-2-D use the *description* fields of topics. On contrary, IRNLP-EN-EN-3-DN and IRNLP-EN-EN-4 include both *description* and *narrative* fields for queries. The results of all runs are retrieved by the aggregation model that combines ranked lists form the BDLM, RDLM, BSLM, and RSLM models. The aggregation thresholds of θ are set to 150 and 200 because they showed the best performance in our test of using NTCIR-8 GeoTime corpus.

Table 4. Summited Runs

RUN	Topic Source	Aggregation	Aggregation Threshold (θ)
IRNLP-EN-EN-1-D	description only	(BD, RD, BS, & RS) LM	150
IRNLP-EN-EN-2-D	description only	(BD, RD, BS, & RS) LM	200
IRNLP-EN-EN-3-DN	description & narrative	(BD, RD, BS, & RS) LM	200
IRNLP-EN-EN-4-DN	description & narrative	(BD, RD, BS, & RS) LM	140

4.3 Results from NTCIR-9

All four of our submitted runs are slightly below the median of the English-only subtask across all five metrics (MAP, Q, nDCG@10, nDCG@100, and nDCG@1000). Table 5 shows the results in different metrics for each of our runs and the median of teams in the NTCIR-9 GeoTime English subtask. Our highest result for each metric is shown in boldface.

We also added the two baselines (BDLM-D and BDLM-DN) to confirm the effectiveness of using semantic roles. BDLM-D is a baseline using BDLM and the only *description*. In BDLM-DN, we use the both of *description* and *narrative*.

Table 5. Overall metrics for submitted NTCIR GeoTime Runs

RUN	MAP	Q	nDCG @10	nDCG @100	nDCG @1000
BDLM-D	0.2710	0.2844	0.3972	0.4003	0.5400
IRNLP-EN-EN-01-D	0.2999	0.3242	0.4257	0.4237	0.5448
IRNLP-EN-EN-02-D	0.2981	0.3225	0.4257	0.4215	0.5430
BDLM-DN	0.2924	0.2959	0.4002	0.3984	0.5460
IRNLP-EN-EN-03-DN	**0.3128**	**0.3354**	0.4343	**0.4281**	**0.5544**
IRNLP-EN-EN-04-DN	0.3123	0.3351	**0.4358**	0.4270	0.5538
Median of NTCIR-9	0.3326	0.3512	0.4591	0.4563	0.5772

In Table 5, IRNLP-EN-EN-03-DN slightly surpassed the others in all cases except nDCG@10. We used both *description* and *narrative* and set the threshold θ to 200. Rather than using description only, combining description and narrative together was better in our proposed method. We think that it is because the narrative fields enrich the query models.

We looked at the performance differences per topics. Topic GeoTime-0026, -0033, -0034, -0039, and -0046 show good performance over 0.7 in terms of nDCG@10 and overcome the baselines not using semantic roles On the other hand, GeoTime-0027, -0028, -0035, -0037, -0044, and -0045 have poor results below 0.3. Details are illustrated in Figure 5.

Table 6 shows the topics that gave high performance in terms of nDCG@10, whose scores are over 0.7, and their scores are higher than those of the baselines. The topics are analyzed without errors by the proposed heuristics based on the natural language processing, and those verbs are related to the activities or states of agents clearly (e.g. “murder”, “hijack”, kill”, and so on). The terms are also not ambiguous because they are proper nouns or very specific number of theme (e.g. “4 people” in GeoTime-0033).

Table 6. The topics showing high performances (nDCG@10 > 0.7 and higher scores than the baselines)

Topic	Topic description
GeoTime-0026	Where and when did the space shuttle Columbia disaster take place?
GeoTime-0033	When and where were 4 people murdered and many others sickened by arsenic poisoning?
GeoTime-0034	When and from what airport was an ANA plane hijacked and a pilot killed?
GeoTime-0039	When and where did a nuclear submarine sink, killing over 100 crew members?
GeoTime-0046	Where and when did presidential debate between Bush and Kelly hold?

When we reviewed the topics with the low performance (nDCG@10 < 0.4), they showed many errors in the analysis of topics (e.g., GeoTime-0028, -0045, and -0050). Furthermore, the verbs were related to the existence or occurrence of agent or theme (e.g., GeoTime-0035 and -0044). They sometimes require inference or term expansion, as in GeoTime-0037, -0042, and -0044. The topics are listed in Table 7.

Table 7. Topics showing low performances (nDCG@1000 < 0.4)

Topic	Topic description
GeoTime-0028	When and where were the Washington beltway snipers arrested?
GeoTime-0035	When and where did a pipeline explosion occur in Africa killing over 500 people?
GeoTime-0037	What fatal accident occurred near (geographical coordinates 5°52′12″N 5°45′00″E / 5.870°N 5.750°E / 5.870; 5.750), which killed hundreds of people, and when did it occur?
GeoTime-0042	Describe the name of the country of Middle East whose King died in 1999.
GeoTime-0043	When was the last time the New England Patriots won the Super Bowl?
GeoTime-0044	Describe when and where deadly earthquakes happened in South America?
GeoTime-0045	When the European Central Bank was established and where is its headquarter?
GeoTime-0050	When and where was CAFTA, the Central America Free Trade Agreement signed?

5. CONCLUSION

In this paper, we propose a new geo-temporal information retrieval method that utilizes semantic role labeling and rank aggregation. We believe that it is useful to analyze documents for semantic roles around main predicates of sentences and generate language models after the analysis. While the SRL-based method is not always superior across different topics, they complement the usual language modeling approach and hence warrant the proposed rank aggregation method.

Because this research is an initial study for our basic idea (i.e. geo-temporal information retrieval based on semantic role labeling and rank aggregation), there is much to be done to improve its effectiveness. We only used the terms that exist in topics for our model, but found through an analysis of the result that term expansion and weighting are necessary. The use of external knowledge or resources is also required (especially to a topic, GeoTime-0037). Parameter optimization for the rank aggregation part is also in our agenda. Furthermore, automatic topic analysis needs to be improved.

6. ACKNOWLEDGMENTS

This research was supported by the MKE (The Ministry of Knowledge Economy), Korea, under the ITRC (Information Technology Research Center) support program supervised by the NIPA (National IT Industry Promotion Agency) (NIPA-2011-C1090-1111-0008)

7. REFERENCES

[1] J. A. Aslam and M. Montague, "Models for metasearch," in *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '01)*, New Orleans, Louisiana, USA, 2001.

[2] X. Carreras and L. Màrquez, "Introduction to the CoNLL-2005 Shared Task: Semantic Role Labeling," in *Proceedings of the Ninth Conference on Computational Natural Language Learning (CoNLL-2005)*, Ann Arbor, Michigan, USA, 2005.

[3] R. Collobert and J. Weston, "SENNA," NEC Laboratories America, Inc., 6 February 2011. [Online]. Available: http://ml.nec-labs.com/senna/. [Accessed 22 September 2011].

[4] C. Dwork, R. Kumar, M. Naor and D. Sivakumar, "Rank aggregation methods for the Web," in *Proceedings of the 10th international conference on World Wide Web (WWW '01)*, Hong Kong, Hong Kong, 2001.

[5] R. Fagin, R. Kumar and D. Sivakumar, "Efficient similarity search and classification via rank aggregation," in *Proceedings of the 2003 ACM SIGMOD international conference on Management of data (SIGMOD '03)*, San Diego, California, USA, 2003.

[6] E. A. Fox and J. A. Shaw, "Combination of multiple searches," in *Proceedings of the 2nd Text REtrieval Conference (TREC-2)*, Gaithersburg, Maryland, USA, 1994.

[7] F. Gey, R. R. Larson, N. Kando, J. Machado and T. Sakai, "NTCIR-GeoTime Overview: Evaluating Geographic and Temporal Search," in *Proceedings of the 8th NTCIR Workshop Meeting on Evaluation of Information Access Technologies (NTCIR-8),* Tokyo, Japan, 2010.

[8] D. Gildea and D. Jurafsky, "Automatic labeling of semantic roles," *Computational Linguistics*, vol. 28, no. 3, pp. 245-288, 2002.

[9] L. L. Hill, *Georeferencing: the geographic associations of information*, Cambridge, MA, USA: MIT Press, 2006.

[10] B. R. Kirkwood and J. A. C. Sterne, *Essential medical statistics*, Hoboken, NJ, USA: Wiley-Blackwell, 2003.

[11] D. Klein and C. D. Manning, "Accurate unlexicalized parsing," in *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics (ACL '03)*, Sapporo, Japan, 2003.

[12] R. R. Larson, "Geographic information retrieval and spatial browsing," *GIS and libraries: patrons, maps and spatial information*, pp. 81-124, 1996.

[13] Y.-T. Liu, T.-Y. Liu, T. Qin, Z.-M. Ma and H. Li, "Supervised rank aggregation," in *Proceedings of the 16th international conference on World Wide Web (WWW '07)*, Banff, Alberta, Canada, 2007.

[14] J. Machado, J. Borbinha and B. Martins, "Experiments with Geo-Temporal Expressions Filtering and Query Expansion at Document and Phrase Context Resolution," in *Proceedings of the 8th NTCIR Workshop Meeting on Evaluation of Information Access Technologies (NTCIR-8)*, Tokyo, Japan, 2010.

[15] T. Mandl, P. Carvalho, G. Di Nunzio, F. Gey, R. Larson, D. Santos and C. Womser-Hacker, "GeoCLEF 2008: The CLEF 2008 Cross-Language Geographic Information Retrieval Track Overview," in *Evaluating Systems for Multilingual and Multimodal Information Access*, Springer Berlin / Heidelberg, 2009, pp. 808-821.

[16] R. Manmatha, T. Rath and F. Feng, "Modeling score distributions for combining the outputs of search engines," in *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '01),* New Orleans, Louisiana, USA, 2001.

[17] D. Metzler and W. Croft, "Combining the language model and inference network approaches to retrieval," Information Processing & Management, vol. 40, no. 5, pp. 735-750, 2004.

[18] M. Palmer, D. Gildea and N. Xue, "Semantic Rolrepresente Labeling," *Synthesis Lectures on Human Language Technologies*, vol. 3, no. 1, pp. 1-103, 2010.

[19] M. Palmer, D. Gildea and P. Kingsbury, "The Proposition Bank: An Annotated Corpus of Semantic Roles," *Computational Linguistic*, vol. 31, no. 1, pp. 71-106, 2005.

[20] J. M. Ponte and W. B. Croft, "A language modeling approach to information retrieval," in *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '98)*, Melbourne, Australia, 1998.

[21] J. Strötgen, M. Gertz and P. Popov, "Extraction and exploration of spatio-temporal information in documents," in *Proceedings of the 6th Workshop on Geographic Information Retrieval (GIR '10)*, Zurich, Switzerland, 2010.

[22] M. M. Sufyan Beg, "Parallel rank aggregation for the World Wide Web," in *Proceedings of International Conference on Intelligent Sensing and Information Processing, 2004 (ICISIP '04)*, Chennai, India, 2007.

[23] "Timex Annotation System," 2 March 2007. [Online]. Available: http://ilps.science.uva.nl/resources/timextag. [Accessed 28 September 2011].

[24] C. C. Vogt and G. W. Cottrell, "Fusion Via a Linear Combination of Scores," *Information Retrieval*, vol. 1, no. 3, pp. 151-173, 1999.

[25] C. Zhai and J. Lafferty, "A study of smoothing methods for language models applied to Ad Hoc information retrieval," in *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '01),* New Orleans, Louisiana, USA, 2001.

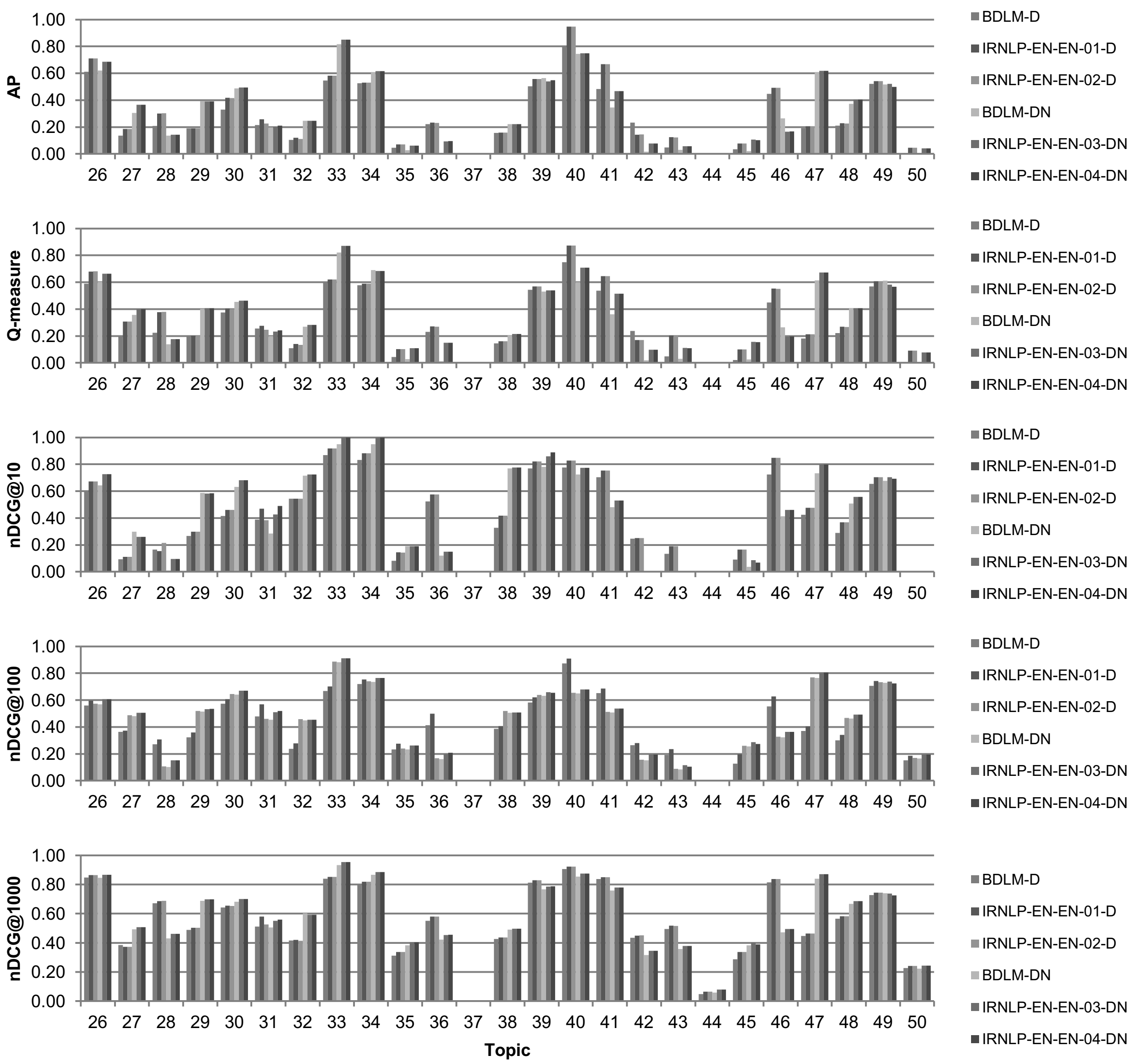

Figure 5. Formal result per topic (from top to bottom: AP, Q-measure, nDCG@10, nDCG@100, and nDCG@1000