

# Map Search via a Factor Graph Model

Qi Zhang, Jihua Kang, Yeyun Gong, Huan Chen, Yaqian Zhou, Xuanjing Huang  
School of Computer Science, Fudan University  
825 Zhangheng Road, Shanghai, P.R.China  
{qz, 12201240059, 12110240006, 12210240054, zhouyaqian,  
xjhuang}@fudan.edu.cn

## ABSTRACT

Map search has received considerable attention in recent years. With map search, users can specify target locations with textual queries. However, these queries do not always include well-formed addresses or place names. They may contain transpositions, misspellings, fragments and so on. Queries may significantly differ from items stored in the spatial database. In this paper, we propose to connect this task to the semi-structured retrieval problem. A novel factor graph-based semi-structured retrieval framework is introduced to incorporate concept weighting, attribute selection, and word-based similarity metrics together. We randomly sampled a number of queries from logs of a commercial map search engine and manually labeled their categories and relevant results for analysis and evaluation. The results of several experimental comparisons demonstrate that our method outperforms both state-of-the-art semi-structured retrieval methods and some commercial systems in retrieving freeform location queries.

## Categories and Subject Descriptors

H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval - Information Search and Retrieval

## Keywords

Map Search, Query Representation, Factor Graph Model

## 1. INTRODUCTION

In recent years, along with the quick expansion of mobile Internet and the use of smartphones, map search has become one of the most popular applications. With map search, users can retrieve geographic information from spatial databases using textual queries. Queries can be formal postal addresses or informal textual descriptions of category, name, location, or geographic coordinates. To satisfy increasing requirements, both research and industry

communities have paid attention to this task [13, 17, 22, 32, 36, 39].

With the purpose of understanding the needs of users, we analyze query logs of a commercial map search engine and find that many queries contain transpositions, misspellings, conflicting information, and similar variations. Consider the following query examples which may be used to retrieve the restaurant named “Alexander’s Steakhouse”, located at “410330 North Wolfe Road, Cupertino, CA 95014”:

- “CA 95014, 410330 North Wolfe Road, Cupertino” - transposition of address sequence.
- “alexander steakhouse 410330 north wolfe road” - combination of restaurant name and location.
- “north wolfe alexander steakhouse cupertino” - inexact location with transposition of city name and restaurant name
- “alexandar steakhouse in cupertino ca” - misspellings with extra words(“in”)
- “alexander steakhouse north wolfe sunnyvale” - city name mistake

From the above examples, we can observe that freeform textual queries may significantly differ from the items stored in the spatial database. Moreover, different countries and territories may have different address formats. Textual attributes in some spatial databases are even more complex. Attributes may contain unstructured descriptions or multiple elements that cannot be easily structured. For example, “165 Lujiazui ring road, Shangri-La Hotel across the way” is a valid postal address in China. Hence, the primary target of this work is to handle the matching problem between freeform textual queries and items in semi-structured spatial databases.

Previous work devoted to this task can be roughly divided into two directions: spatial keyword search and geo intention discovering. Studies belonging to the spatial keyword search category consider both distance and textual relevance for location-based services [39, 14, 38, 26, 17]. They have proposed numerous specialized index structures, which combine inverted indexes with space-filling curves [12] and tree-based methods [14, 25, 39], to deal with the efficiency problem. Christoforaki et al. [13] introduced coarse-grained spatial structures to the inverted indexes. However, queries with misspellings, conflicting information, transpositions, and other issues are much less discussed in these studies.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).  
*CIKM'13*, Oct. 27–Nov. 1, 2013, San Francisco, CA, USA.  
Copyright 2013 ACM 978-1-4503-2263-8/13/10 ...\$15.00.  
<http://dx.doi.org/10.1145/2505515.2505674>.

On the other hand, a number of researchers use unsupervised, semi-supervised, and supervised methods to parse a query into a semantic interpretation [31, 32, 36, 15]. Some studies focus on the location disambiguation task and used rule-based [3], ontological-based [32], and learning-based [1] methods to resolve it. The CLEF 2007 Cross-Language Geographic Information Retrieval Track [28] defined semantic components of queries. Since automatically analyzing semantic interpretation of a query is still challenging, if directly creating intermediate interpretation of queries for map searches, the performance of retrieval methods may suffer from the low performance.

In this work, we propose to connect the map search task to the semi-structured retrieval problem and directly model the relationships between query terms and locations. The proposed method is based on factor graphs. The query concept, which models dependency between arbitrary query terms, is treated as the basic unit in this model. To handle variations of queries and verbal terms, concept weighting is processed by factor functions. We also incorporate the mapping relationship of attributes and concepts into factor functions to handle attribute selection. Experimental evaluations demonstrate that by integrating these characteristics, the proposed method can significantly improve the effectiveness of map searches.

Overall, the contributions of this paper can be summarized as fourfold. First, we propose to model the map search task as a semi-structured retrieval task. Second, a novel factor graph-based semi-structured model is introduced to deal with textual issues in map searches by combining concept weighting and attribute selection in a unified framework. Third, a word-based similarity metric is proposed to match names between the query and items in a spatial database. Fourth, a detailed analysis of query logs is provided to show the specific characteristics of map search queries.

The remainder of the paper is organized as follows: In section 2, we review a number of related works and the state-of-the-art approaches in related areas. Section 3 presents the proposed method. Experimental results in test collections and analyses are shown in section 4. Section 5 concludes this paper.

## 2. RELATED WORK

The proposed approach builds on contributions from several research communities: (1) spatio-textual index structure, (2) geo intention discovering, and (3) semi-structured information retrieval. In the following of this section, we give brief description of the previous works on these areas.

### 2.1 Spatial Keyword Search

Spatial keyword search approaches are based on input location and keywords to find all relevant POIs. There are many previous studies focusing on this task [39, 16, 14, 38, 26]. Zhou et al. [39] described a hybrid index structure and three different combining schemes to handle both textual and location aware queries. The index structure integrated inverted files and R\*-trees. Felipe et al. [16] defined the top- $k$  spatial keyword search problem and proposed IR<sup>2</sup>-Tree, which also combined R-Trees and signature files to answer spatial keyword queries. Göbel et al. [19] also proposed an R-tree based approach, of which index structure was optimised for a single criterion adding special treatment for

the other criterion at the leaf nodes. These methods are mainly focusing on index structures for efficiency.

Cao et al. [9] introduced another problem of retrieving a group of spatial web objects and proposed approximate solutions based on IR-Tree. Fan et al. [17] studied spatio-textual similarity search problem on *regions-of-interest*(ROIs). They proposed grid-based signatures and efficient hybrid filtering algorithms. Roy and Chakrabarti [4] introduced the problem of location-aware type-ahead search on spatial databases. They proposed an integrated Trie tree with a spatial data structure for this task.

We can observe that most of previous spatial keyword search methods mainly focused on index structures to process spatial overlap. However, as can be seen from the example given in the introduction section, textual issues are also challenging in practice. Christoforaki et al. [13] also reported that text-first could beat both R\*-tree methods and some index structures. Hence, in this work, we restrict our attention to complex textual issues.

### 2.2 Geo Intention Discovering

In order to satisfy the information needs of users, various approaches have been proposed to discover geographic intention from user queries. The cross-language evaluation forum even proposed a cross-language geographic information retrieval track in 2007 (GeoCLEF) [28]. A number of unsupervised, semi-supervised, and supervised methods have been proposed to parse queries to semantic interpretations [31, 32, 22, 36, 15].

Datha et al. [15] proposed a hybrid algorithm to extract entity names and attributes from queries. Spatial processing was used to determine viable partial interpretations. Joshi et al. [22] also introduced spatial constraints to resolve the ambiguities problem.

Sengar et al. [32] proposed to find a direct mapping from subsequences in the text query to specific entities in a spatial database. Given a textual query, a list of multiple approximate matches between query subsequences and entity attributes are calculated based on pre-computed indexes. Then, their algorithm tried to generate a set of candidate interpretations. Finally, a ranked list of interpretations is returned as final results.

Yi et al. [36] proposed city language model and methods using the model on three tasks: 1) identify users' implicit geo intent; 2) determine whether the geo-intent is localized around the current geographic location of user; 3) predict cities from queries.

The goal of these investigations are related with us. However, in this paper, the proposed factor graph based method directly models the relation between query and entries in a spatial database without any intermediate steps. Although, geographic intention is not explicitly dealt with under this framework, one can easily determine it through final result.

### 2.3 Semi-Structured Information Retrieval

The proposed method is also related to the research on structured and semi-structured information retrieval. The goal of this task is to allow users to enter freeform queries to retrieve complex semi-structured XML data, structured database, and RDF resources. The task has been addressed from both IR and database perspectives by various approaches[30, 27, 23, 11, 2, 18, 8].

Table 1: Descriptions of query categories, examples, and percentages.

Category	Description	Example	Percentage
POI	The intentions of the queries are to find a particular POI.	“410330 North Wolfe Road”	86.2%
List	The intentions of the queries are to find POIs belonging to a type or a chain store.	“Starbucks, Palo Alto”	13.8%
Name	Queries contain only the name of the target location.	“Philz Coffee”	56.5%
Address	Queries contain only the address of the target location.	“748 Van Ness Ave San Francisco”	18.8%
Complex	Queries contain more than one attributes (place name, address, category, and so on).	“Philz Coffee 748 Van Ness Ave”	24.7%
Verbose	Queries contain verbal terms, which are used as grammatical components for communication between humans.	“ <i>Help me locate Philz Coffee in Van Ness Ave</i> ”	29.5%
Typo	Queries contain misspelled street, place name, category, or terms.	“ <i>Phils</i> ” instead of Philz	16.8%
Transposition	Some fields of addresses or names are transposed comparing with the standard formats.	“95014 CA, Van Ness Ave 748”	5.2%
Fragment	Address or place names are not complete.	“Philz Coffee, Van Ness Ave” 748 is omitted	62.9%
Conflict	Queries contain conflict information provided by different attributes.	“Philz Coffee 748 <i>North Wolfe Ave</i> ” instead of “Van Ness Ave”	3.5%

Ogilvie and Callan [30] adapted a language model for XML component retrieval. The method assigned different weights for each element based on the length of text content or the importance for retrieval. Lu et al. [27] extended document level field-weighted retrieval function BM25F to element level retrieval function BM25E. Each element was treated as a document. Kim et al. [23] proposed a probabilistic method which infers the mapping between each query term and XML element based on collection statistics. Probabilistic semi-structured retrieval model use the mapping probability as weights to combine each element into a document score.

Chen et al. [11] gave an overview of this task in the tutorial of SIGMOD 2009. Agrawal et al. [2] studied the task of retrieving relevant information from structured databases for web search queries. They proposed search engine-integrated approach, which used web documents to enhance the performance. Ganti et al. [18] studied the problem of keyword search over entity databases, and proposed an approach to a keyword query to a structured query. Bicer et al. [8] focused on keyword search result ranking and proposed a relevance-based language models for ranking aggregated structured results.

In this work, we connect the map search task to semi-structured information retrieving task and proposed a novel factor-graph based approach to model the similarity between query and entries of semi-structured spatial database.

### 3. PRELIMINARIES

#### 3.1 Problem Formulation

Given a query  $q$ , which contains  $n$  terms  $\{t_1, t_2, \dots, t_n\}$ , the aim of map search is to find a list of related *POIs* (point of interests) from a spatial database. A spatial database consists of a huge number of *POIs* which have *geometry* and

*textual* attributes. Geometry attributes include longitude, latitude, altitude, target shape description, and so on. Textual attributes contain street address, city, state, name, category, landmark, ZIP code, et al. We use  $p_i$  to represent the  $i$ th POI in the spatial database.  $\{att_{i1}, att_{i2}, \dots, att_{im}\}$  represents attributes set of POI  $p_i$ .

In this work, we make no assumptions about the form and structure of inputs. It means that users can specify a target location using either well-formed or ill-formed address, place name, category and any other textual contents as user wishes. Hence, efficiently matching subsequences of queries with attributes of POIs is a challenging problem. For example, “San Francisco 748 Van Ness Ave” would be a legal query, and the POI named “Philz Coffee”, which is located in this address, should be retrieved as the corresponding result. All the queries described in the introduction section are also can be used to retrieve “Alexander’s Steakhouse”.

#### 3.2 Query Analysis

To better understand the intention of users, analyzing search query logs is one of the commonest way. Most of the previous analyses on map search focus on discovering the characteristics of distance, search session, type of target place, and so on [21, 35]. In this work, we examine query logs from the perspective of retrieving. We collected query logs from Nov. 1st, 2012 to Nov. 10th, 2012 with the help of a commercial map search engine in China. For analysis, we sampled 10,000 queries from query logs and manually labeled their categories.

In this paper, we work on the following questions: 1) Whether the intention of user is to find a particular location or a set of locations? 2) What kinds of information are provided through queries? 3) Whether queries are exactly the same as items in spatial database?. To answer these questions, we propose to classify queries into 10 categories, which can be classified into three groups. The description

of the query categories, examples and percentage of queries for each category are shown in Table 1.

From the statistic, we can observe that over 86% queries belonging to POI category. It means that users often already have a clear target place when they enter the query. For the distribution of attributes in map search queries, we observe that 24.7% queries contain more than one attribute. For example, the query “alexander steakhouse north wolfe cupertino” contains sub-queries for three attributes: POI name, road, and city name. Among all List queries, about 81.8% of them are name query and the other 18.2% queries contains not only a name of POI but also other information. However, among the POI queries, the distributions of Name, Address, and Complex categories are very close to the distributions of all queries, which are 52.4%, 19.8%, and 27.8% respectively.

We also study the categories of queries variants. From the percentages of queries belonging to “Verbose”, “Typo”, “Fragment”, and “Conflict” categories, we can observe that most of the queries are not strictly equal to the items in spatial database. About 62.9% queries do not contain complete place names or addresses. Verbose queries also occur frequently, which are around 29.5% among all queries. Since only 3.5% queries contain conflict information in total, most of information provided by users can be directly used as constraints for retrieval results. Around 16.8% queries contain typos. We think that the longer length of map search queries is one of the main reason.

## 4. OUR APPROACH

From the definition and analysis of the task, we can observe that the problem of matching subsequences of queries with attributes of POIs can be converted into semi-structured search task. To handle fragment, misspelling, verbose, and other issues, we propose a novel factor graph based semi-structured search model, which combines concept weighting and attribute selection together. According to the definition given by Kschischang et al. [24], a factor graph is a bipartite graph, which has a *variable node* for each variable  $x_i$ , a *factor node* for each local function  $\phi_j$ , and an edge-connecting variable node  $x_i$  to factor node  $\phi_j$  if and only if  $x_i$  is an argument of  $\phi_j$ .

Figure 1 shows the graph structure used in this work. Inspired by hypergraph model, which was proposed by Bendersky and Croft [5] for processing higher-order term dependencies, in this work, we also treat concept as the basic unit for retrieving. Vertex  $c_i$  represents the variable of concept  $i$ . It models dependency between a subset of query terms. In order to show the relationships between concepts and query terms, we use dashed lines to connect them. Query terms are not directly incorporated into calculation. Vertex  $q$  and  $POI$  represent the query and POI respectively. Vertex  $att_j$  represents the  $j$ -th attribute of the POI. Edges in the graph are based on two kinds of factors  $\phi_e(c, att)$  and  $\phi_e(q, POI)$ . In following sections, we introduce details of concepts, edges, and factors used in this work.

### 4.1 Concepts

Under this framework, concept is treated as the basic unit for retrieving. Structures group concepts belonging to the same type together. In previous works, several structures have been proved to be useful for document or passage level

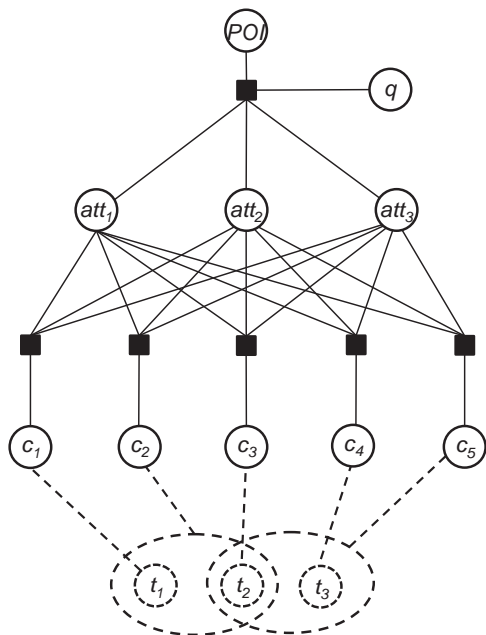


Figure 1: The factor graph used in this work for textual based location search task.

retrieval task. In this work, we propose to use following structures for the location search task:

**UNI-Structure** The unigram (UNI) structure treats the individual term  $t_i$  of query as concepts. It has been successfully used in bag-of-words models. Since Chinese is one of the languages which have a nontrivial word segmentation process, we used both character (QTC) and word (QTW) based term generation methods.

**BI-Structure** The bigram (BI) structure contains adjacent term pairs  $t_i t_{i+1}$  of query as concepts. Similar with UNI-Structure, character (QTC) and word (QTW) based term generation methods are also both used in here.

**NE-Structure** The named entity (NE) structure contains identified state, city name, street, organization and other types of data within query terms. Since Chinese word segmenter may have conflicts with named entity extractor, we only use NE-Structure with character based term generation method.

### 4.2 Factors

As shown in Figure 1, two kinds of factors are used in this work: attribute factor ( $\phi_e(c, att)$ ) and POI factor ( $\phi_e(c, POI)$ ). Attribute factor assigns a score to the occurrences of concept  $c$  in all attributes of a POI, regardless of the other query concepts. Since the order of terms is important for location search, we incorporate POI factor to assign scores according to the number of common term sequence between query and attributes of POI.

To model the occurrences of a concept  $c$  in a text fragment  $X$ , we follow previous works on log-linear retrieval models [5, 6]. We estimate the occurrence matching function  $f(c, X)$  using a log of the language modeling. In order to overcome

**Table 2: Features used for concept weighting and their descriptions**

Feature	Description
$GF(c)$	Frequency of concept $c$ in Google n-grams
$QF(c)$	Frequency of concept $c$ was used as a portion of a query (extracted from a large query log)
$NF(c)$	Frequency of concept $c$ in the POI name attribute of the collection
$CB(c)$	Concept $c$ is a city name (=1) or not (=0)
$SB(c)$	Concept $c$ is a state name (=1) or not (=0)
$TB(c)$	Concept $c$ is a category name (=1) or not (=0)

the zero probability problem, Dirichlet smoothing is also used in this work[37].

$$f(c, X) \triangleq \log \frac{tf(c, X) + \mu \frac{tf(c, C)}{|C|}}{\mu + |X|}, \quad (1)$$

where  $tf(c, X)$  function represents the number of occurrences of the concept  $c$  in a text fragment or a collection;  $\mu$  is the smoothing parameter;  $|X|$  and  $|C|$  represent the total number of terms in  $X$  and the collection respectively.

Based on occurrence function  $f(c, X)$ , attribute factors are defined as follows:

$$\phi_e(c, att) \triangleq \exp \left( \omega(c)P(att|c)f(c, att) \right), \quad (2)$$

where  $\omega(c)$  represents the importance weight of concept  $c$ ;  $P(att|c)$  represents the mapping probability between concept  $c$  and attribute  $att$  based on collection statistics;  $f(c, att)$  represents occurrence matching function.  $\omega(c)$  is assigned by several weighting features, of which detailed explanations are given below.

From analyzing the query log, we observe that queries may contain one or more attributes such as POI name, street name, city name, and so on. Since these entities may contain multiple terms, term sequences should be taken into consideration. In this work, we use POI factor ( $\phi_e(q, POI)$ ) to incorporate the number and order of common terms between query and attributes. Formally,

$$\phi_e(q, POI) \triangleq \exp \left( \lambda \sum_{i=1}^m Jaro_{word}(q, att_i) \right), \quad (3)$$

where  $q$  is input query, which can be represent by concepts belonging to *UNI-Structure*;  $Jaro_{word}$  represents a similar metric, which is extended from *Jaro* metric [20] and will be discussed in the following section;  $\lambda$  is the weighting parameter between two kinds of factor functions.

With the definitions of attribute, POI factor and factor graph ranking function in Equation 1, we can explicitly rewrite the relevant score  $Score(q, POI)$  between query and a POI as:

$$\begin{aligned} Score(q, POI) &= \prod_{e \in \Pi_e} \phi_e(q, POI) \stackrel{rank}{=} \sum_{e \in \Pi_e} \log \left( \phi_e(q, POI) \right) \\ &\triangleq \sum_{c \in \Pi_C} \sum_{i=1}^m \left( \omega(c)P(att_i|c)f(c, att_i) \right) \\ &\quad + \lambda \sum_{i=1}^m Jaro_{word}(q, att_i), \end{aligned} \quad (4)$$

where  $\Pi_C$  represents the entire concept set.

### 4.3 Concept Weighting $\omega(c)$

According to the analysis of query log, we can observe that about 29.5% queries contain verbal terms, which have grammatical meaning for communication between humans. Previous works have proposed several frameworks to address this problem [6]. In this work, we also use weighting features and linear weighted combination to model concept weight as follows

$$\omega(c) = \sum_{\varphi \in \Phi} \lambda_{\varphi} \varphi(c), \quad (5)$$

where  $\Phi$  represents weighting feature set;  $\varphi$  is weighting feature. Table 2 presents the features used for concept weighting. These features can be grouped into two sets: frequencies and occurrences. Moreover, both of them can be efficiently calculated for even large scale collections.

### 4.4 Attribute Mapping Probability $P(att_i|c)$

Since user queries may contain one or more information belonging to different attributes and do not have explicit structure information, we propose to follow PRM-S[23], which estimates the probability based on statistic from attributes. Using Bayes' theorem, we can estimate the posterior attribute mapping probability  $P(att_i|c)$  by combing the prior probability  $P(att_i)$  and the probability of a concept occurring in a given attribute type  $P(c|att_i)$ .

$$P(att_i|c) = \frac{P(c|att_i)P(att_i)}{P(w)} = \frac{P(c|att_i)P(att_i)}{\sum_{k=1}^m P(c|att_k)P(att_k)}, \quad (6)$$

where  $P(c|att_i)$  can be estimated based on bigram language model  $P_{att_i}(t_k|t_{k-1})$  for each attribute  $att_i$ ,  $P(att_i)$  represents the prior probability of attribute  $att_i$  being mapped into any query concept.

### 4.5 Similarity Metric $Jaro_{word}(q, att_i)$

*Jaro* metric [20] is used to measure similarities between two strings. It is based on the number and order of common characters. In this work, to emphasize the impact of the sequence of common terms between query and attributes, we extend it to word level *Jaro<sub>word</sub>*, which is based on the number and order of the common words between two strings. Given string  $s = a_1 \dots a_K$  and string  $t = b_1 \dots b_L$ , define a word  $a_i$  in  $s$  to be *common with*  $t$  there is  $b_j, sim(a_i, b_j) > \delta$ . Let  $s' = a'_1 \dots a'_K$ , be words in  $s$  which are common with  $t$  (in the same order they appear in  $s$ ) and let  $t' = b'_1 \dots b'_L$ , be analogous; Then define a *transposition* for  $s', t'$  to be a position  $i$  such that  $sim(a'_i, b'_i) < \delta$ . Let  $T_{s', t'}$  be one-half

the number of transpositions for  $s'$  and  $t'$ . The  $Jaro_{word}$  similarity metric for  $s$  and  $t$  is

$$Jaro_{word}(s, t) = \frac{1}{3} \cdot \left( \frac{|s'|}{|s|} + \frac{|t'|}{|t|} + \frac{|s' - T_{s',t'}|}{|s'|} \right), \quad (7)$$

where the similarity between words  $sim(a_i, b_j)$  can be different metrics, such as edit-distance metrics, token-based distance metrics, hybrid methods, or phonemic similarity metrics.

## 4.6 Parameter Training

To applying this framework, we need to estimate parameters  $\lambda_\phi$  in Equation 6, and assign value for  $\lambda$  in Equation 5. In this work, we also propose to use coordinate ascent (CA) algorithm [29] to do it. The CA algorithm has demonstrated its simplicity and effectiveness though several previous works [5, 7]. Given a target optimization metric, it iteratively optimizes a series of one-dimensional line searches. In each iteration, one parameter is processed, while all other parameters are fixed.

We perform the optimization in two stages. Firstly, we optimize  $\lambda_\phi$  used in *local* factor function for concept weighting. Then we fix the parameters for weights and optimize parameters in global factor function. Features  $GF(c)$ ,  $QF(c)$ , and  $NF(c)$ , which are described in Table 2 are normalize by their respective maximum frequencies. The whole iteration is performed until the performance gain is below a given threshold.

## 5. EXPERIMENTS

In this section, we present experimental results. We compare the proposed factor graph based map search (FGMS) method with several semi-structured retrieval models, a query parsing based method and three online map search services. Evaluation queries are randomly sampled from a query log. In the following sections, we firstly describe the corpus, evaluation queries, and golden standard construction strategies. Then we present comparative results to demonstrate the effectiveness of the proposed method.

### 5.1 Corpus

The spatial database used in this work contains a total of 10.8 million POIs. Each POI in the database has nine attributes: name, address, city, state, category, ZIP code, telephone number, longitude, latitude. The average length of “Name” and “Address” are 9.17 and 16.26 characters respectively.

We use queries belonging to four cities Shanghai, Hefei, Nanjing, and Shenyang for evaluation. For each city, we randomly selected a number of queries from the query log. To compare with commercial services, we also convert these queries and check that whether at least one relevant result existed in their spatial databases. Among the queries which all the commercial services and our spatial database contain relevant results, we randomly selected 600 queries for each city. There are 2,400 queries in total. Queries are split into two sets: 400 queries are used to train the parameters, the others are used to evaluate. Five human annotators were instructed to rate the relevance between query and retrieved results. A three-point scale is given for each query and candidate pair by annotators. The following descriptions of the scale are shown to annotators.

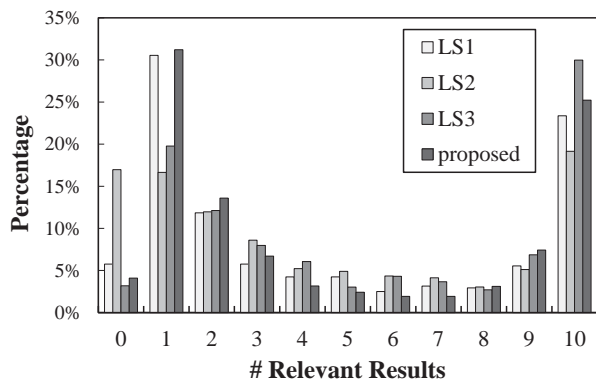


Figure 2: Number of relevant results per query.

- 0: Not relevant - the candidate in the result is not at all relevant to the query.
- 1: Moderately relevant - the candidate in the result is relevant to the query in some parts. For example, for the query “Guilin Mansion”, some POIs’ addresses may contain it.
- 2: Relevant - the candidate in the result is the result describes in the query.

The final score for each result is set to the value which the maximum number of annotators labeled. To compare with commercial online map search services, we also collect top 10 results of the 2,000 evaluation queries from Baidu<sup>1</sup>, Sogou<sup>2</sup>, and Google<sup>3</sup> through APIs provided by themselves. Since different services may use different spatial databases, annotators were also asked to label each of them with the same mechanism.

Even with the description of detail evaluation mechanism, labeling task is also a highly subjective process. To evaluate the quality of corpus, we validate the agreements of human annotations using Cohen’s kappa coefficient. The average  $\kappa$  among all annotators is 0.526. If we omit the difference between 1 and 2 scales, the average  $\kappa$  improves to 0.708. It indicates that the annotations of the corpus are reliable.

Figure 2 illustrates the distribution of the numbers of relevant results per query using different methods. LS1, LS2, and LS3 represent three commercial search services. We treat results whose final scores are 1 and 2 as relevant for the query. The zero number of relevant results means that no relevant results were returned by the method. Since LS2 cannot handle verbal terms well, about 16.9% queries do not have relevant results. We manually remove verbal terms from these queries, relevant results for most of the queries can be returned by LS2. Due to different strategies used by commercial services, the distributions of number of relevant queries are much different from each other.

### 5.2 Experimental Setup

We implement the calculation Eq.(4) with two steps. Firstly, Indri [33] is used to calculate the first part of Eq.(4)  $(\sum_{c \in \Pi_C} \sum_{i=1}^m (\omega(c)P(att_i|c)f(c, att_i)))$ . Then, the top 100

<sup>1</sup><http://developer.baidu.com/map/>

<sup>2</sup><http://map.sogou.com/api/>

<sup>3</sup><http://developers.google.com/maps/>

**Table 3: Retrieval effectiveness comparisons with different retrieval methods and three commercial online services. Best result per column is marked in boldface.**

Methods	Hefei					Nanjing				
	nDCG@10	MAP <sub>12</sub>	MAP <sub>2</sub>	P <sub>12</sub> @10	P <sub>2</sub> @10	nDCG@10	MAP <sub>12</sub>	MAP <sub>2</sub>	P <sub>12</sub> @10	P <sub>2</sub> @10
LS1	0.8393	83.5%	76.1%	79.2%	72.6%	0.8958	89.1%	79.8%	87.8%	78.6%
LS2	0.7116	67.8%	46.5%	61.1%	42.3%	0.7286	72.1%	55.8%	61.4%	50.7%
LS3	0.9232	91.4%	87.0%	<b>87.5%</b>	83.1%	0.9284	<b>92.2%</b>	84.3%	86.1%	79.1%
CRFs	0.8865	82.8%	82.17%	84.7%	83.9%	0.8642	79.4%	79.1%	89.6%	88.8%
PRMS	0.7522	63.9%	63.4%	72.6%	62.1%	0.7140	64.5%	53.1%	71.2%	56.9%
HLM	0.6417	62.9%	52.3%	55.8%	47.2%	0.6335	61.4%	49.4%	59.1%	51.6%
FGMS	<b>0.9529</b>	<b>91.7%</b>	<b>90.0%</b>	87.1%	<b>86.4%</b>	<b>0.9557</b>	91.1%	<b>90.5%</b>	<b>89.8%</b>	<b>89.2%</b>

Methods	Shanghai					Shenyang				
	nDCG@10	MAP <sub>12</sub>	MAP <sub>2</sub>	P <sub>12</sub> @10	P <sub>2</sub> @10	nDCG@10	MAP <sub>12</sub>	MAP <sub>2</sub>	P <sub>12</sub> @10	P <sub>2</sub> @10
LS1	0.8235	62.2%	37.6%	60.3%	36.4%	0.8257	81.8%	71.1%	80.4%	69.7%
LS2	0.7648	56.9%	27.4%	45.6%	22.0%	0.6980	71.8%	61.7%	68.1%	57.3%
LS3	0.8802	80.2%	<b>69.8%</b>	87.9%	63.9%	0.8579	85.4%	81.7%	<b>85.4%</b>	81.4%
CRFs	0.6231	46.2%	43.3%	79.0%	71.7%	0.8456	75.5%	75.5%	84.2%	81.5%
PRMS	0.7614	66.7%	56.2%	81.0%	63.6%	0.6974	67.8%	59.8%	72.7%	64.9%
HLM	0.6718	57.1%	51.0%	86.5%	<b>71.4%</b>	0.6371	59.2%	53.5%	63.8%	53.2%
FGMS	<b>0.9014</b>	<b>85.0%</b>	66.7%	<b>89.4%</b>	69.4%	<b>0.9486</b>	<b>89.3%</b>	<b>87.1%</b>	84.5%	<b>82.2%</b>

retrieved results are reranked based on the whole equation. Indri<sup>4</sup> is an open source search engine and provides a rich structured query language. Hence concepts and weighting schemes can be easily implemented.

To evaluate the retrieval performance, we employ the following standard performance measures: *Normalized Discounted Cumulative Gain* at 10 (nDCG@10), *Mean Average Precision* (MAP) and *Precision* at 10 (P@10). The graded relevance score in nDCG is set based the relevance score in ground truth. For the evaluation metric *Precision*, we use two methods: P<sub>2</sub>@10 and P<sub>12</sub>@10. P<sub>2</sub>@10 is used to measure the precision of top results whose relevance scores are equal to 2. P<sub>12</sub>@10 is used to measure relaxed relevant results whose relevance scores are equal to 1 or 2. The two-tailed t-test is conducted for significance. *MAP* is the same as *Precision*. Both MAP<sub>12</sub> and MAP<sub>2</sub> are used as evaluation metrics. The *ireval* package provided in the Lemur toolkit is used for evaluation and significance test.

### 5.3 Performance Evaluation

In this section, we compare the performance of the proposed method FGMS, with two semi-structured retrieval methods, a query parsing based method and three commercial online map search services. Two semi-structured retrieval methods are used to compare with. **HLM** represents the method proposed by Ogilvie and Callan [30]. It assigns different weights for each element based on the length of text content or the importance for retrieval. In this work, we tune the weights based on training data and distributions of attributes in sampled query log. **PRMS**, which is proposed by Kim et al. [23], uses the mapping probability as weights to combine each attribute into a final score. We re-implement it and tune the parameters with the same training data as described above. **CRFs** represents the query parsing based

method. Firstly, we use conditional random fields to detect street, city, POI name, and other attributes. Then, these identified subsequences are used as queries for corresponding attributes for retrieving and ranking. To train the model, we manually labeled 2,000 queries. The token accuracy achieves 87.10%, which is comparable with the state-of-the-art systems [34, 10].

Table 3 shows the performance results of baseline methods, three commercial services as well as our proposed method. We can observe that the proposed FGMS method is superior to the baseline methods, which only consider weights of attributes. For all the spatial databases of different cities, the result holds with all the evaluation metrics. Query parsing based method, CRFs, is better than PRMS, HLM, and LS2 in most evaluation metrics. However, due to the fact that its performance is relied heavily on the results of attribute detection, the variance of the performance is larger than others. It is also clear from the Table 3 that the PRMS method outperforms the HLM method in most cases. This result demonstrates the effectiveness of attribution weighting based on corpus statistic. In most cases, the retrieval effectiveness is also improved by the concept weighting and word sequence similarity.

Although different commercial services use different spatial databases, some evaluation metrics can also be used to compare with in some degree. From the results, we can observe that the proposed method achieve better performance than commercial search services in most cases, especially under the evaluation metric P<sub>2</sub>@10. The performance of LS3 is the best one among all the commercial services.

To detailedly analyze the performances of methods for different kinds of queries, we compare performances of categories belonging to the same group. Fig. 3, Fig. 4, and Fig. 5 show the results of comparisons. As discussed in Section 3.1, we establish ten query categories, which can be grouped into 3 groups.

<sup>4</sup><http://www.lemurproject.org/indri.php>

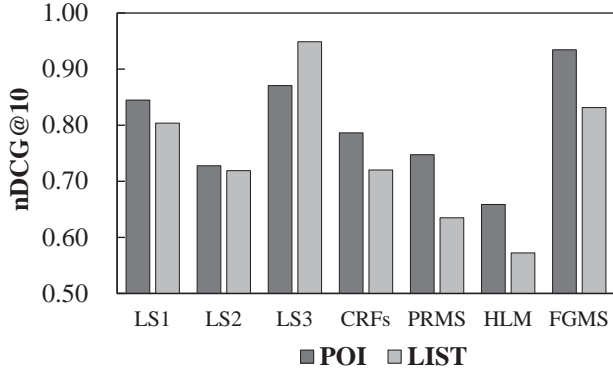


Figure 3: Performances of different methods on POI and List queries.

Fig. 3 depicts the performance comparison of different methods in processing queries belonging to the categories of POI and list. We can observe that except for LS3, the other methods achieve better performance in processing POI queries. For list queries, LS3 achieve the best result. The performance of the proposed method FGMS is significantly better than other methods in processing POI queries. For list queries, FGMS only worse than LS3. HLM and PRMS can not process list queries well either, comparing with other methods.

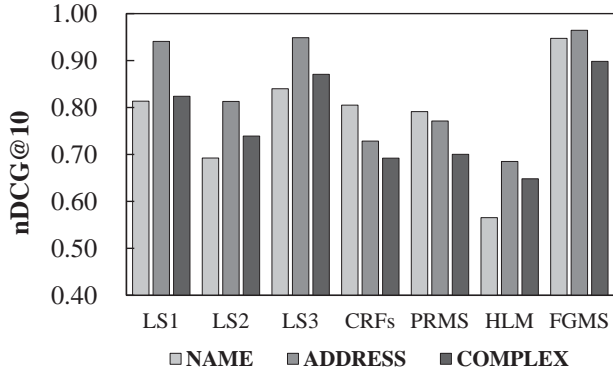


Figure 4: Performances of different methods on Name, Address and Complex queries.

Fig. 4 shows the performance comparisons of different methods on name, address, and complex queries. For address queries, most of the methods, except CRFs and PRMS, achieve better performance than name and complex queries. We think that it is because almost all the address queries belonging to POI category and address queries are usually related to only one POI. As the same reason, performances of complex queries are better than name queries can be explained. Since queries containing addresses usually contain more characters than name queries, the performance of CRFs may suffer from the low entity identification precision in processing address and complex queries.

The performance comparisons of different variants of queries are shown in Fig. 5. We observe that LS1, LS2, LS3, and the proposed FGMS can handle verbose queries well. However, HLM and PRMS are suffer from verbal

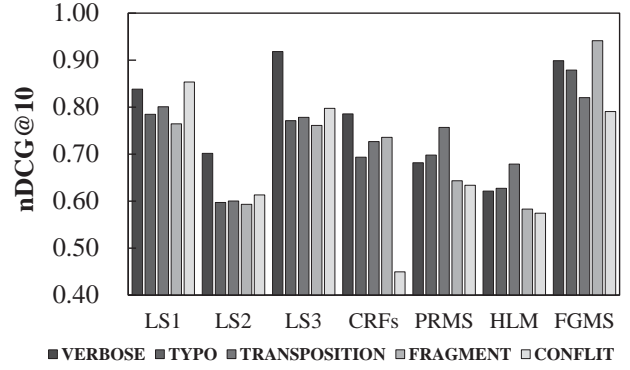


Figure 5: Performances of different methods on Verbose and Normal queries.

terms. We think that concept weighting contribute a lot for this issue. For conflict queries, all three commercial search engines can handle them well comparing with other query variants. LS1 achieve the best result in processing conflict queries. CRFs highly suffers from the conflict information. We think that the hard weighting strategies of CRFs may be the primary reason. Comparing to HLM and PRMS, the proposed FGMS achieve better performance for all variants of queries than them. Hence, the contribution of concept weighting and word based similarity can be demonstrated. According to the static shown in Table 1, a large of queries are not exactly same as the items in spatial database. Hence, methods which can correctly process variants can achieve better performance.

## 5.4 Parameter Analysis

In this section, we analyze the inference of parameters used in the proposed method. In Section 4.6, we proposed to use coordinate ascent algorithm to optimize the parameters, which are used in local and global factor functions. To verify the stableness of parameter estimation, we list the value of parameters and present their inferences to one of the evaluation metric nDCG@10.

Table 4 shows the optimized weights of concept importance features and global factor. The corresponding results of these parameters are shown in Table 3. We observe that among all the features used for concept weighting QF is the highest. It demonstrates that highly distinct concepts can be found from query logs. The optimized weight  $\lambda$  for global factor is 0.08. We also note that all the parameters are assigned positive weight. This indicates all the features can contribute to the retrieval performance.

Fig. 6 provides analyses of robustness of weights for concept importance features. It shows the nDCG@10 scores of different weights for GF, QF, NF, CB, SB, and TB in training data. When varying weights for one feature, weights of the other features are set to the optimized values listed in Table 4. We observe that although the weight of GF is relatively small, the performance drops quickly without it. Consistently with previous conclusion, the effectiveness gains achieved by QF also demonstrate its importance. Fig. 6 also demonstrates the effectiveness of each concept features. From the perspective of difference with and without a feature, the performance order is QF > GF > NF > TB > CB > SB.



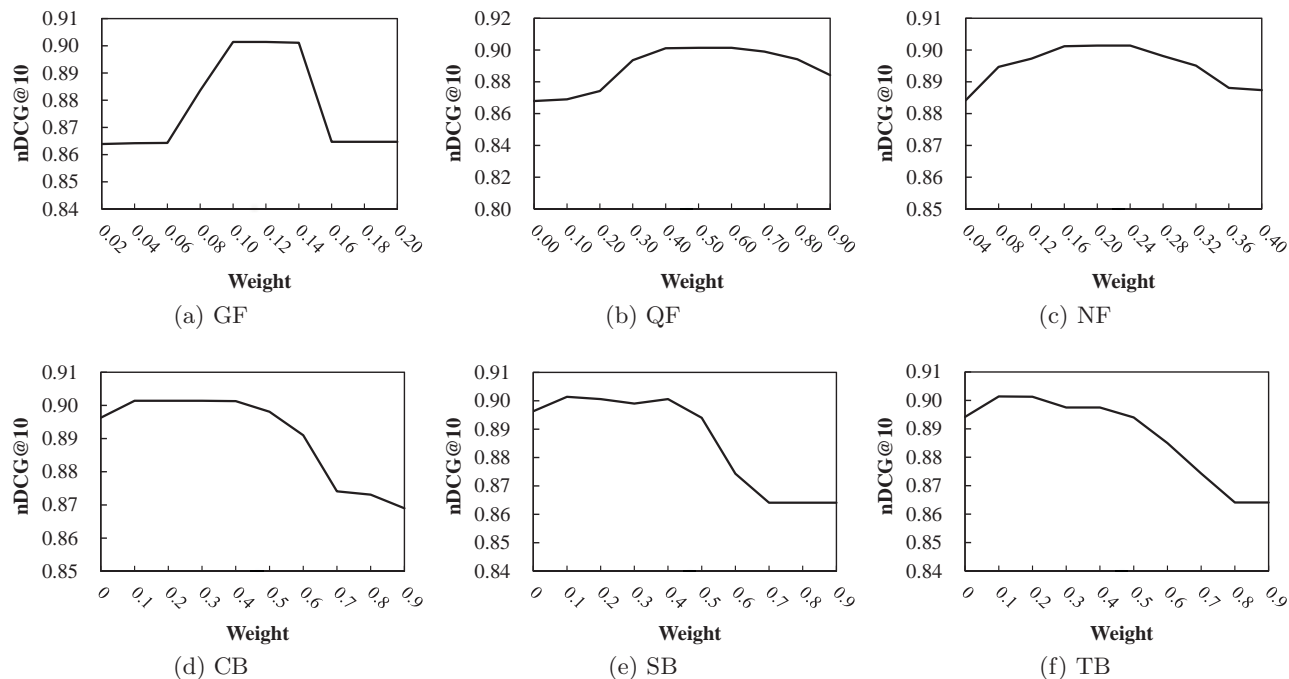


Figure 6: Robustness evaluation of concept importance features.

Table 4: The optimized parameterizations of local and global factors.

$\varphi$	Weight
GF	0.12
QF	0.52
NF	0.24
CB	0.22
SB	0.08
TB	0.12
$\lambda$	0.08

## 6. CONCLUSIONS

In this paper, we focused on the location search task with freeform queries. We converted the problem as a semi-structured information retrieval task and proposed a novel factor graph based map search method, which incorporated concept weighting and attribute selection together. To analyze the intention of users, we collected a number of query logs with the help of a commercial online map search company in China. We defined several query categories and analyzed the distributions of different categories. Several experimental comparisons demonstrated that the proposed method outperformed baseline methods and some commercial systems for retrieving unstructured location queries. The results also indicated that concept weighting, attributes importance, and word based similarity metric benefited the performance gains.

## 7. ACKNOWLEDGEMENT

The authors wish to thank the anonymous reviewers for their helpful comments and Suntec for providing us the

valuable query logs. This work was partially funded by 973 Program (2010CB327900), Shanghai Leading Academic Discipline Project (B114), National Natural Science Foundation of China (61003092, 61073069), National Major Science and Technology Special Project of China (2014ZX03006005), Key Projects in the National Science & Technology Pillar Program(2012BAH18B01) and Shanghai Municipal Science and Technology Commission (12511504500).

## 8. REFERENCES

- [1] R. J. Agrawal and J. G. Shanahan. Location disambiguation in local searches using gradient boosted decision trees. In *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems, GIS '10*, 2010.
- [2] S. Agrawal, K. Chakrabarti, S. Chaudhuri, V. Ganti, A. C. Konig, and D. Xin. Exploiting web search engines to search structured databases. In *Proceedings of the 18th international conference on World wide web, WWW '09*, 2009.
- [3] G. Andogah, G. Bouma, J. Nerbonne, and E. Koster. Geographical scope resolution. In *Methodologies and Resources for Processing Spatial Language Workshop at the 6th Conference on Language Resources and Evaluation, LREC*, 2008.
- [4] S. Basu Roy and K. Chakrabarti. Location-aware type ahead search on spatial databases: semantics and efficiency. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of data, SIGMOD '11*, 2011.
- [5] M. Bendersky and W. B. Croft. Modeling higher-order term dependencies in information retrieval using query hypergraphs. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval, SIGIR '12*, 2012.
- [6] M. Bendersky, D. Metzler, and W. B. Croft. Learning concept importance using a weighted dependence model. In *Proceedings of the third ACM international conference on Web search and data mining, WSDM '10*, 2010.

- [7] M. Bendersky, D. Metzler, and W. B. Croft. Parameterized concept weighting in verbose queries. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, SIGIR '11, 2011.
- [8] V. Bicer, T. Tran, and R. Nedkov. Ranking support for keyword search on structured data using relevance models. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, CIKM '11, 2011.
- [9] X. Cao, G. Cong, C. S. Jensen, and B. C. Ooi. Collective spatial keyword querying. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of data*, SIGMOD '11, 2011.
- [10] W. Chen, Y. Zhang, and H. Isahara. Chinese named entity recognition with conditional random fields. In *Proceedings of the Fifth SIGHAN Workshop on Chinese Language Processing*, pages 118–121, Sydney, Australia, July 2006. Association for Computational Linguistics.
- [11] Y. Chen, W. Wang, Z. Liu, and X. Lin. Keyword search on structured and semi-structured data. In *Proceedings of the 2009 ACM SIGMOD International Conference on Management of data*, SIGMOD '09, 2009.
- [12] Y.-Y. Chen, T. Suel, and A. Markowetz. Efficient query processing in geographic web search engines. In *Proceedings of the 2006 ACM SIGMOD international conference on Management of data*, SIGMOD '06, 2006.
- [13] M. Christoforaki, J. He, C. Dimopoulos, A. Markowetz, and T. Suel. Text vs. space: efficient geo-search query processing. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, CIKM '11, 2011.
- [14] G. Cong, C. S. Jensen, and D. Wu. Efficient retrieval of the top-k most relevant spatial web objects. *Proc. VLDB Endow.*, 2(1), Aug. 2009.
- [15] N. Datha, T. Joshi, J. Joy, and V. Sengar. Custom local search. In *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, GIS '09, 2009.
- [16] I. De Felipe, V. Hristidis, and N. Rishe. Keyword search on spatial databases. In *Proceedings of the 2008 IEEE 24th International Conference on Data Engineering*, ICDE '08, 2008.
- [17] J. Fan, G. Li, L. Zhou, S. Chen, and J. Hu. Seal: spatio-textual similarity search. *Proc. VLDB Endow.*, 5(9), May 2012.
- [18] V. Ganti, Y. He, and D. Xin. Keyword++: a framework to improve keyword search over entity databases. *Proc. VLDB Endow.*, 3(1-2), Sept. 2010.
- [19] R. Göbel, A. Henrich, R. Niemann, and D. Blank. A hybrid index structure for geo-textual searches. In *Proceedings of the 18th ACM conference on Information and knowledge management*, CIKM '09, 2009.
- [20] M. Jaro. Probabilistic linkage of large public health data files. *Stat Med*, 14(5-7):491–8, 1995.
- [21] R. Jones, W. V. Zhang, B. Rey, P. Jhala, and E. Stipp. Geographic intention and modification in web search. *Int. J. Geogr. Inf. Sci.*, 22(3), Jan. 2008.
- [22] T. Joshi, J. Joy, T. Kellner, U. Khurana, A. Kumaran, and V. Sengar. Crosslingual location search. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '08, 2008.
- [23] J. Kim, X. Xue, and W. B. Croft. A probabilistic retrieval model for semistructured data. In *Proceedings of the 31th European Conference on IR Research on Advances in Information Retrieval*, ECIR '09, 2009.
- [24] F. Kschischang, B. Frey, and H.-A. Loeliger. Factor graphs and the sum-product algorithm. *Information Theory, IEEE Transactions on*, 47(2):498–519, feb 2001.
- [25] Z. Li, K. C. Lee, B. Zheng, W.-C. Lee, D. L. Lee, and X. Wang. Ir-tree: An efficient index for geographic document search. *IEEE Transactions on Knowledge and Data Engineering*, 23:585–599, 2011.
- [26] J. Lu, Y. Lu, and G. Cong. Reverse spatial and textual k nearest neighbor search. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of data*, SIGMOD '11, 2011.
- [27] W. Lu, S. Robertson, and A. MacFarlane. Field-weighted xml retrieval based on bm25. In *Proceedings of the 4th international conference on Initiative for the Evaluation of XML Retrieval*, INEX'05, 2006.
- [28] T. Mandl, F. Gey, G. Nunzio, N. Ferro, R. Larson, M. Sanderson, D. Santos, C. Womser-Hacker, and X. Xie. Geoclef 2007: The clef 2007 cross-language geographic information retrieval track overview. In C. Peters, V. Jijkoun, T. Mandl, H. Müller, D. W. Oard, A. Peñas, V. Petras, and D. Santos, editors, *Advances in Multilingual and Multimodal Information Retrieval*. 2008.
- [29] D. Metzler and W. Bruce Croft. Linear feature-based models for information retrieval. *Inf. Retr.*, 10(3), June 2007.
- [30] P. Ogilvie and J. Callan. Hierarchical language models for xml component retrieval. In *Proceedings of the Third international conference on Initiative for the Evaluation of XML Retrieval*, INEX'04, 2005.
- [31] C. Peters, V. Jijkoun, T. Mandl, H. Müller, D. W. Oard, A. Peñas, V. Petras, and D. Santos, editors. *Advances in Multilingual and Multimodal Information Retrieval, 8th Workshop of the Cross-Language Evaluation Forum, CLEF 2007, Budapest, Hungary, September 19-21, 2007, Revised Selected Papers*, volume 5152 of *Lecture Notes in Computer Science*, 2008.
- [32] V. Sengar, T. Joshi, J. Joy, S. Prakash, and K. Toyama. Robust location search from text queries. In *Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems*, GIS '07, 2007.
- [33] T. Strohmaier, D. Metzler, H. Turtle, and W. B. Croft. Indri: a language-model based search engine for complex queries. Technical report, in *Proceedings of the International Conference on Intelligent Analysis*, 2005.
- [34] L. Wang, S. Li, D. F. Wong, and L. S. Chao. A joint chinese named entity recognition and disambiguation system. In *Proceedings of the Second CIPS-SIGHAN Joint Conference on Chinese Language Processing*, 2012.
- [35] X. Xiao, Q. Luo, Z. Li, X. Xie, and W.-Y. Ma. A large-scale study on map search logs. *ACM Trans. Web*, 4(3), July 2010.
- [36] X. Yi, H. Raghavan, and C. Leggetter. Discovering users' specific geo intention in web search. In *Proceedings of the 18th international conference on World wide web*, WWW '09, 2009.
- [37] 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, 2001.
- [38] D. Zhang, Y. M. Chee, A. Mondal, A. K. H. Tung, and M. Kitsuregawa. Keyword search in spatial databases: Towards searching by document. In *Proceedings of the 2009 IEEE International Conference on Data Engineering*, ICDE '09, 2009.
- [39] Y. Zhou, X. Xie, C. Wang, Y. Gong, and W.-Y. Ma. Hybrid index structures for location-based web search. In *Proceedings of the 14th ACM international conference on Information and knowledge management*, CIKM '05, 2005.