Proximity²-aware Ranking for Textual, Temporal, and Geographic Queries (extended version)*

Jannik Strötgen and Michael Gertz

Institute of Computer Science, Heidelberg University,
Im Neuenheimer Feld 348, 69120 Heidelberg, Germany
{jannik.stroetgen,michael.gertz}@informatik.uni-heidelberg.de
http://dbs.ifi.uni-heidelberg.de/

Abstract. Temporal and geographic information needs are frequent and important but not well served by standard IR systems. There are neither good ways to add temporal or geographic constraints to a normal text query, nor are geographic and temporal expressions in the documents interpreted as such kind of information, i.e., their semantics is not exploited. Recent approaches address such needs by extracting and normalizing temporal and geographic expressions from documents. They calculate specific scores for the temporal and/or geographic parts of a query. However, all approaches assume independence between the different query parts.
In this paper, we present a new model to rank documents according to combined textual, temporal, and geographic queries. In this model, the independence assumption between the query parts is eliminated by calculating different proximity scores. Thus, documents are regarded to be more relevant if terms and expressions satisfying the different query parts occur close to each other in a document. In addition, we present a second type of proximity feature addressing the problem of sparse results. For this, we determine the temporal and geographic distance between expressions in a document and the queried time interval and geographic region. This allows to take into account documents containing expressions close to the time interval or region of interest. As our evaluations based on the NTCIR-GeoTime data show, our proposed model outperforms baseline models that do not use either of proximity information.

1 Introduction

In many types of documents, temporal and geographic information plays a pivotal role. Unfortunately, users' temporal and geographic information needs are not well handled by standard IR systems although such aspects of queries are important in many search scenarios. For example, Nunes et al. report that 1.5% of

*This technical report is an extended version of [18]: Jannik Strötgen and Michael Gertz: Proximity²-aware Ranking for Textual, Temporal, and Geographic Queries. In *CIKM'13*, pages 739–744, 2013.

the queries in an analyzed query log contain explicit temporal information [11]. Zhang et al. attest the importance of geographic information by reporting that 12.7% of the queries in an analyzed query log contain some kind of geographic information [22].

There probably would be even more search queries that include temporal and geographic information needs if there were better ways to properly query for documents whose content is constrained to specific time intervals or geographic regions. However, in queries as well as in documents, temporal and geographic expressions are usually treated in the same way as regular terms. Thus, their meaning is lost and cannot be exploited to satisfy respective information needs.

A typical search scenario with a combination of a textual, temporal, and geographic information need is to query for documents about events, which are composed of a specific time and some specific place. Assume, for example, the information need "world records between 1965 and 1974 in Central Europe". Here, one is faced with two problems: (i) the time interval and the geographic region have to be interpreted as such kind of information, and (ii) temporal and geographic expressions in the documents have to be verified if they belong to the specified interval and region, respectively. However, if temporal and geographic expressions are not identified and normalized, a search engine cannot assign different relevance scores to different documents. Consider the two simple documents "In 1972, he set a world record in Munich" and "He set a world record in Beijing in 2008". By identifying temporal and geographic expressions, it is possible to use the knowledge that Munich is located in Central Europe while Beijing is not, and that 1972 is within the interval 1965 to 1974, while 2008 is not. This issue is even more problematic if relative expressions such as "ten years later" or "in the following month" occur in a document.

To address the above shortcomings, there recently have been approaches to incorporate temporal information [2], geographic information [13], and both [9] into retrieval models. However, all approaches assume that the textual, geographic, and temporal information needs formulated in a query are independent of each other. Thus, for such queries, the models calculate independent scores for each part and finally combine them into a single score for ranking documents. However, this independence assumption is problematic because the *proximity among expressions in a document* satisfying the query terms is disregarded. Similar to the previous example, assume the two simple documents (A) "He set a world record in 1972 ... he died in 2008" and (B) "He set a world record in 2008 ... he was born in 1972", (A) should be given a higher relevance score for the example query due to the proximity between "world record" and "1972" in the document.

Another aspect not considered by related approaches is that the *spatial and temporal proximity* of expressions in documents to the temporal and spatial query terms is not considered at all. Assume the simple query "world record 1990 Germany". Also assume there is no document satisfying both the temporal and spatial information needs but there are three documents (A) "In 1991 he set a world record in Germany", (B) "In 1990 he set a world record in France", and (C) “In 1992 he set a world record in Japan”. Both documents (A) and (B) seem to be close to satisfying the information need as 1990 is temporally close to 1991, and France is spatially close to Germany. Both are temporally and spatially in closer proximity than “1992” and “Japan” in document (C).

In this paper, we present a novel ranking approach that effectively considers both the proximity of text, temporal, and geographic expressions in documents and the spatial and temporal proximity of expressions to query terms. For this, we build on the well-known and widely used ranking model Okapi BM25 [14] and extend it to incorporate the above two proximity features. For determining temporal and geographic proximity, we employ standard distance metrics for time and space as well as spatial information associated with geographic expressions in documents. Using the semantics of distances and proximity in space and time naturally allows to increase the number of ranked documents because documents not fully satisfying the temporal and geographic queries can be judged based on their distance to the interval/region of interest. Both temporal and geographic expressions detected in documents are suitable encoded for efficient look-ups to determine documents relevant to queries that formulate temporal and spatial information needs.

In summary, the main contributions of this paper are:

– An IR model that addresses textual, temporal, and geographic information needs, and which – in contrast to previous works – takes into account the dependency between the three query parts by using proximity information for the final ranking.
– A proximity feature, as part of the IR model, for processing temporal and geographic information needs by calculating temporal and geographic distances between expressions in documents and the query interval/region if the query is not directly matched by documents.
– An extensive evaluation of the proposed ranking model using datasets and query patterns adopted from the NTCIR-GeoTime challenge.

The remainder of the paper is structured as follows. After a brief review of related work in the following section, we detail the problem statement and assumptions in Section 3. The description of our ranking model in Section 4 is followed by a presentation of the evaluation in Section 5.

2 Related Work

Temporal and geographic information retrieval are often considered separately. Current research trends and challenges in temporal IR such as temporal clustering of search results and temporal querying are discussed in [1]. While there are some approaches to improve search results by taking into account document creation times, e.g., to favor recent documents [3, 7], there is only little work using temporal information mentioned in documents for querying. Berberich et al. do so by integrating temporal information into a language model for addressing temporal information needs [2]. In contrast to using temporal information for querying, there is more work in geographic IR to use geographic information extracted from documents for querying document collections, e.g., STEWARD [8], SPIRIT [13], and several works that appeared in the Geographic Information Retrieval Workshop series.

Recently, there have been approaches to combine temporal and geographic IR. For example, event-centric search and exploration of document collections with events being considered appropriate combinations of temporal and geographic expressions occurring in documents is presented in [16]. However, the authors do not provide a ranking approach for their framework. In particular, none of the above approaches consider any proximity aspects of expressions in the documents matching temporal/geographic parts of search queries.

In addition to the methods outlined above, the geographic and temporal information retrieval challenges NTCIR-GeoTime were organized [5, 6]. In comparison to our work, the focus of these challenges is on querying for temporal and geographic answers, i.e., to process queries of the form "when and where did something happen". As pointed out by the organizers, the results show that "semantic questions require semantic processing to deliver results beyond bag-of-words search" [4]. That is, temporal and geographic information embedded in documents should be handled in a special way. Most similar to our work is a Lucene extension to process temporal and geographic queries [9]. However, they assume that the textual, temporal, and geographic parts of a query are independent of each other and do not take into account the proximity in the documents between terms satisfying the different query parts. This is a crucial weakness as proximity of query terms in documents plays already an important role in standard (commercial) search engines. The same independence assumption for query terms as done in the Lucene extension is also made in the methods proposed in [2] and [13] for temporal and geographic queries, respectively.

Clearly, there has been substantial work on showing that using information about the proximity of terms matching query terms in documents significantly improves search result, e.g., by adding a proximity score to Okapi BM25 [14]. In our work, we extend these ideas in two ways. First, we develop a proximity measure for terms and expressions satisfying the textual, temporal, and geographic parts of a query and thus eliminate the independence assumption adopted in previous approaches on temporal and/or geographic IR. For this, we use the results of the detailed analysis of different methods to determine the proximity between query terms described in [19]. Second, taking the full semantics of temporal and geographic expressions in documents and queries into account, as suggested in the context of the NTCIR-GeoTime challenges, our approach furthermore adds temporal and geographic proximity aspects to the ranking approach.

3 Problem Statement and Model Assumptions

In this section, we formulate the problem statement, describe some basic concepts regarding temporal and geographic information embedded in documents, and define assumptions for our proximity^2-aware ranking model for textual, temporal, and geographic information needs. The model itself will be introduced in Section 4.

> **Problem Statement:** Given a document collection D and a search query composed of a textual, a temporal, and a geographic part, return a list of documents $d_i \in D$ ranked by a score measuring how well the combined information need is satisfied. The score should consider the documents' relevance on all parts of the information need and a proximity score covering the distance between terms satisfying the different parts of the information need in the documents.

3.1 Temporal and Geographic Information Extraction

Key to our proposed approach is that in a preprocessing step for a given corpus, temporal and geographic expressions in documents are identified as such and normalized in a way that allows for efficient comparison and matching. As indicated in the introduction, the main reason for temporal and geographic information not being well handled by standard search engines is that respective expressions in documents are usually treated as regular terms, that is, without any further semantics.

To accomplish the above preprocessing tasks, for temporal expressions so-called temporal taggers are used. A temporal tagger aims at identifying terms in a document (or query) that correspond to temporal concepts. These can be either explicit expressions, such as "December 25, 2009" or "March, 2013", relative expressions, e.g., "on Monday" or "last year", or implicit temporal expressions, such as "Christmas 2009". While explicit expressions are easy to normalize into some standard format, e.g., based on TimeML [12], normalizing relative expressions requires some reference time, which can either be the document creation time or some explicit expression in the document. Implicit expressions require some background knowledge, for example, the names and dates of holidays. For the above expressions, for example, a temporal tagger would normalize "December 25, 2009" and "Christmas 2009" both to the same date value "2009-12-25". Note that expressions can also be time intervals with normalized start and end times (see, e.g., [12]). The result of a temporal tagger, when applied to a document, is basically a set of triples, each triple consisting of the term(s) forming the temporal expression t, the offset $p(t)$ of the expression in the document, and the normalized value $v(t)$.

Similarly, for geographic information embedded in documents (and queries), geo-taggers are used to detect and normalize respective expressions. Normalization typically associates a spatial object, such as a point represented by latitude and longitude values or a bounding box, with each expression. Some geo-taggers such as Yahoo Placemaker [21] also provide further useful concept information about the geographic expressions found in a document. For example, for the expression "Munich" it would also give the hierarchy "Munich, Germany, Europe". Especially this additional information is useful for determining the proximity of geographic expressions based on their spatial distance, e.g., to a location mentioned in a query. Consequently, for a geographic expression g found in a document, a geo-tagger returns the expression g, its offset $p(g)$ in the document, and its normalized value $v(g)$, which can be a complex object such as a point or bounding box.

3.2 Model Assumptions

Based on the explanations given in the previous section, we make the following assumptions for our model.

Profiles. Given a document collection D, all documents $d_i \in D$ are preprocessed with a temporal tagger and a geo-tagger. Thus, the temporal and geographic expressions in the documents are extracted, normalized to their standard values, and organized in *temporal* and *geographic document profiles*, respectively as:

$$tdp(d) = \{\langle v(t)_1, p(t)_1\rangle, ..., \langle v(t)_n, p(t)_n\rangle\}$$
$$gdp(d) = \{\langle v(g)_1, p(g)_1\rangle, ..., \langle v(g)_m, p(g)_m\rangle\}$$

In Section 5, we will give more details on how these profiles are computed and managed for a given document collection in a preprocessing step for subsequent efficient lookup and ranking tasks required by our framework.

Queries. A query consists of a textual part q_{text} (terms), a temporal part q_{temp} (one or more time intervals), and a geographic part q_{geo} (one or more geographic regions specified by, e.g., bounding boxes). Thus, we define a query as:

$$q = \{q_{text}, q_{temp}, q_{geo}\}$$

It should be noted that the user can specify such a query in different ways, depending on what query interface is provided. For a normal textual query, geographic and temporal expression (including time intervals such as "1999 to 2011") are identified and normalized, very much in the same way as expressions in documents are handled. One can also envision a graphical query interface in which the user specifies a point location or a bounding box plus some time interval using a time-slider. Here, one would already obtain normalized values for respective query components.

The document profiles are used to evaluate q_{temp} and q_{geo} and to determine the temporal and geographic proximity – based on normalized values – between expressions in the documents and the query parts. Again, implementation details on how to efficiently process q_{temp} and q_{geo} will be given in Section 5.

4 Proximity^2-aware Ranking Model

Based on the assumptions described in the previous section, we now incrementally develop our proximity^2-aware ranking model for textual, temporal, and geographic information needs formulated in search queries. For this, we first describe the key characteristics of the model. For a search query, we then detail the calculations of the textual score (Section 4.2), the temporal and the geo scores (Section 4.3), and the term proximity score (Section 4.4). Finally, we describe how the single scores are combined into one overall ranking score.

4.1 Model Characteristics

The key characteristics of the model can be summarized as follows:

- For the individual components q_{text}, q_{temp}, and q_{geo} present in a search query, single scores are calculated.
- Given a document, based on the distances between terms and expressions in the document satisfying the different query parts, a score is calculated (*term proximity score*).
- There will typically be documents not directly satisfying the q_{temp} and q_{geo} parts of a query. Assume, for example, if q_{temp} = "March 2013" but in a document the only temporal expression found is "December 2012". For such documents, still positive temporal and geographic scores can be calculated. This is because the temporal and geographic distances between expressions in the documents and the time interval and region specified in a query are taken into account (*temporal and geographic proximity*).

4.2 Textual Ranking

One part of our proximity[2]-aware ranking model is to calculate a score s_{text} for the textual part q_{text} of a query. For this, we use Okapi BM25 [15], a well-known standard measure for ranking documents according to a textual query. This measure is mainly based on the term frequency $c(w, d)$ and the inverse document frequency (first fraction in Equation 1, with $df(w)$ being the number of documents containing term w). For the text part q_{text} of a query and a document $d \in D$ with $|D| = N$, it is defined as follows:

$$\begin{aligned} s_{text}(q_{text}, d) := \sum_{w \in q_{text} \cap d} & ln \frac{N - df(w) + 0.5}{df(w) + 0.5} \\ & \times \frac{(k_1 + 1) \times c(w, d)}{k_1((1 - b) + b\frac{d_{len}}{D_{avgl}}) + c(w, d)} \end{aligned} \tag{1}$$

Note that the score is length-normalized using the length d_{len} of a document d and the average document length D_{avgl} of all documents in D. The parameters $k_1 \in [1.2, 2.0]$ and $b = 0.75$ calibrate term frequency scaling and length normalization scaling [10]. For every document $d_i \in D$, this formula determines a textual score $s_{text}(q_{text}, d_i)$ representing the relevance of document d_i with respect to q_{text}. Based on these key concepts of this ranking formula, in the following, we develop our ranking functions for q_{temp} and q_{geo}.

4.3 Temporal and Geographic Ranking

Similar to the score s_{text}, we want to calculate the scores s_{temp} and s_{geo} representing how well a document satisfies the other two query parts q_{temp} and q_{geo}, respectively. A difference between validating q_x (with $q_x \in \{q_{temp}, q_{geo}\}$) and validating q_{text} is that q_x may consist of one or more time intervals/geographic regions and q_{text} consists of one or more terms. More importantly, the regular terms considered in s_{text} and intervals/regions have different characteristics. While the terms matching q_{text} directly occur "as is" in documents (after preprocessing such as stemming), expressions that may match q_x have to be validated based on their normalized values and, in addition, do not necessarily have to completely match q_x. These differences have to be taken into account when calculating s_{temp} and s_{geo} following the idea of computing s_{text}.

Given a query q_x and a document profile $xdp(d)$ of a document d (with $xdp(d) \in \{tdp(d), gdp(d)\}$), the normalized values of expressions in $xdp(d)$ can (i) be inside q_x, (ii) overlap q_x, or (iii) be outside q_x. Assume, for example, the normalized query time interval $q_{temp} = [1965, 1974]$. The expression "1972" corresponds to case (i), "1960s" to case (ii), and "1960" to case (iii). Note that for case (i), an expression may cover different parts of a query interval/region. For example, "September 1972" and "1972" are both in q_{temp} but cover different parts of it due to their different granularities. Berberich et al. assume for their approach to satisfy temporal information needs that the more of the query time interval is covered by a temporal expression, the more relevant the temporal expression [2]. However, we argue that it is only important whether an expression is within q_x or not, and that the coverage of q_x can be better determined based on *all* expressions in $xdp(d)$, as we will justify below.

Temporal and Geographic Proximity Due to the three different ways how expressions in a profile $xdp(d)$ are related to a query part q_x, we cannot simply use the term frequency as for regular terms matching q_{text}. Thus, we calculate the *weighted value frequency*, vf, which aggregates the *value weight* vw of every expression $x \in xdp(d)$:

$$vf(q_x, d) := \sum_{v(x) \in xdp(d)} vw(v(x), q_x), \text{ with} \tag{2}$$

$$vw(v(x), q_x) := \begin{cases} 1, & \text{if } v(x) \text{ is in } q_x \quad \text{(i)} \\ \frac{|v(x) \cap q_x|}{|v(x)|}, & \text{if } v(x) \text{ overlaps } q_x \quad \text{(ii)} \\ exp^{-\frac{\delta(v(x), q_x)}{|q_x|}}, & \text{if } v(x) \text{ is outside } q_x \quad \text{(iii)} \end{cases}$$

The first two cases (i) and (ii) are straightforward: if $v(x)$ is inside q_x, we want vw to be 1. If $v(x)$ overlaps q_x, we want vw to represent the proportion of $v(x)$ being inside of q_x, i.e., $\frac{|v(x) \cap q_x|}{|v(x)|}$. For example, given $q_{temp} = [1965, 1974]$, then $vw(\text{"1960s"}) = \frac{1}{2}$ and $vw(\text{"20th century"}) = \frac{1}{10}$.

For the third case (iii), however, that is, if $v(x)$ is outside q_x, we do not want vw to be simply 0 as we want to distinguish if $v(x)$ is (temporally/spatially) close to q_x or not. Thus, we introduce the first important proximity parameter of our model to calculate the distance δ between the normalized value $v(x)$ and q_x. This allows to score also documents with a $s_x > 0$ (with $s_x \in \{s_{temp}, s_{geo}\}$) that do not contain expressions directly satisfying q_x. We use the distance δ in relation to the "size" of q_x, which is denoted $|q_x|$. If q_x is a temporal expression, then, depending on the granularity of q_x, it can be the number of days, months, or years covered by q_x. Similarly, if q_x is a geographic expression, the size of q_x is simply the area described by q_x (based on its normalized value). Given two expressions q_{x1} and q_{x2} of the same type, intuitively, the following condition should hold:

if $\delta(v(x), q_{x1}) = \delta(v(x), q_{x2})$ **and** $|q_{x1}| < |q_{x2}|$
then $vw(\delta(v(x), q_{x1})) < vw(\delta(v(x), q_{x2})$

In other words, the same distance between a normalized value $v(x)$ and a normalized query value should result in a lower value weight if the size of the query interval/region is smaller.

For example, assume $v(t) =$ 1972-09-03 and two temporal query parts $q_1 =$ [1972-08-01, 1972-08-31] and $q_2 =$ [1972-08-30, 1972-08-31]. The distance to $v(t)$ is the same for both queries (3 days). However, due to the larger interval of interest formulated by q_1 (31 days), the distance of 3 days is less relevant than in the second case, where the length of q_2 is smaller (2 days). Examples for normalized geographic expressions found in documents and two query regions are devised similarly in our framework, based on the shortest distance between respective regions and the area of regions. For geographic expressions, it obviously becomes even simpler in case only geographic points (as normalized values) are considered.

The desired behavior of the value weight function $vw(v(x), q_x)$ with $\delta > 0$ can be described as follows: the smaller $\frac{\delta(v(x),q_x)}{|q_x|}$, the lower $vw(v(x), q_x)$ with its first derivative being negative and its second derivative being positive. This concave behavior is obtained by an exponential term of the form $exp^{-\frac{\delta(v(x),q_x)}{|q_x|}}$, so that we can summarize the behavior of the value weight function $vw(v(x), q_x)$ as defined in Equation 2.

In summary, an important ingredient of our novel ranking model is that for the temporal and geographic ranking functions, we use the weighted value frequency instead of the standard term frequency for regular terms. This approach appropriately considers the semantics of temporal and geographic expressions in terms of proximity of time intervals and geographic regions, respectively, based on well-defined distance metrics for time and space.

Coverage of the Query Interval/Region In the Okapi BM25 for the textual ranking score s_{text}, the second important feature besides the term frequency is the inverse document frequency. It carries information about how characteristic a document is for a query with respect to the document collection. For our modifications to BM25 for calculating the temporal and geographic scores s_{temp} and s_{geo}, we combine information about the document collection with information about the coverage of the query interval/region. Given q_x (with $q_x \in \{q_{temp}, q_{geo}\}$) and a document profile $xdp(d)$ (with $xdp(d) \in \{tdp(d), gdp(d)\}$), we calculate the ratio of distinct normalized values in $xdp(d)$ and the number of distinct normalized values in the combined document profile of all documents

$xdp(D)$ overlapping q_x. To avoid that the coverage is zero or undefined if a document or the document collection contains no normalized values overlapping with q_x, we add 0.5 to both counts. This is important since temporal and geographic scores should be positive in both cases for the temporal and geographic proximity introduced above to work effectively. By this, the coverage of a document without values overlapping with q_x is larger than 1 and the coverage is the same for all documents if no values in the document collection overlap with q_x.

$$coverage(d, q_x) := \frac{count_{dist}(v(x) \in xdp(d) : v(x) \cap q_x \neq \emptyset) + 0.5}{count_{dist}(v(x) \in xdp(D) : v(x) \cap q_x \neq \emptyset) + 0.5} \tag{3}$$

For example, given a temporal query "August 1972" and two documents with the first containing some temporal expressions referring to "1972-08-01" and the second containing some expressions referring to "1972-08-07" and "1972-08", respectively. In addition, in the corpus, there are ten distinct normalized values of temporal expressions that (partially) match the temporal query. Then, the temporal coverage of the first document is $\frac{1.5}{10.5}$ and the temporal coverage of the second document is $\frac{2.5}{10.5}$.

In our opinion, when being faced with a temporal or geographic query formulated as time intervals or geographic regions, the most relevant document does not necessarily cover the whole interval or region but contains many different normalized values in the interval or region of interest compared to other documents. Thus, we use Equation 3 as corpus-dependent coverage instead of using the plain coverage of q_x or the inverse document frequency as for terms.

Temporal and Geographic Scores Replacing the inverse document frequency by the coverage and the term frequency c by the weighted value frequency vf in Equation 1, the temporal and geographic scores s_{temp} and s_{geo} are now calculated as follows:

$$\begin{aligned} s_{temp}(q_{temp}, d) := \sum_{v \in q_{temp}} & coverage(d, v) \\ & \times \frac{(k_1 + 1) \times vf(v, d)}{k_1((1 - b) + b\frac{d_{len}}{D_{avgl}}) + vf(v, d)} \end{aligned} \tag{4}$$

$$\begin{aligned} s_{geo}(q_{geo}, d) := \sum_{v \in q_{geo}} & coverage(d, v) \\ & \times \frac{(k_1 + 1) \times vf(v, d)}{k_1((1 - b) + b\frac{d_{len}}{D_{avgl}}) + vf(v, d)} \end{aligned} \tag{5}$$

In the same way as s_{text} is defined in Equation 1, these scores are length-normalized, and the parameters k_1 and b are used to calibrate the scaling behavior.

Thus far, we now have scores to rank the individual components q_{text}, q_{temp} and q_{geo}, where for ranking q_{temp} and q_{geo} we introduced temporal and geographic proximity measures based on the distance of time intervals and geographic regions, respectively. We now turn to the second type of proximity measure, the term proximity, as another important ingredient to our ranking model.

4.4 Term Proximity Ranking

The relevance scores described in the previous sections represent independent scores for the query parts q_{text}, q_{temp}, and q_{geo} with respect to a document. While previous approaches combine such independent scores into a final ranking score for a document, we argue that this independence assumption is problematic. As illustrated in the example in Section 1, information about the proximity in a document between terms and expressions satisfying q_{text}, q_{temp}, and q_{geo} should be considered to reward documents in which the proximity among matching expressions is small, and to penalize documents where such a proximity is large.

Tao and Zhai analyzed different ways to measure the proximity between query terms matching a textual query in documents [19]. In their comparison of five measures, the minimum pair distance (shortest distance of two different query terms, independent of the number of query terms) performed best. Although we are faced with a slightly different problem here, because the terms and expressions for which we want to measure the proximity are of different types, we use their study as basis for developing the function to calculate the proximity score s_{prox}. For this, we transfer the minimum pair distance into a *minimum triple distance*. Given a document, such a distance then is naturally defined as the shortest distance among a term w of q_{text}, a temporal expression t satisfying q_{temp}, and a geographic expression g satisfying q_{geo}, denoted $prox(w, t, g)$. Clearly, the closer w, t, and g are together in a document, the higher should be the ranking for that document with respect to the query.

In contrast to the original proximity measure for two terms, there is no need that the three terms/expressions w, t, and g, respectively, occur within a few tokens, but it should be awarded if they occur within a few sentences. Thus, instead of the original concave function, we use the following proximity transformation function (containing cubic terms in both nominator and denominator):

$$s_{prox} := exp^{\frac{ln(0.5) \times prox(w,t,g)^3}{50^3}} \tag{6}$$

The behavior of Equation 6 is shown in Figure 1. Assuming a typical sentence length of 20 to 25 tokens [10], the function only slightly penalizes proximities within one or two sentences, but significantly penalizes proximities larger than three sentences since s_{prox} is convex for proximities smaller than 50 tokens and concave for larger proximities.

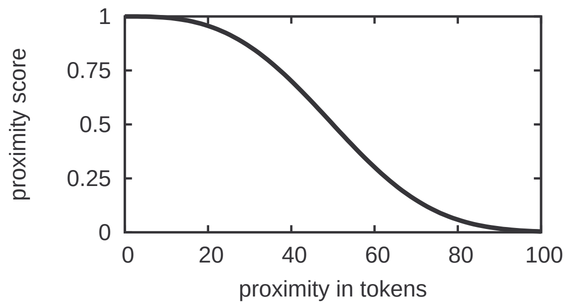

Fig. 1. Ideal shape of the proximity transformation function.

4.5 Combined Ranking

Having defined the separate scores for the textual, temporal, geographic, and proximity ranking, we are now finally faced with the same problem as similar approaches, namely, how to combine the single scores in a meaningful way. A typical way that also allows to specify weightings for the single scores is to use a linear combination. For this, we first normalize the s_{text}, s_{temp}, and s_{geo} scores by the maximum score for the given query, denoted $\hat{s}_{text}(q)$ etc. Thus, for each query the highest textual, temporal, and geographic scores is set to 1. The proximity score is already normalized, as described in the previous section. We therefore obtain the following score for a query q and document d:

$$\begin{aligned} s(q,d) :=& (1 - \alpha_t - \alpha_g) \frac{s_{text}(q,d)}{\hat{s}_{text}(q)} + \alpha_t \frac{s_{temp}(q,d)}{\hat{s}_{temp}(q)} + \\ & \alpha_g \frac{s_{geo}(q,d)}{\hat{s}_{geo}(q)} + \beta s_{prox}(q,d) \end{aligned} \tag{7}$$

The weights α_t and α_g are used to weight the importance of the three query components q_{text}, q_{temp}, and q_{geo}. In addition, β is used to weight the importance of the proximity measure. In our evaluation, which is detailed in the next section, we show the impact of varying the β parameter and analyze the influence of the proximity feature of our model.

5 Evaluation

Evaluating a combination of textual, temporal, and geographic information needs is difficult since there are no benchmarks from IR challenges such as TREC [20] in which a query consists of a textual, a temporal, and a geographic part. However, recently, the NTCIR-GeoTime challenges addressed a similar problem, namely ranking documents of a given document collection for a query having temporal and/or geographic aspects [5, 6].

topic	explicit constraint	positive judgment	topic	explicit constraint	positive judgment
0001		9	0014	geo, time	31
0002	geo	335	0015	time	71
0003		5	0016		320
0004		38	0017	geo	24
0005		8	0018	time	58
0006	geo	112	0019		79
0007		8	0020		9
0008	geo	172	0021	time	3
0009		49	0022	time	15
0010		10	0023	geo	27
0011	time	96	0024		48
0012		36	0025	geo	19
0013	geo	18			

Table 1. NTCIR-8 GeoTime topics with explicit constraints and the number of positive judgments.

5.1 NTCIR-8 GeoTime Data

For our evaluation, we used the NTCIR-8 GeoTime dataset consisting of 25 queries (called topics). As document collection, the 2002 to 2005 articles of the New York Times corpus are used (315,417 documents).[1] In addition to the topics, relevance judgments for each topic are also publicly available.[2] In the context of the GeoTime challenge, for each query the top-100 ranked documents of each system of the participating teams were judged resulting in 17,423 judgments in total. Many of the topics in the GeoTime data are of the form *"where and when happened X"*, but there are also some queries with explicit temporal constraints, explicit geographic constraints, or both. Table 1 gives an overview of the types of explicit constraints as well as the number of relevant judged documents for each topic. While we will have a closer look on the single topics when analyzing the results in Section 5.4, the varying numbers of positive judgments directly indicate the different levels of difficulty.

Due to the reproducibility and comparability of our evaluation results and due to the lack of other, more suitable benchmarks, we performed our evaluation based on the publicly available NTCIR-8 GeoTime data. However, to be able to process all queries with our proximity^2-aware ranking model, we had to slightly adapt the model for the queries without explicit temporal and/or geographic constraints. We describe these model adaptations in Section 5.2, where we also discuss how the parameters in our model are set. Then, in Section 5.3, we outline our document preprocessing, the used index structures, and the query processing. The evaluation itself is finally presented in Section 5.4.

[1] The New York Times corpus is available from the linguistic data consortium (http://www.ldc.upenn.edu/).

[2] The topics and relevance judgments are available from the Japanese National Institute of Informatics (http://research.nii.ac.jp/ntcir/).

5.2 Model Adaptation and Parameters

To be able to process all GeoTime queries and not only those with explicit temporal and geographic constraints, we adapt our proximity^2-aware model in the following way: In the absence of explicit temporal or geographic constraints, no s_{temp} and s_{geo} are calculated, respectively. However, since all queries have at least a latent temporal and geographic aspect ("when" and "where"), we calculate s_{prox} between terms matching the textual query and all temporal and geographic expressions in the documents.

The parameters for s_{text}, i.e., for the BM25 model, are set to standard values ($k_1 = 1.2$ and $b = 0.75$). The α-parameters of Equation 7 for weighting the single scores s_{text}, s_{temp}, and s_{geo} are set as follows: if a temporal and a geographic constraint are specified, α_t and α_g are set to 0.2, otherwise, they are set to 0. This is motivated by the intuition that the textual relevance is more important than the temporal and the geographic relevance on its own. If a document satisfies either the temporal or the geographic constraint in addition to q_{text}, it should be considered more relevant than a document not satisfying q_{text} but both the temporal and the geographic constraints. In terms of the GeoTime judgments, the former document would be considered as partially relevant while the latter document would be considered as not relevant [5].

5.3 Implementation and Indexing

Preprocessing: As described in Section 3.1, it is crucial for our model that a temporal tagger and a geo-tagger are applied in a preprocessing step to extract and normalize temporal and geographic expressions from the documents. For these two tasks, we use HeidelTime [17] and Yahoo! Placemaker [21], respectively. In addition, the Porter stemmer is applied to the documents and stop words are removed. Note that these preprocessing steps do not influence the efficiency of the query processing and that all tasks are performed on a document level and can thus be parallelized.

Query processing: The textual query is processed in the same way as the documents, i.e., using the Porter stemmer and removing stop words. For the temporal query, we assume that the intervals are specified using normalized values, e.g., "2001-11 to 2001-12". For the geographic query, we assume that query regions are formulated using their bounding box information. These assumptions are suitable for our evaluation since we translated the original descriptive GeoTime topics into q_{text}, q_{geo}, and q_{temp} as shown in Table 2 and as will be detailed in Section 5.4. Note, however, that other types of query interfaces are possible, e.g., a map interface and a time-slider, or the textual formulation of the temporal and geographic queries. In the latter case, the text query could be analyzed with the temporal and geo-taggers, and separated into q_{text}, q_{temp}, and q_{geo} automatically.

Strategy and indexes: For efficiency reasons, we run the following strategies for calculating the combined ranking score $s(q, d)$ detailed in Equation 7:

Topic	Original description, q_{text} for BL-text (underlined), q_{text} for BL-bool and proximity[2]-aware models (bold)	q_{geo} and q_{temp}
01	When and where did **Astrid Lindgren die**?	
02	When and where did **Hurricane Katrina** make **landfall** in the United States?	United States
03	When and where did **Paul Nitze die**?	
04	When and where did the **SARS epidemic begin**?	
05	When and where did **Katharine Hepburn die**?	
06	When and where did **anti-government demonstrations** occur in Uzbekistan?	Uzbekistam
07	How old was **Max Schmeling** when he **died**, and where did he die?	
08	When and where did **Chechen rebels** take Russians **hostage** in a **theatre**?	Russia
09	When and where did **Rosa Parks die**?	
10	When was the **decision** made on **siting** the **ITER** and where is it to be **built**?	
11	Describe when and where **train accidents** occurred which had **fatalities** in the period 2002 to 2005.	2002 to 2005
12	When and where did **Yasser Arafat die**?	
13	What **Portuguese colony** was **transferred** to China and when?	China
14	When and where did a **volcano erupt** in Africa during 2002?	2002 Africa
15	What American **football team won** the **Superbowl** in 2002, and where was the game played?	2002
16	When and where were the last three **Winter Olympics** held?	
17	When and where was a **candidate** for **president** of a democratic South American country **kidnapped** by a **rebel group**?	South America
18	What date was a **country** was **invaded** by the **United States** in 2002?	2002
19	When and where did the **funeral** of **Queen Elizabeth** (the Queen Mother) take place?	
20	What **country** is the most **recent** to **join** the **UN** and when did it join?	
21	When and where were the 2010 **Winter Olympics host** city location **announced**?	2010
22	When and where did a **massive earthquake** occur in December 2003?	2003-12
23	When did the largest **expansion** of the **European Union** take place, and which countries became members?	Europe
24	When and what **country** has **banned cell phones**?	
25	How long after the **Sumatra earthquake** did the **tsunami** hit Sri Lanka?	Sri Lanka

Table 2. GeoTime-1 original topic descriptions and translated queries used for the different models. For q_{geo}, we show the names of the regions for better readability although we use their bounding boxes.

- We first calculate s_{text} and calculate s_{geo}, s_{temp}, and s_{prox} only for the top-k documents of the text query. For our evaluation, we set k to 2000, and thus perform a re-ranking of the top-2000 ranked documents according to s_{text}.
- The weighted value frequency detailed in Equation 2 for expressions not satisfying q_{temp} or q_{geo} is only calculated for those documents that do not have any normalized values of temporal/geographic expressions directly satisfying q_{temp} or q_{geo}. This allows for a much more efficient temporal and geographic query processing based on the indexes described next.

The indexes used to process queries with textual, temporal, and geographic constraints efficiently, are as follows:

- For the stemmed words, a standard inverted index with term frequency information is used. Additionally, document frequency and document length information are indexed. This is already sufficient to calculate s_{text} as described in Section 4.2.
- For the normalized values of temporal expressions, we create multiple inverted indexes – namely year-, month- and day-level inverted indexes with value frequency information. Note that all fine-grained values are additionally included in the indexes for coarser granularities, e.g., an expression normalized to "1972-08-01" is listed in the day-level index, and as "1972-08" and "1972" in the month- and year-level indexes. Depending on the granularity of the query, this allows to directly determine which documents satisfy q_{temp}.
- For the normalized values of geographic expressions, we use an R-tree to index their latitude/longitude information. This allows to efficiently evaluate topological predicates (see Section 4.3.1) and to retrieve all geographic entities satisfying q_{geo}. Additionally, an inverted index is used to return all documents containing expressions referring to these geographic entities.
- For calculating the term proximities, we additionally index the position information for each term/document and value/document pair.
- For calculating the weighted value frequencies in cases when a top-k textually ranked document does not directly satisfy q_{temp} or q_{geo}, we directly access the temporal/geographic document profile of the document and iterate over all temporal/geographic expressions to calculate vw as described in Equation 2. Note that there are usually much fewer temporal and geographic expressions than regular terms in a document.

5.4 Evaluation Results

In Section 5.4, we present two baselines and the parameters used in our newly developed ranking model. Then, we detail and analyze the evaluation results in Section 5.4.

Baseline and Advanced Models We use the following two baseline models as comparison to the proximity^2-aware ranking model:

- **BL-text**: The most simple way to handle temporal and geographic information needs is to include them in the textual part of the query without treating the temporal and geographic expressions neither in the query nor in the documents in a special way. The BL-text queries are shown in Table 2 (underlined).
- **BL-bool**: Once temporal and geographic expressions are extracted and normalized as detailed in Section 3.1, and the temporal and geographic information needs are formally described as time intervals and regions, respectively, we rank the documents according to q_{text}, and use q_{temp} and q_{geo} as boolean constraints. This is a very strong baseline, which already uses the semantics of temporal and geographic expressions – a feature usually not used by standard search engines. Thus, all documents not satisfying the temporal and geographic information needs are discarded from the results. If there are no temporal or geographic constraints, documents without any temporal or geographic expression are discarded.

In addition, we use our newly introduced ranking model with different values for β, including $\beta = 0$, i.e., without using the proximity information for the final ranking. The queries used as q_{text}, q_{geo}, and q_{temp} are also given in Table 2.

Evaluation Results The following evaluation metrics are used to compare the different models with each other: precision at k (P@k), average precision at k (AP@k) and normalized discounted cumulative gain at k documents (nDCG@k). AP and nDCG have also been used to evaluate the systems of the NTCIR-8 GeoTime participants [5]. Note that some of the documents ranked top-100 by any of our used methods do not have any judgment (neither relevant nor irrelevant) from the GeoTime challenge. We set the judgment of those documents to "irrelevant" motivated by the fact that on average, there are almost 700 judgments per topic, for documents, which have been retrieved as relevant by other systems. In addition, this allows the for simpler validation and comparability of our evaluation results.

In Table 3, the evaluation results of the two baseline models and of our ranking model with different β-weights for the term proximity are presented. Independent of the evaluation metrics and the number of documents (k), the first baseline BL-text is outperformed by the second baseline BL-bool. This shows how important it is to consider the semantics of temporal and geographic expressions, i.e., to extract and normalize temporal and geographic expressions and to not consider them as regular terms.

The proximity^2-aware ranking model outperforms both baseline approaches. The best results are achieved with β set to 0.5, i.e., a medium weighting of the term proximity feature. The results demonstrate in particular that the improvements over both baselines are most remarkable when evaluating the top ranked documents ($k = 5$ and $k = 10$). Since the relevance of the top-ranked documents is most crucial for search engines, this shows the importance of taking into account the term proximity between regular terms satisfying q_{text} and expressions

method	precision (P@k)				average precision (AP@k)					nDCG@k				
	@5	@10	@20	@50	@5	@10	@20	@50	@100	@5	@10	@20	@50	@100
BL-text	44.8	42.0	36.2	29.9	35.6	34.3	27.7	25.6	23.6	45.0	44.8	44.2	46.8	47.2
BL-bool	48.0	44.0	38.6	32.7	40.0	37.6	30.7	29.6	27.7	49.1	47.8	47.2	51.0	52.1
β=0	47.2	42.8	36.6	30.1	39.2	36.8	28.8	26.5	25.6	48.1	46.6	45.1	46.7	48.7
β=0.1	49.6	44.0	39.0	31.2	42.0	38.8	32.2	29.4	28.4	50.5	48.1	47.8	49.3	51.0
β=0.3	**51.2**	45.6	41.4	**33.3**	43.3	39.1	34.3	31.6	**28.9**	52.0	49.5	50.3	52.2	**52.6**
β=0.5	**51.2**	**46.8**	**41.6**	32.9	**44.9**	**40.6**	**34.5**	**31.7**	28.8	**53.2**	**51.1**	**51.2**	**52.9**	**52.6**
β=0.7	49.6	**46.8**	41.4	32.4	43.8	40.1	33.9	30.7	27.8	51.6	50.5	50.5	51.8	51.5
β=0.9	49.6	**46.8**	41.4	32.3	43.6	40.1	34.1	30.7	27.7	51.5	50.5	50.3	51.5	50.9

Table 3. Evaluation results on all 25 NTCIR-8 GeoTime topics.

method	precision (P@k)				average precision (AP@k)					nDCG@k				
	@5	@10	@20	@50	@5	@10	@20	@50	@100	@5	@10	@20	@50	@100
BL-text	65.7	64.3	57.1	46.9	60.6	56.4	42.8	39.2	33.1	68.7	66.9	61.5	64.0	60.5
BL-bool	71.4	67.1	59.3	51.1	63.4	56.5	45.7	44.7	38.8	73.4	69.8	63.9	69.3	66.7
β=0	68.6	64.3	57.9	46.3	58.0	53.2	43.1	41.0	39.0	68.8	66.0	61.4	63.8	64.5
β=0.1	**77.1**	67.1	62.1	49.4	**66.5**	58.1	50.1	47.0	**44.2**	**76.6**	69.9	66.2	68.7	**68.8**
β=0.3	74.3	67.1	64.3	**52.6**	63.3	56.1	51.5	**49.9**	42.7	73.5	68.9	67.0	**70.9**	68.4
β=0.5	74.3	**68.6**	63.6	50.9	66.2	**59.0**	50.6	48.2	40.4	75.0	**70.7**	67.1	69.6	66.1
β=0.7	74.3	**68.6**	63.6	50.0	66.3	**59.0**	50.6	47.2	39.3	74.9	70.6	67.0	68.7	64.4
β=0.9	74.3	67.1	**65.0**	49.4	65.8	58.0	**53.3**	47.8	39.8	74.7	69.5	**67.9**	68.2	63.5

Table 4. Evaluation results on GeoTime topics with explicit geographic constraints.

satisfying q_{temp} and q_{geo} in addition to considering the semantics of temporal and geographic expressions (BL-bool).

Since the GeoTime topics are very heterogeneous, we split the results into four groups for a more detailed analysis: topics with explicit geographic constraints (Table 4), topics with explicit temporal constraints (Table 5), the topic with explicit temporal and geographic constraints (Table 6), and topics without explicit constraints (Table 7).

The differences between the results for topics with geographic and temporal constraints are huge. However, these differences are not due to our model but due to the different topic difficulty. There are many more documents judged as relevant for the topics with explicit geographic constraints than for those with explicit temporal constraints (see Table 1). All topics with explicit temporal constraints were among the most difficult topics in the data set as an analysis by the GeoTime organizers showed [5]. Despite these differences, on both topic sets, most of the observations discovered from the whole data set hold: (i) BL-bool outperforms BL-text, (ii) the proximity^2-aware model outperforms both baselines in particular for the top-ranked documents ($k = 5$, $k = 10$) and with a medium β-weight. The improvements for the top ranked documents are again in particular remarkable for the topics with explicit temporal constraints.

In Table 6, the results for topic GeoTime-0014 are presented – the only topic with explicit temporal and explicit geographic constraints. The huge differences

	precision (P@k)				average precision (AP@k)					nDCG@k				
method	@5	@10	@20	@50	@5	@10	@20	@50	@100	@5	@10	@20	@50	@100
BL-text	20.0	16.0	14.0	11.2	6.9	5.7	4.2	4.1	3.8	15.0	13.8	14.2	16.5	16.6
BL-bool	12.0	12.0	**20.0**	16.4	6.0	6.4	11.4	**12.2**	**11.4**	11.5	11.9	18.6	22.9	23.3
β=0	12.0	10.0	12.0	12.0	9.1	7.3	5.8	4.5	4.1	12.8	11.3	12.1	12.0	14.3
β=0.1	12.0	12.0	14.0	13.6	9.1	8.5	6.1	5.5	5.3	12.8	12.6	14.0	16.4	17.3
β=0.3	20.0	18.0	18.0	16.0	15.1	12.2	10.3	9.3	9.1	21.4	19.5	21.5	23.2	25.2
β=0.5	20.0	22.0	**20.0**	17.2	16.7	14.5	**11.6**	10.7	9.7	22.1	22.9	**23.9**	27.3	26.4
β=0.7	**24.0**	22.0	19.0	17.2	19.9	15.0	11.1	10.8	9.9	24.7	23.1	23.3	27.4	26.6
β=0.9	**24.0**	**26.0**	19.0	**17.6**	**20.7**	**17.1**	9.7	10.4	9.3	**25.0**	**25.8**	23.5	**27.7**	**26.7**

Table 5. Evaluation results on GeoTime topics with explicit temporal constraints.

	precision (P@k)				average precision (AP@k)					nDCG@k				
method	@5	@10	@20	@50	@5	@10	@20	@50	@100	@5	@10	@20	@50	@100
BL-text	20.0	40.0	35.0	18.0	4.0	14.6	10.0	6.0	7.2	13.1	30.6	30.2	27.2	32.3
BL-bool	**100**	**100**	55.0	**36.0**	**100**	**100**	31.5	24.6	30.3	**100**	**100**	68.4	66.2	**77.1**
β=0	**100**	90.0	50.0	22.0	**100**	90.0	27.5	13.9	13.7	**100**	93.6	64.1	50.4	52.0
β=0.1	**100**	**100**	55.0	24.0	**100**	**100**	32.7	15.9	17.5	**100**	**100**	68.5	53.5	60.3
β=0.3	**100**	**100**	**75.0**	34.0	**100**	**100**	55.3	31.0	30.8	**100**	**100**	82.2	66.4	68.3
β=0.5	**100**	**100**	**75.0**	**36.0**	**100**	**100**	**55.7**	**33.2**	**31.7**	**100**	**100**	**82.3**	**68.6**	68.6
β=0.7	**100**	**100**	**75.0**	**36.0**	**100**	**100**	54.8	32.5	31.1	**100**	**100**	**82.3**	68.5	68.5
β=0.9	**100**	**100**	**75.0**	**36.0**	**100**	**100**	54.5	32.2	30.7	**100**	**100**	82.2	68.5	68.5

Table 6. Evaluation results on the GeoTime topic with explicit temporal and geographic constraints.

	precision (P@k)				average precision (AP@k)					nDCG@k				
method	@5	@10	@20	@50	@5	@10	@20	@50	@100	@5	@10	@20	@50	@100
BL-text	45.0	**40.0**	33.3	28.8	35.6	35.0	30.1	28.3	27.7	46.4	46.1	47.8	50.9	53.5
BL-bool	45.0	39.2	32.9	28.5	35.6	34.5	30.0	28.4	27.9	46.4	45.6	47.5	50.7	53.4
β=0	45.0	**40.0**	33.3	28.8	35.6	35.0	30.1	28.3	27.7	46.4	46.1	47.8	50.9	53.5
β=0.1	45.0	39.2	34.6	28.5	36.6	35.0	32.6	30.1	29.6	47.0	45.9	49.3	51.4	53.8
β=0.3	**46.7**	**40.0**	**35.0**	**29.2**	38.6	35.3	32.5	30.2	29.0	48.2	46.5	49.9	52.2	53.4
β=0.5	**46.7**	**40.0**	**35.0**	28.7	**39.6**	**35.8**	**33.0**	**30.7**	**29.8**	**49.4**	**47.3**	**50.8**	**52.6**	**54.2**
β=0.7	41.7	**40.0**	**35.0**	28.2	35.9	34.6	31.9	29.3	28.4	45.1	46.1	49.5	50.8	52.9
β=0.9	41.7	39.2	34.2	28.2	35.6	34.3	31.3	29.1	28.0	45.0	45.5	48.5	50.2	52.1

Table 7. Evaluation results on the GeoTime topics without explicit temporal or geographic constraints.

between the two baselines show how important it is to include the semantics of temporal and geographic expressions into the ranking model when dealing with explicit temporal and geographic constraints. While the baseline with boolean constraints already achieves very good results for top ranked documents, the results for more documents ($k \geq 20$) can further be improved by integrating proximity information into the model. The best evaluation results are again achieved with a medium β-weighting of 0.5.

After having demonstrated the value of our model when being faced with explicit temporal and/or geographic expressions, we finally analyze if proximity information also helps to improve the retrieval quality when being faced with implicit temporal and geographic constraints only ("when" and "where"). As Table 7 shows, our ranking model with $\beta = 0.5$ achieves the best results on this subset and outperforms the baselines. Thus, we can summarize the evaluation results with the following three main observations:

- Considering the semantics of temporal and geographic expressions helps to improve satisfying information needs with explicit and/or implicit temporal and geographic constraints.
- Taking into account term proximity information between regular terms satisfying a text query and expressions satisfying temporal and geographic queries helps to further improve ranking results.
- In particular, the top-ranked documents benefit from the term proximity information.

When analyzing the search results after the evaluation, we made the following observation: Since all documents are from the New York Times corpus and thus news documents, the document creation time (dct) plays an important role throughout the whole text of the documents (cf. [17]). Thus, if the dct satisfies q_{temp}, the temporal constraint could be softened for the term proximity calculation since the dct is latently available in the whole news article. However, although we evaluated our newly developed model based on a dataset with news articles, we did not develop our model for news documents only.

6 Conclusions and Ongoing Work

In this paper, we presented a new model to rank documents for a combination of textual, temporal, and geographic information needs. In addition to calculating single scores for each part, we developed a proximity measure to determine the distance in the document between terms and expressions satisfying different query parts. Thus, our model eliminates the counter-intuitive assumption of independence between the query components, which distinguishes our model from previous works. As our evaluation demonstrates, the extraction and normalization of temporal and geographic expressions is a prerequisite to satisfy temporal and geographic information needs. Taking into account the proximity improves ranking results even further.

Currently, we are applying our model to identify temporal and geographic highlights, i.e., to answer queries that contain temporal and geographic constraints but no textual parts. This allows to determine documents that are most relevant for time-region combinations such as “Munich 1972”. This approach could be the basis of a geo-temporal event search engine.

Acknowledgements

We thank Fredric Gey and Ray Larson from UC Berkeley for sharing the NTCIR Geo-Time data.

References

1. O. Alonso, J. Strötgen, R. Baeza-Yates, and M. Gertz. Temporal Information Retrieval: Challenges and Opportunities. In *TWAW'11*, pages 1–8, 2011.
2. K. Berberich, S. J. Bedathur, O. Alonso, and G. Weikum. A Language Modeling Approach for Temporal Information Needs. In *ECIR'10*, pages 13–25, 2010.
3. W. Dakka, L. Gravano, and P. G. Ipeirotis. Answering General Time Sensitive Queries. In *CIKM'08*, pages 1437–1438, 2008.
4. F. Gey, N. Kando, and R. Larson. The Crucial Role of Semantic Discovery and Markup in Geo-temporal Search. In *ESAIR '10*, pages 5–6, 2010.
5. F. Gey, R. Larson, N. Kando, J. Machado, and T. Sakai. NTCIR-GeoTime Overview: Evaluating Geographic and Temporal Search. In *NTCIR-8*, 2010.
6. F. Gey, R. Larson, J. Machado, and M. Yoshioka. NTCIR9-GeoTime Overview - Evaluating Geographic and Temporal Search: Round 2. In *NTCIR-9*, 2011.
7. X. Li and W. B. Croft. Time-based Language Models. In *CIKM'03*, pages 469–475, 2003.
8. M. D. Lieberman, H. Samet, J. Sankaranarayanan, and J. Sperling. STEWARD: Architecture of a Spatio-textual Search Engine. In *GIS '07*, pages 186–193, 2007.
9. J. Machado, B. Martins, and J. Borbinha. LGTE: Lucene Extensions for Geo-Temporal Information Retrieval. In *GIIW'09*, 2009.
10. C. D. Manning, P. Raghavan, and H. Schütze. *Introduction to Information Retrieval*. Cambridge University Press, 2008.
11. S. Nunes, C. Ribeiro, and G. David. Use of Temporal Expressions in Web Search. In *ECIR'08*, pages 580–584, 2008.
12. J. Pustejovsky, J. M. Castaño, R. Ingria, R. Sauri, R. J. Gaizauskas, A. Setzer, G. Katz, and D. R. Radev. TimeML: Robust Specification of Event and Temporal Expressions in Text. In *New Directions in Question Answering*, pages 28–34, 2003.
13. R. S. Purves et al. The Design and Implementation of SPIRIT: a Spatially Aware Search Engine for Information Retrieval on the Internet. *International Journal of Geographical Information Science*, 21(7):717–745, 2007.
14. Y. Rasolofo and J. Savoy. Term Proximity Scoring for Keyword-based Retrieval Systems. In *ECIR'03*, pages 207–218, 2003.
15. S. E. Robertson, S. Walker, S. Jones, M. Hancock-Beaulieu, and M. Gatford. Okapi at TREC-3. In *TREC-3*, 1994.
16. J. Strötgen and M. Gertz. Event-centric Search and Exploration in Document Collections. In *JCDL'12*, pages 223–232, 2012.

17. J. Strötgen and M. Gertz. Multilingual and Cross-domain Temporal Tagging. *Language Resources and Evaluation*, 47(2):269–298, 2013.
18. J. Strötgen and M. Gertz. Proximity^2-aware Ranking for Textual, Temporal, and Geographic Queries. In *CIKM'13*, pages 739–744, 2013.
19. T. Tao and C. Zhai. An exploration of proximity measures in information retrieval. In *SIGIR'07*, pages 295–302, 2007.
20. TExt Retrival Conference (TREC). http://trec.nist.gov.
21. Yahoo! Placemaker. http://developer.yahoo.com/geo/placemaker/.
22. W. V. Zhang, B. Rey, E. Stipp, and R. Jones. Geomodification in Query Rewriting. In *GIR'06*, 2006.