A Query-Aware Document Ranking Method for Geographic Information Retrieval

Bo Yu and Guoray Cai

College of Information Sciences and Technology
Pennsylvania State University
University Park, PA 16802, USA

{byu, cai}@ist.psu.edu

ABSTRACT

Geographically oriented search must consider both the thematic and geographic dimensions of relevance when matching documents to queries. We propose a dynamic document ranking scheme to combine the thematic and geographic relevance measures on a per-query basis. Query specificity is introduced to determine the relative weights of different sources of ranking evidence for each query. A preliminary evaluation comparing with human judgment shows that our method to distinguish different types of geo-referenced queries based on query specificity is promising to address the issue of relevance combination in GIR document ranking. In addition, we explore the possibility of using Dempster-Shafer's theory to combine the two different sources of ranking evidence.

Categories and Subject Descriptors

H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing – *indexing methods* H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval – *retrieval models, selection process*

General Terms

Algorithms, Design

Keywords

Geographic search, relevance ranking, query specificity, Dempster-Shafer's theory

1. INTRODUCTION

Ranking documents by their relevance to queries is a common way for search engines to return retrieval results to users. In doing so, a single relevance score must be calculated based on all the measurable evidence. In traditional informational retrieval models (e.g. vector space model), relevance is represented by the similarity metric between a document and a query. The assumptions underlying these models are that (1) the content of a document can be represented by a vector of concept-bearing terms and (2) the similarity between the term vectors of a document and a query measures the relevance of the document in response to the query. However, in geographical information retrieval research, relevance is judged also by the spatial relationships (such as contain, overlap, intersect, connect, near etc.) between documents and queries in geographical space, in addition to similarity in thematic (non-geographic) space. To deal with this complication, the vector space model has recently been extended by a geographic dimension of relevance, resulting in a hybrid geo-thematic IR model, called GeoVSM [4]. GeoVSM explicitly represents and reasons document relevance in two subspaces (geographical space and thematic space), producing two relevance scores respectively. Due to the relative independence and complex interactions between the two document subspaces, it remains unclear how a document ranking method could take these two types of relevance and combine them to retrieve documents in an effective manner. Our current work will focus on this challenging problem and suggest solutions.

There have been several alternative schemes to combine two types of relevance scores into a final ranking score. The most common scheme is the weighted sum of individual scores [1, 4, 8]:

$$\mathrm{Rel}(q,d) = \omega_T * \mathrm{Rel}_T(q,d) + \omega_G * \mathrm{Rel}_G(q,d) \quad (1)$$

where q is a user query, d is a document. Rel_T and Rel_G are functions to calculate the thematic and geographic relevance scores. ω_T and ω_G are weights of these two individual relevance scores respectively. Other combination functions (such as the product or the maximum of two individual scores) are also proposed in [8] to compute the final ranking scores.

In contrast to computational approaches to determining ranking scores, other methods focus on visualizing multiple dimensions of relevance to facilitate human judgment on document relevance. Their assumption is that ranking algorithms cannot provide the final judgment on relevance, as searchers themselves must be the ultimate judges to the relevance of documents [5]. Therefore, they do not combine the different sources of relevance to generate a single ranked list for each query. Instead, they employ multidimensional visualization mechanisms to represent the degrees of relevance and facilitate the users to find relevant documents by themselves. GeoVIBE [3] is one of the attempts to explore the possibilities of visualizing document similarities in both geographic and thematic domains. The system consists of two types of browsing windows, GeoView and VibeView, which display documents based on geographic and thematic relevance respectively. Hobona et al. propose an approach to visualize the degrees of relevance in a three-dimensional visualization

GIR'07, November 9, 2007, Lisbon, Portugal.

environment [5]. In their model, they represent the thematic, geographic, and temporal similarities of a geospatial data set in a 3D vector space.

We believe that relevance ranking on geographical documents is a much more complex problem, and previous methods only provide some partial solutions with their own limitations. On the one hand, visualization approaches provide more freedom for users to express their information search needs and make decisions on the relevance, but sophisticated multidimensional visual interfaces can also cause extra cognitive loads and require more learning efforts. On the other hand, automated computation of document ranking lists comply with users' current search habits and are easier to use for ordinary users, although they are less flexible and adaptive to search contexts. Our work generally follows the later approach, while proposing methods to make relevance ranking more adaptive to queries.

Most of current computational ranking methods in GIR have employed a static way to combine the thematic and geographic relevance, with no concern about the different contexts of users' search behavior. In these methods, no matter what kinds of queries are imposed, the ranking algorithm uses a fixed formula to represent the relative importance of thematic and geographic relevance. However, in reality, geographic and thematic relevance measures may have varying degrees of influence on the document ranking depending on the nature of the queries. In some situations, users may care more about the geographic constrains of documents, therefore the geographic relevance scores should value more than thematic ones; and in other situations, uses may prefer the results that give higher weights to thematic relevance over the geographic ones. Based on the experiments with the GeoCLEF collection, Andrade and Silva [1] also found that some queries are more influenced by geographic relevance than others, and that the optimal balance between thematic and geographic ranking is query-dependent.

We propose a relevance ranking method that generates ranked lists of documents based on their combined thematic and geographic relevance scores. Unlike previous work where fixed functions are used in combining the thematic and geographic relevance measures, our method uses a functional form that is dynamically determined by analyzing users' search context on a given query. The key idea is to measure the relative importance of thematic and geographic relevance through analyzing query specificity. We define query specificity as a measure of how specific (or general) a query is. Those terms of a query contribute to either thematic specificity or geographic specificity. Thematic specificity is a measure which attempts to capture how specific (or general) the thematic terms in a query are. Geographic specificity describes how specific (or general) the geographic scope in a query is. Our assumption is that users tend to use more specific thematic terms if they prefer the results of thematic relevance, while they tend to use more specific geographic scopes if they are more confident with geographic relevance scores. Consider the two queries "*Hilton hotels in central Pennsylvania*" and "*hotels along the Atherton Street in State College, PA*". Comparing with the latter, the former query is using more specific thematic terms and broader geographic scopes, which reflects that the user is concerned more about the thematic relevance than the geographic relevance, therefore the thematic relevance should be weighted more and the geographic relevance should be weighted less. Therefore, in our ranking method, we propose the algorithms to measure the thematic and geographic specificities of user queries and use them as indicators to judge the relative weights of thematic and geographic relevance.

Additionally, we explore the possibility of using Dempster-Shafer's theory [10] to combine the two different sources of ranking evidence based on the measurements of query specificity. This theory has been used to combine the relevance obtained by content and link analysis on the Web [9], or the visual and textual features in the image retrieval systems [2]. However, it has not been used in GIR systems. By considering thematic and geographic relevance measures as two independent sources of evidence for document ranking and thematic and geographic specificity as certainty measurement for each source of evidence, we attempt to adopt this theory to calculate combined relevance scores and compare it with the method of weighted sum.

This paper is organized as follows. In Section 2, we discuss how to measure the query specificities and propose our dynamic combination scheme based on Dempster-Shafer's theory of evidence. Section 3 reports experimental setup and results. Section 4 concludes the paper and addresses future research directions.

2. METHOD

Our ranking method is based on the hybrid geo-thematic IR model – GeoVSM [4]. In this model, each document in the collection ($D=\{d_1,d_2,\dots,d_n\}$) will be indexed both by a footprint (in geographical coordinate space) and by a term vector (in vector space). The footprint only represents the geographical scope of the document (GS_{d_i}), and the term vector only represents the thematic scope of the document (TS_{d_i}). Similarly, a query q also has a geographic scope GS_q and a thematic scope TS_q . The document ranking in this model can be performed in three steps:

1. find a function $\mathrm{Rel}_T(TS_q, TS_{d_i})$ to return the thematic similarity score between q and d_i ;
2. find a function $\mathrm{Rel}_G(GS_q, GS_{d_i})$ to return the score of geographic relevance between q and d_i ;
3. find a rank $(R,\le)$, $R \subseteq D$, so that
   $d_i \le d_j \Leftrightarrow f(\mathrm{Rel}_T(q,d_i),\mathrm{Rel}_G(q,d_i)) \le f(\mathrm{Rel}_T(q,d_j),\mathrm{Rel}_G(q,d_j))$
   where $d_i, d_j \in R$. f is a function combining both $\mathrm{Rel}_T(q,d_i)$ and $\mathrm{Rel}_G(q,d_i)$.

In this paper, we mainly focus on the third step. Given the thematic and geographic relevance between documents and queries, we investigate on how the query specificity can be calculated and incorporated to combine the two types of relevance measures effectively.

2.1 Computation of Query Specificity

2.1.1 Computing Thematic Specificity

Since the meaning of a query can be predicted from the meaning of its composing terms, we can measure the thematic specificity of a query based on the specificity of each term it includes. Assuming that each term is independent of the others, the contribution of each term's specificity is added to the query's specificity.

Let $Q = \{t_1, t_2, \ldots, t_m\}$ be a query, and t_k be its *k*-th composing term, the thematic specificity, $Spec_T$, of query Q is given by:

$$Spec_T = \sum_{t \in Q} \omega_t Spec_t \quad (2)$$

where ω_t is the weight for each term.

After having defined the thematic specificity of a query, we need to identify the method to measure the specificity of each term. Here, we will combine collection-based computation and ontology-based computation [12] together to calculate the term specificity.

Collection-based computation measures the term specificity based on the density of relevant documents in the whole collection. The fewer documents a term is related to, the higher possibility it has to include specific information. The formula we will use is IDF (Inverse Document Frequency) [6]:

$$Spec_t = -\log\left(\frac{N_t + 1}{N}\right) \quad (3)$$

where N_t is the number of documents containing term t, and N is the total number of documents in the collection.

Collection-based computation is dependent on the overall statistical properties of the document collection; therefore, it is not sufficient to quantify the term specificity if the document collection has a small population or a biased distribution. To address this problem, we complement it with ontology-based computation, which employs conceptual information in ontology structures to determine term specificity. The aim of ontology-based method is, given a query term t and an ontology structure (e.g. WordNet) as inputs, to calculate the term specificity for t by interpreting the conceptual information related to t, such as the number of senses, synonyms, and children or the depth in the structure. The computation is based on the notion that the less senses, synonyms and children a term has and the deeper it appears in the ontology structure, then the more specific the term is. Specifically, we adopt Zakos and Verma's computation framework [12] to calculate the ontology-based term specificity. In their framework, they hold the different types of conceptual information for a term in a conceptual term matrix (CTM). Given a term t, they first extract conceptual information representatives (number of senses, number of synonyms, level number, and number of children) of t from WordNet and store them in the CTM, then they weight the values based on the importance of different information types. Finally, they combine the weighted values to give a final score $CTM(t)$ of the specificity for term t.

After calculating these two kinds of specificity values for each term, the final measure of term specificity is the product of these two values. Then the thematic specificity for query Q can be calculated by:

$$Spec_T = -\sum_{t \in Q} \omega_t * CTM(t) * \log\left(\frac{N_t + 1}{N}\right) \quad (4)$$

2.1.2 Computing Geographic Specificity

Since we have associated each geo-referenced query with a geographic scope in the two-dimension geographic space, we can define the geographic specificity as the ratio of the area of the geographic scope to the whole coverage of the document collection.

Let Q be a geo-referenced query, and G_Q be the geometry representative of the associated geographic scope of Q, then the geographic specificity, $Spec_G$, of query Q is given by:

$$Spec_G = -\log\left(\frac{Area(G_Q)}{Area(G_D)}\right) \quad (5)$$

where $Area(G_Q)$ is the area of the geographic scope of Q, and $Area(G_D)$ is the total area of the geographic coverage of all documents in the collection.

2.2 Combination of Relevance

Given the measurement of thematic and geographic specificities, our next task is to incorporate them in the combination of two types of relevance scores.

A straightforward way is considering them as the weights of relevance scores respectively and using the linear weighted sum to compute the final relevance scores (see Equation 1). Since the values of query specificity range from $(0, +\infty)$, we need to normalize them into $(0,1)$ as follows:

$$\omega_T = 1 - \left(\frac{1}{\ln(e + Spec_T)}\right), \quad \omega_G = 1 - \left(\frac{1}{\ln(e + Spec_G)}\right) \quad (6)$$

In addition, we attempt to employ the Dempster-Shafer theory [10] to combine the two sources of relevance. Dempster-Shafer theory is a mathematical theory of evidence which is used to combine separate pieces of information evidence to calculate the probability of an event. One of the advantages of the Dempster-Shafer framework is that it allows one to specify a degree of uncertainty $m(\Theta)$ for each source of evidence instead of being forced to provide prior probabilities. If in a body of evidence $m(\Theta) = 0$, it means we have very high confidence in this body of evidence and no uncertainty. In our ranking context, we can consider thematic and geographic relevance as two independent pieces of evidence for the document ranking and our measures of query specificity can be used directly to calculate the degrees of uncertainty for thematic and geographic ranking evidence.

According to this theory, we consider the set of documents in the collection as the frame of discernment $\Theta = \{d_1, d_2, \ldots, d_n\}$. The thematic and geographic relevance are considered as two bodies of evidence that will be combined into a single body of evidence in the frame of discernment Θ. When two bodies of evidence are defined in the same frame of discernment, we can combine them using Dempster's combination rule, under the condition that the two bodies are independent of each other. The original equation to combine the two pieces of evidence is generally computationally expensive. Following Jose [7], we can reduce the computational complexity to a particular case where we have positive evidence for singleton hypotheses only. In other words, we have positive belief for $\{d_1\}, \{d_2\}, \ldots, \{d_n\}$ and Θ only. Let m_T, m_G be the probability mass functions of thematic and geographic ranking evidence respectively, the combination rule can be simplified as:

$$m(\{d_i\}) \propto m_T(\{d_i\}) * m_G(\{d_i\}) + m_T(\Theta) * m_G(\{d_i\}) + m_T(\{d_i\}) * m_G(\Theta) \quad (7)$$

where $m_T(\{d_i\})$ and $m_G(\{d_i\})$ are the thematic and geographic relevance scores for document d_i.

Besides, we need to assign to each source of evidence a measure of uncertainty. Based on our notion that thematic/geographic specificity can reflect the user's confidences on thematic /geographic relevance, we can consider them as indicators of certainty for each source of evidence. As a result, $m_T(\Theta)$ and $m_G(\Theta)$ can be calculated by:

$$m_T(\Theta) = \frac{1}{\ln(e + Spec_T)}, \quad m_G(\Theta) = \frac{1}{\ln(e + Spec_G)} \quad (8)$$

In sum, given the relevance scores $m_T(d_i)$ and $m_G(d_i)$ from both thematic and geographic evidence for each document d_i and the query specificity values $Spec_T$ and $Spec_G$ computed as described in Section 4.1, the final combined relevance score for each document can be computed using Equation 7 and 8.

3. EVALUATION

3.1 Experimental Setup

To evaluate our proposed scheme for query-aware relevance combination in document ranking, we run experiments on a web-based data set, which consists of thousands of documents, which are either news stories or blog entities extracted from different websites. The documents were processed with geographic names extracted. Only documents containing at least one geographic name were kept in the evaluation collection, which totally has 3488 distinct geographic names, 62% of which are geographic names in US.

For each document, we indexed it using both thematic keywords and geographic scopes. First, all terms were filtered by the SMART 571 stop word list and stemmed using the Porter stemming algorithm. A standard inverted file was used as the indexing structure. On the other hand, the geographic names contained in each document were parsed, geo-coded and merged to calculate the geographic scopes. The geographic scopes have two different geometric types: polygons (continents, countries, and other administrative areas) and points (cities and towns). The linear objects, such as roads, rivers were not parsed in this experiment. Each document may have multiple geographic scopes, which were stored and indexed in a PostgreSQL database with PostGIS extension. Documents in these two index structures can be matched based on their unique document IDs.

When a query is processed, it is first parsed into two parts: the thematic terms and the geographic scope. The former is used to retrieve documents with thematic relevance scores, and the latter is used to retrieve documents with geographic relevance scores. We use Lucene, a well-known open source keyword-matching IR system, to calculate the thematic relevance scores based on the standard TF-IDF (Term Frequency-Inverse Document Frequency) model. The geographic relevance scores are calculated with both topological relationship and metric proximity concerned. When the geographic scope of a query is a polygon, we will first test whether the geographic scope of a document is inside the query's scope. If so, the geographic relevance score will be 1, or else the Euclidean distance between these two scopes will be calculated and normalized to compute the geographic relevance score. If the geographic scope of a query is a point, the scores will be computed merely based on distances. This is quite similar to the functions provided by Andrade and Silva in [1].

After the two types of relevance scores are computed, the thematic and geographic specificities for the query are calculated using the equations we proposed in Section 2. Then the documents with two relevance scores can be merged following our combination scheme and the final ranked list of documents will be returned.

3.2 Experiment Design

To evaluate whether our dynamic ranking scheme can provide better results, we implemented three different ranking methods for each query in the experiment. First, the thematic and geographic relevance scores are combined using the static weighted sum (Equation 1). The weights are manually tuned to $\omega_T = 0.47, \omega_G = 0.53$ with the best average performance in the document collection and fixed for all the queries in the experiment (R1). Second, we still used the weighted sum to calculate the final ranking scores, but the weights are dynamically decided by using thematic and geographic specificities for each query (Equation 4 & 5) (R2). Last, we use the Dempster-Shafer theory (Equation 7) to calculate the final ranking scores, which is also on a per-query basis (R3).

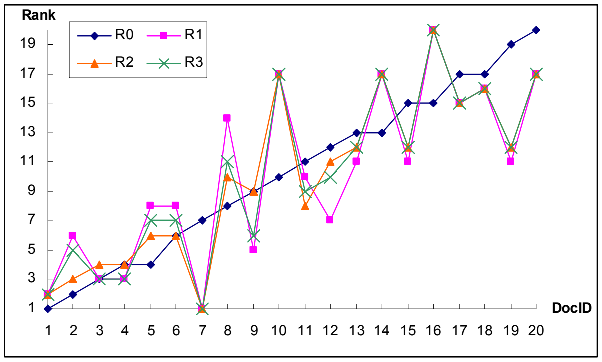
To investigate whether our dynamic ranking methods can provide better performance, we compared the results of these three methods with human judgment. We recruited two volunteers from our research group to participate in the evaluation. Based on the contents of the document collection, we formed 12 queries with different types of distribution of thematic and geographic specificities. Because we assume these queries be proposed by the users with their own search intentions, we asked the evaluators to preview the sample queries and modify them with their familiar topics and locations (Some of the sample queries are listed in Table 1.). Then we selected 20 candidate documents for each query (some of the queries are associated with less than 20 documents if they are too specific), and distributed them to the evaluators in random orders. The evaluators were asked to rank the documents according to their own experience. The ranking results produced by the evaluators were cross-checked and merged to form the baseline for comparison.

Table 1. Query examples in the evaluation

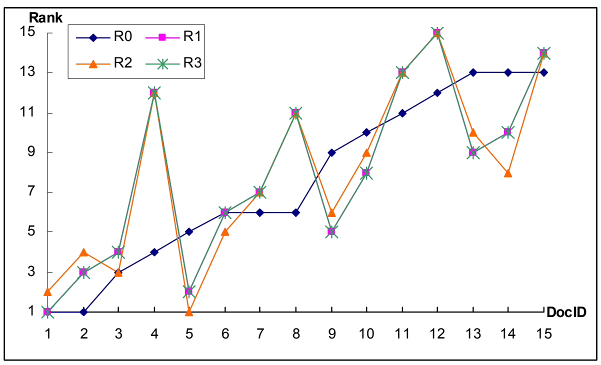
Type	Query	Thematic Specificity	Geographic Specificity
1	animals in Pennsylvania	general	general
2	dogs in Pennsylvania	specific	general
3	animals near Pittsburgh, PA	general	specific
4	dogs near Pittsburgh, PA	specific	specific

3.3 Results and Analysis

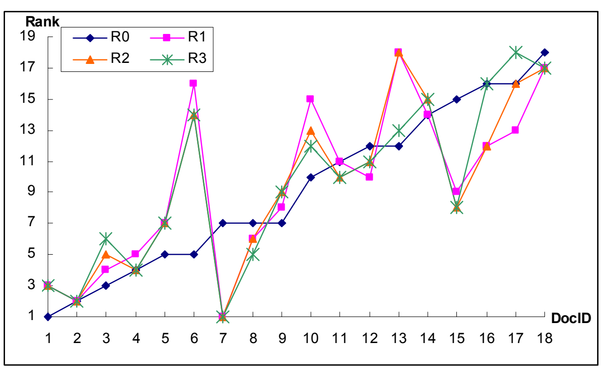
Figure 1 shows some ranking results of our experiments. R0 stands for the results of evaluators' judgment, R1, R2, and R3 stand for the ranking results from our three experimental methods respectively. The first graph (A) shows the results from a query in type 1 (lower thematic and lower geographic specificities). The second graph (B) shows the results from a query from type 2 (higher thematic but lower geographic specificities). The other two graphs (C, D) show the results of queries in the other two types. For each graph, the x-axis is the document numbers in the ascending order of human ranks, and the y-axis is the ranks for each document using different methods respectively.



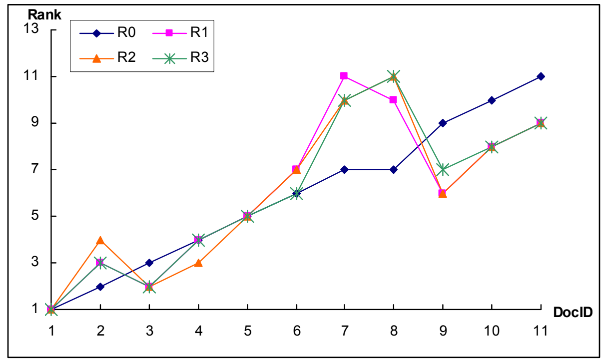
A. animals in Pennsylvania



B. dogs in Pennsylvania



C. animals near Pittsburgh, PA



D. dogs near Pittsburgh, PA

Figure 1. Ranking results for four types of sample queries.

To compare the ranking results, we calculated the Spearman rank correlation coefficient ρ [11] between ranking results of each method and the human results (Table 2). Spearman's ρ is one of the most common methods to compare two rankings on the same set of items in statistics. If the agreement between two rankings is perfect, the value is 1. If the one ranking is completely diverse of the other, the value is -1. The increasing values between them imply increasing agreement between the rankings. Table 2 shows the average Spearman's ρ values for each method in four types of queries.

Table 2. Average Spearman's ρ values

Type	R1, R0	R2, R0	R3, R0
1	0.558	0.698	0.665
2	0.607	0.534	0.607
3	0.472	0.548	0.616
4	0.680	0.650	0.726

To test whether the coefficient values in Table 2 are significantly different from each other, we compared them for various levels of significance (Table 3).

Table 3. Levels of Significance

Type	R1, R0	R2, R0	R3, R0
1	1%	0.5%	0.5%
2	2.5%	5%	2.5%
3	5%	2.5%	0.5%
4	2.5%	2.5%	1%

From Table 3, we can see that, for queries in Type 1 and 3, dynamic combination schemes (R2, R3) show higher levels of significance, which means that the agreements between their results and human judgment are more significant than the static one (R1). For queries in Type 4, R2 has lower ρ value than R1, but they are in the same level of significance, which means the difference is not critical. R3 continues to outperform the other two schemes. In Type 2, the result of R1 is significantly better than that of R2 and comparable with that of R3. This is probably because the manually tuned weights for R1 best fit the queries in Type 2. However, since the weights are fixed, it is unable to provide the same level of performance for the other types of queries. In the average, we can conclude that the dynamic combination schemes (R2, R3) provide more flexibility and have obvious performance improvement over the static one (R1).

Comparing the results of dynamic weighted sum (R2) with the results of Dempster-Shafer method (R3), we can see that R3 outperformed R2 in most situations. The results of R3 show higher levels of significance than those of R2 except the queries in Type 1. Overall, the improvement of R3 is significant comparing with R2 based on our experimental data. It reflects that Dempster-Shafer method can be considered as an effective way to combine different sources of relevance in GIR ranking.

Another interesting finding in Figure 1 is that, for almost all the documents, the system ranks in R1, R2, and R3, are either all above the human ranks, or all below the human ranks, i.e. although the different combination schemes can have quite different levels of agreement with human judgment, their ranking results cannot be extremely different from each other. This is because that the thematic and geographic relevance scores themselves are still the major factors to determine the final ranking results. Dynamic combination parameters and schemes can be used to adjust and improve the ranking results. However, the maximum optimization can only be achieved with the effective computation of thematic and geographic relevance scores.

4. CONCLUSIONS

In this paper, we have presented a query-aware ranking method to dynamically combine thematic and geographic relevance in GIR. We believe that features of user queries, which may reflect the users' underlying search intention, can be used to calculate the relative weights of different sources of ranking evidence on the fly. The feature that we chose to investigate is query specificity. The preliminary evaluation shows that the different distribution of thematic and geographic specificities in user queries can affect the performance of ranking methods and our method to distinguish different types of geo-referenced queries based on their query specificity is promising to address the issue of multiple relevance combination for document ranking in GIR.

In practice, we propose two dynamic schemes to combine the thematic and geographic relevance scores: the dynamic weighted sum (R2) and the Dempster-Shafer's method of evidence combination (R3). Our experimental result shows that Dempster-Shafer's method is superior to the weighted sum method in most situations, which has higher level of agreement with human judgment.

Currently, the experiment sample is quite small and the validity of human judgment in the evaluation is questionable. In the future work, we will supplement and test our ranking method in other GIR test environments (e.g. GeoCLEF) and compare its effectiveness with other system implementations.

Also, we will investigate other features of user queries which may help GIR systems better understand the users' information search contexts and provide better retrieval performance. One potential is to understand a series of user queries together in a specified search context, i.e. queries can be grouped corresponding to different searching tasks. The idea that problem solving is the underlying motivation for information search can further improve the ranking performance in GIR.

5. ACKNOWLEDGEMENTS

This work is partially supported by the National Science Foundation under Grants No EIA-0306845. We thank Haibin Liu and Kun Chen who helped on the IR experiment presented here.

6. REFERENCES

[1] Andrade, L and Silva, M. J. Relevance Ranking for Geographic IR. *Proceedings of the workshop on Geographic Information Retrieval*, *SIGIR 06,* Seattle, USA, 2006.

[2] Aslandogan, Y. A. and Yu, C. T. Diogenes: A Web Search Agent for Content Based Indexing of Personal Images. *Proceedings of the ACM SIGIR 00*, Athens, Greece, 2000.

[3] Cai, G. GeoVIBE: A Visual Interface to Geographic Digital Library. *Proceedings of the 1st Visual Interfaces to Digital Libraries Workshop*, Roanoke, VA, USA, 2001.

[4] Cai, G. GeoVSM: An Integrated Retrieval Model For Geographical Information. *Lecture Notes on Computer Science 2478: Geographical Information Science: Second International Conference on GIScience*, Baltimore, MD, USA, 2002, 65-79.

[5] Hobona, G., James, P. and Fairbairn, D. Multidimensional Visualization of Degrees of Relevance of Geographic Data. *International journal of geographic information science*, 20(5), 2006, 469-490.

[6] Jones, K. S. A Statistical Interpretation of Term Specificity and its Application to Retrieval. *Journal of Documentation*, 28(1), 1972, 111-121.

[7] Jose, J. M. An Integrated Approach for Multimedia Information Retrieval. *Ph.D. Thesis*, Robert Gordon University, 1998.

[8] Martins, B., Silva, M. J., and Andrade, L. Indexing and Ranking in Geo-IR Systems. *Proceedings of the workshop on Geographic Information Retrieval, CIKM 05*, Bremen, Germany, 2005.

[9] Plachouras, V. Dempster-Shafer Theory for a Query-Biased Combination of Evidence on the Web. *Information Retrieval*, 8, 2005, 197-218.

[10] Sentz, K. and Ferson, S. Combination of Evidence in Dempster-Shafer Theory, *Technical Report SAND2002-0835*, Sandia National Laboratories, Albuquerque, NM, 2002.

[11] Spearman's Rank Correlation Coefficient. Online at: http://en.wikipedia.org/wiki/Spearman's_rank_correlation_coefficient.

[12] Zakos, J. and Verma, B. Concept-Based Term Weighting for Web Information Retrieval. *Proceedings of the 6th International Conference on Computational Intelligence and Multimedia Applications (ICCIMA05')*, Las Vegas, USA, 2005, 173-178.