# Experiments with Semantic-flavored Query Reformulation of Geo-Temporal Queries

Nuno Cardoso
University of Lisbon
Faculty of Sciences, LaSIGE
ncardoso@xldb.di.fc.ul.pt

Mário J. Silva
University of Lisbon
Faculty of Sciences, LaSIGE
mjs@di.fc.ul.pt

## ABSTRACT

We present our participation in the NTCIR GeoTime evaluation task with a semantically-flavored geographic information retrieval system. Our approach relies on a thorough interpretation of the user intent by recognising and grounding entities and relationships from query terms, extracting additional information using external knowledge resources and geographic ontologies, and reformulating the query with reasoned answers. Our experiments aimed to observe the impact of semantic-based reformulated queries on the retrieval performance.

## Categories and Subject Descriptors

H.3 [**Information Storage and Retrieval**]: H.3.3 Information Search and Retrieval

## General Terms

Algorithms, Design, Evaluation

## Keywords

Geographical Information Retrieval, Query Reformulation, Information Extraction

## 1. INTRODUCTION

As user information needs become more elaborate and context-aware, classic information retrieval (IR) approaches show significative limitations on returning relevant documents, as term-statistic approaches focus on what the user said (given by the query terms), not on what the user wanted (given by the information need expressed in query terms).

Users typically describe their simple information needs to IR systems using short queries. For more complex information needs involving entities such as places, organisations or persons, relationships between them such as "located in" or "published by," the user might formulate elaborated queries that can resemble more as a question rather than a list of keywords. Our belief is that, if a user is interested for instance in knowing "which Swedish writers died in Stockholm?," he shouldn't have to state his information need in short queries that old classic IR systems handle, but rather state it in natural language. The IR system should take the burden of understanding his intentions by taking advantage of the semantic content in the query, and reasoning through a result list that indeed matches what the user needs.

We are developing a geographic information retrieval (GIR) system that handles complex queries, specifically queries that contain geographic criteria that define a geographic area of interest, such as "restaurants in Stockholm." Its query reformulation module does not work directly with terms, but with the *entities* represented by those terms. External knowledge resources, such as Wikipedia, DBpedia and geographic ontologies, are used to extract information about entities, their properties and relationships among them, and find answers matching the user information need. The initial query is afterwards reformulated using the extracted information, and submitted to the retrieval module.

The GikiP pilot task in 2008 [12], and the GikiCLEF track in 2009 [11], focused precisely on the reasoning step needed for such demanding queries, rather than the retrieval step. The task proposed to participants was addressing geographically challenging topics by using Wikipedia as an information resource, and returning answers given by Wikipedia pages. Our participation on both evaluation tasks helped to shape the semantic approaches within the GIR system described in this paper.

The NTCIR GeoTime task presents an ad-hoc IR evaluation task where topics are elaborated as questions with a strong geographic and temporal bias [5]. The task challenges participants to develop a system with robust retrieval *and* reasoning capabilities. We participated in the NTCIR GeoTemporal task with the goal of measuring the impact on the retrieval performance using reformulated queries generated by our semantic-based query reformulation module, compared to simple, short query strings.

The rest of this paper is organised as follows: Section 2 overviews our GIR system. Section 3 describes our experiments and submitted runs. Section 4 presents and analyses our results, and Section 5 concludes the paper.

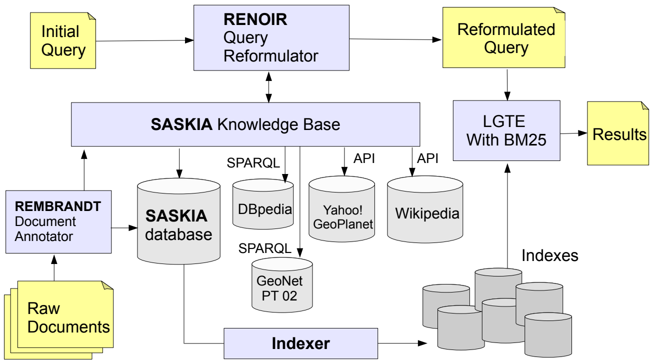


**Figure 1: Overview of the GIR architecture**

## 2. SYSTEM DESCRIPTION

Figure 1 presents the architecture of our GIR system. There are five main modules: i) a semantic query reformulation module, RENOIR, handles and reformulates user queries; ii) a document annotator tool, REMBRANDT [2], recognises and grounds all entities from documents; iii) a knowledge base, SASKIA, is the access point for all knowledge resources, iv) an indexer, which generates a standard term index and selective indexes for each entity type, and v) a retrieval engine, LGTE (Lucene with GeoTemporal Extensions) [8], which retrieves and ranks results.

As knowledge resources, we use a local copy of the English and Portuguese Wikipedia snapshots (article texts and SQL dumps), a local copy of the DBpedia dataset [1] and the geographic ontologies GeoNetPT-02 [7] (for the Portuguese territory) and the Yahoo!'s GeoPlanet™web-service [14]. DBpedia and GeoNetPT-02 are both loaded to an OpenLink Virtuoso triple-store server, which provides a SPARQL query interface.

### 2.1 Query parsing

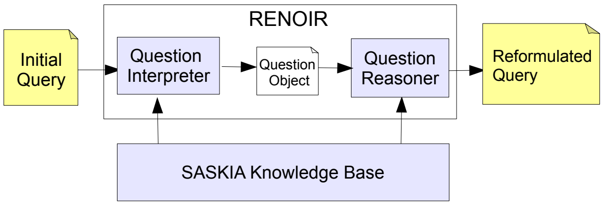


**Figure 2: RENOIR query reformulation module**

The NTCIR GeoTime topics are handled by RENOIR, which is detailed in Figure 2. The initial task of RENOIR is performed by a question interpreter (QI), which recognises entities and expressions from the topic title, and grounds them using unique identifiers, such as DBpedia resource URLs ands GeoPlanet™WOEIDs – Where On Earth IDs [15]. With this information, the QI generates a *question object* which is composed of the following attributes:

**Subject,** grounding information for the expression in the user query that represents the expected answer type, as for example "Swedish writers." The subject can be represented by i) DBpedia resources that have a property `rdf:type` for a value `skos:Concept` [9] as in `http://dbpedia.org/resource/Category:Swedish_writers`, ii) a DBpedia ontology class as in `http://dbpedia.org/ontology/Writer`, or iii) a semantic classification such as `PERSON/INDIVIDUAL`, as defined in the HAREM categorization [13], in this preferential order.

**Expected Answer Type (EAT),** list of properties that the final set of answers must have.

**Conditions,** list of filtering criteria on the subject, such as a geographic scope or a temporal expression (for example, "died in Stockholm" or "born in 2002"). A condition may contain i) a DBpedia ontology property, as in http://dbpedia.org/ontology/deathPlace, ii) an operator such as `BEFORE` or `BETWEEN`, and iii) a referent object, which may be represented by a grounded entity (such as `http://dbpedia.org/resource/Stockholm`), a generic named entity (such as the year 2002) or subject (as in "Cities of Sweden," grounded to `http://dbpedia.org/resource/Category:Cities_in_Sweden`).

For the example question "Which Swedish writers died in Stockholm?," the question interpreter would start with a first set of pattern rules that detects "Swedish writers" as a subject and grounds it to the DBpedia resource `http://dbpedia.org/resource/Category:Swedish_writers`, which is derived from the corresponding Wikipedia's category page.

Another set of rules to detect question type matches the "Which <subject>" pattern and assigns the subject to the EAT, that is, the answers must contain the property `Category:Swedish_writers`. In a different scenario where the question started with the "Where and when" pattern, as in a significative amount of GeoTime topic titles, the EAT was then assigned to the generic HAREM categories `LOCAL` and `TIME`, not to a subject.

Finally, a set of pattern rules interprets the expression "died in Stockholm" into a new condition, which contains the DBpedia property `http://dbpedia.org/ontology/deathPlace` and a referent entity `http://dbpedia.org/resource/Stockholm`.

The final step of RENOIR is the question reasoner (QR), which aims to resolve the question (given by the question object) into a list of answers. Depending on the elements present in the question object, the QR decides the best strategy to obtain those answers, which may involve a list of SPARQL queries to the SASKIA knowledge base.

In the given example, for a question object with an EAT grounded to `http://dbpedia.org/resource/Category:Swedish_writers`, a single condition described by a property `dbpedia-owl:deathPlace` and a referent geographic entity `http://dbpedia.org/resource/Stockholm`, the QR module issues the following SPARQL query to the DBpedia:

```
SELECT DISTINCT ?swedishWriters WHERE {
 { ?swedishWriters skos:subject
   <http://dbpedia.org/resource/Category:Swedish_writers>
 } UNION
 {  ?swedishWriters skos:subject
    ?category.
    ?category skos:broader
   <http://dbpedia.org/resource/Category:Swedish_writers>
 }
 ?swedishWriters dbpedia-owl:deathPlace
 <http://dbpedia.org/resource/Stockholm>
}
```

With the DBpedia v3.5.1 dataset, this SPARQL query returns 21 DBpedia resources, including `http://dbpedia.org/resource/Astrid_Lindgren`.

## 2.2 Document parsing

REMBRANDT (`http://xldb.di.fc.ul.pt/Rembrandt/`) is a named-entity recognition software which is used to annotate documents by classifying named entities (NEs) and assigning them identifiers composed by Wikipedia and DBpedia URLs. Its classification strategy begins by mapping NEs to their corresponding Wikipedia and DBpedia pages, using DBpedia's ontology classes and Wikipedia categories to infer the semantic classification. Then, REMBRANDT applies a set of manually generated language-dependent rules, which represent the internal and external evidence for NEs for a given language, as in "*city of X*" or "*X, Inc.*" This set of rules disambiguates NEs with more than one semantic classification, classifies NEs that were not mapped to a Wikipedia or DBpedia page. Figure 3 shows a text excerpt tagged by REMBRANDT.

```
<NE ID="204" S="59" T="24" C1="PERSON" C2="INDIVIDUAL"
WK="Astrid_Lindgren" DB="Astrid_Lindgren" >Astrid Lind-
gren</NE>, the Swedish writer whose rollicking, anarchic
books about <NE ID="206" S="59" T="36" C1="MARTERPIECE"
C2="REPRODUCED" WK="Pippi_Longstocking" >Pippi Long-
stocking</NE> horrified a generation of parents and
captivated <NE ID="207" S="59" T="45" C1="NUMBER"
C2="TEXTUAL" >millions</NE> of children around the
globe, died in her sleep <NE ID="208" S="59" T="56"
C1="DATETIME" C2="CALENDAR" C3="DATE" TG="?:" >Mon-
day</EM> at her home in <NE ID="209" S="59" T="61"
C1="LOCAL" C2="HUMAN" WK="Stockholm" DB="Stockholm"
>Stockholm</NE>, <NE ID="210" S="59" T="63" C1="LOCAL"
C2="HUMAN" C3="COUNTRY" WK="Sweden" DB="Sweden" >Sweden
</NE>.  She was <NE ID="211" S="60" T="2" C1="NUMBER"
C2="NUMERAL" >94</NE>.
```

**Figure 3: Excerpt of a document tagged by REMBRANDT, `NYT_ENG_20020128.0134`.**

REMBRANDT also generates *geographic signatures* and *temporal signatures* for tagged documents. A geographic signature of a document is the surrogate of the NEs found in the document that were grounded as a geographic place, where each place is expanded upwards to the country level, following the strategy proposed by Li et al. [6].

Figure 4 presents an excerpt of a geographic signature of a document. The geographic signature is composed by a list of `<Place>` elements, which has a `count` attribute that stores the document frequency of that place, and a WOEID (plus a GeoNetPT-02 ID if the place is within the Portuguese territory) as an identifier. Each `<Place>` element contains the different NEs used in the document to designate it (`<NE>`), the place's entity name (`<Entity>`), HAREM's semantic classifications under the `LOCAL` category (`<Type>` and `<Subtype>`), the DBpedia class (`<DBpediaClass>`), the ancestor's WOEIDs and entity name (`<Ancestor>`), and centroid/bounding box information given by GeoPlanet.

```
<GeoSignature version="1.0" totalcount="25">
<Doc id="555397" original_id="NYT_ENG_20020128.0134"
 lang="en" />
(...)
  <Place count="1" woeid="906057">
    <NE id="1059384">Stockholm</NE>
    <Name>Stockholm</Name>
    <Type>@HUMANO</Type>
    <DBpediaClass>Area</DBpediaClass>
    <Ancestor woeid="12587478">Stockholm Kommun</Ancestor>
    <Ancestor woeid="2347067">Stockholm</Ancestor>
    <Ancestor woeid="23424954">Sweden</Ancestor>
    <Centroid>
       <Latitude>59.332169</Latitude>
       <Longitude>18.062429</Longitude>
    </Centroid>
    <BoundingBox>
      <southWest>
       <Latitude>58.877621</Latitude>
       <Longitude>17.171261</Longitude>
      </southWest>
      <northEast>
       <Latitude>59.786720</Latitude>
       <Longitude>18.953581</Longitude>
      </northEast>
    </BoundingBox>
  </Place>
(...)
</GeoSignature>
```

**Figure 4: Excerpt of the geographic signature generated for the document shown in Figure 3.**

```
<TimeSignature version="1.0" totalcount="16" >
  <Doc id="555397" original_id="NYT_ENG_20020128.0134"
 lang="en" />
  <DocDateCreated>20020128</DocDateCreated>
 <Time count="1">
   <NE id="3648" lang="en">28th January, 2002</NE>
   <TimeGrounding>!:+Y2002M01D28</TimeGrounding>
   <Index type="date">20020128</Index>
 </Time>
 <Time count="1">
   <NE id="3652" lang="en">29th January</NE>
   <TimeGrounding>!:+M01D29</TimeGrounding>
   <Index type="date">20020129</Index>
 </Time>
</TimeSignature>
```

**Figure 5: Excerpt of the temporal signature generated for the document shown in Figure 3.**

Likewise, a temporal signature of a document is the surrogate of the NEs of category `DATETIME` found in the document that were grounded into a temporal expression. Figure 5 illustrates a temporal signature of a document generated by REMBRANDT. The temporal signature starts with a `<DocDateCreated>` element that contains the creation date of the document (inferred by the published date given in the NYT collection). Each distinct NE is represented in a `<Time>` element, with a `count` attribute storing its document frequency. The `<NE>` element contains the NE terms. The `<TimeGrounding>` element contains the temporal expression that grounds the NE in an easier format to process. The `<Index>` element rounds the temporal expressions to a day threshold, and if necessary fills the missing date elements by looking at the surrounding time expressions and selecting the most popular values. For example, in the time expression "29th January" from the time signature of Figure 5, the `<Index>` field includes the year 2002, which was inferred from the other time expressions in the document which explicitly refer to a year.

All tagged documents and associated geographic and temporal signatures are stored in the SASKIA knowledge base. SASKIA is the access point to the information extracted from the document collection, and SASKIA's API also facilitates access to external knowledge resources, such as Wikipedia and DBpedia.

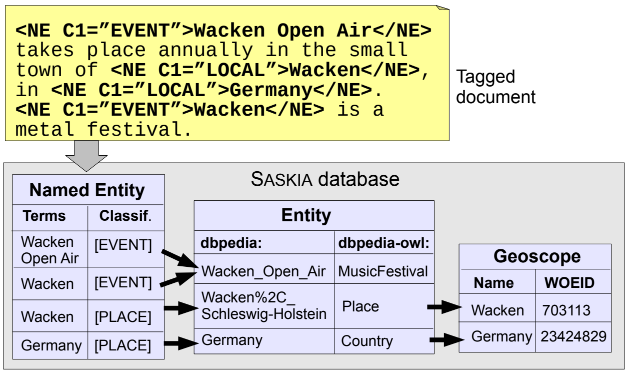


**Figure 6: Filling the SASKIA database with semantic information from tagged documents**

Figure 6 illustrates how SASKIA stores the semantic information from the documents tagged by REMBRANDT. Each NE is stored in a NE table, together with its terms and semantic classification. If the NE was successfully grounded to a DBpedia resource, that information is stored in an ENTITY table. SASKIA's database also has a GEOSCOPE table, which encompasses all grounded information for entities that are geographic places, such as centroids, bounding boxes, ancestors or neighbors, as well as WOEIDs and GeoNetPT-02 identifiers. This document post-processing step is required for the index generation, described in the following section.

The NE/ENTITY separation is needed, as a given entity can be referred to by several NEs (for example, the entity `http://dbpedia.org/resource/Wacken_Open_Air` can be represented by the NEs "Wacken" and "Wacken Open Air"). Likewise, the NE table distinguishes between NEs with the same terms and/or same semantic classification, but with different grounded entities. For example the NE "Wacken" may refer to a place or an event (the disambiguation was previously made by REMBRANDT). Moreover, RENOIR can therefore have access to a list of different NE representations for a given entity, which can be useful in the query reformulation step, to circumvent the different designations given to certain entities by different authors in the document collection.

## 2.3 Index Generation and Document Retrieval

The retrieval and ranking step is performed by LGTE, a version of Lucene with extensions for geographic and temporal ranking measures. LGTE implements the BM25 weighting scheme [10]. In our GIR system, LGTE uses three kinds of indexes: i) an index of document terms, ii) an index for NEs, and iii) an index of document signatures.

The index for document terms is a standard inverted index used in classic text retrieval, providing a scheme for reverting to classic IR when RENOIR fails to reformulate a query. The index for NEs has all terms assigned with a semantic classification. For instance, if a document contains the NE "Lindgren" grounded to the entity `http://dbpedia.org/resource/Astrid_Lindgren`, an index for NEs of category `PERSON` includes both the original terms "Lindgren", and the terms "Astrid Lindgren" extracted from the `rdfs:label` DBpedia property of the entity. Lastly, the index for document signatures is the base for LGTE's geographic and temporal similarity rankings. The geographic signature index uses the WOEIDs from each place's entity and ancestors, as illustrated below for a document that refers to the Swedish cities of Stockholm and Gotenburg (the entity name associated to the WOEID is included for clarity):

| Entity | WOEID | DocID | <DocFreq> |
|---|---|---|---|
| Stockholm | 906057 | 555397 | <1> |
| Stockholm Kommu | 12587478 | 555397 | <1> |
| Stockholm | 2347067 | 555397 | <1> |
| Sweden | 23424954 | 555397 | <1> |
| Gothenburg | 890869 | 555397 | <1> |
| Gothenburg | 12587233 | 555397 | <1> |
| Västra Götaland | 20070562 | 555397 | <1> |
| Sweden | 23424954 | 555397 | <1> |

With the geographic signature index, LGTE can generate a simple geographic similarity measure for queries with a geographic scope. For example, if a query geoscope is "Sweden", then all documents that refer to cities in Sweden will provide a partial score with the geographic signature index, since they all have a reference to Sweden's WOEID, 23424954. In the above case, since there are two references to Sweden's WOEID, its document frequency of 2 will generate a higher score.

The temporal signature index has a similar format. For the example in Figure 5, the index entries are:

| Time | DocID | <DocFreq> |
|---|---|---|
| 20020128 | 555397 | <1> |
| 20020129 | 555397 | <1> |

When the temporal signature index doesn't have any grounded temporal expressions, the document creation date is used as the only entry. To compute temporal similarity measures, LGTE can take advantage of the year-month-day order and use wildcards. For instance, a query with a temporal restriction of "January 2002" is grounded by RENOIR and later reformulated as **time-index**:`200201*`, which encompasses all documents with references to a date within that month, including `20020128` and `20020129`.

The retrieval step is invoked when RENOIR issues a reformulated query similar to:

**term**: swedish writers died stockholm "Astrid Lindgren" "Lina Sandell" "Nelly Sachs" "Eyvind Johnson" "Stieg Larsson"
**ne-PERSON-index**: "Astrid Lindgren" "Lina Sandell" "Nelly Sachs" "Eyvind Johnson" "Stieg Larsson"
**ne-LOCAL-index**: Stockholm
**woeid-index**: 906057
**time-index:**2002*

where each term/entity is assigned to a target index. Entities are enclosed in double quotes on all indexes, so that LGTE matches documents that have those terms in that order. In the **term** field, we have query terms targeted for the inverted term index, composed by original query terms and entity terms. In the **ne-PERSON-index** field, we have NEs of type PERSON that were considered by RENOIR as correct answers to the question, and their terms are targeted for the NE PERSON index. In a similar way, the term "Stockholm" is targeted for the **ne-LOCAL-index** field. Finally, as the query has a scope given by the entity "Stockholm," its WOEID will be used in the geographic signature index, **woeid-index**.

In the retrieval process, LGTE returns a single list of results ordered by the sum of the BM25 scores given by multiple indexes. It will be possible to tune the index weights for each retrieval according to each reformulated query, but this functionality hasn't been implemented.

# 3. EXPERIMENTS AND RUNS

Almost all topics consist of a question about the place and time of a certain event, and included at least one entity – person, place, sport events, diseases and hurricanes, a similar distribution to the EAT of the GikiCLEF topics [3]. For instance, 15 out of 25 start with "When and Where," such as topic # 1, "When and where did Astrid Lindgren die?," and such events are denoted using indirect references to other entries ("Astrid Lindgren") and actions ("die").

Our participation had the following goals:

- To observe RENOIR's performance on reasoning towards providing answers for the proposed topics;
- To evaluate how well RENOIR and LGTE are integrated for this reasoning and retrieval task;
- To detect failure points and limitations on the proposed approach.

**Table 1: Submitted runs**

| Run | Description |
|---|---|
| #1 | baseline, plain terms with no expansion |
| #2 | automatic run, with DBpedia ontology lookup |
| #3 | supervised run, with DBpedia ontology lookup |
| #4 | extended run, with DBpedia abstract entities |

Table 1 lists the four runs submitted. The difference between the runs is only in the reformulated queries used, as the LGTE parameter configuration was the same in all runs. Also, we only used the title string of the topics, as we believe it is a more realistic scenario in a retrieval environment with real users.

## 3.1 baseline run

This run is generated from queries with no reformulation, only stop-word removal (including the question words). For instance, the query string used for topic #1 is "astrid lindgren die." This run serves as a baseline, and represents the performance of a standard text retrieval system with a BM25 weighting scheme.

## 3.2 automatic run

This run is generated with reformulated queries from RENOIR, without human intervention. For instance, the query string for the topic #1 was:

```
astrid lindgren die ne-PERSON-INDIVIDUAL-index:"Astrid
Lindgren" ne-LOCAL-HUMAN-DIVISION-index:Stockholm ne-
LOCAL-HUMAN-DIVISION-index:Sweden woeid-index:23424954
woeid-index:906057 tg-index:20020128
```

As described in Section 2.1, the "Where and when" pattern grounds the EAT to the generic HAREM classification of `LOCAL` and `DATETIME`. The entity "Astrid Lindgren" is detected and grounded to `http://dbpedia.org/resource/Astrid_Lindgren`, and the expression "die" is mapped to the DBpedia ontology properties `http://dbpedia.org/property/deathPlace` and `dbpedia-owl:deathDate`. In the QR step, the morphology of this question object (a condition with a single NE classified as a `PERSON` and a mapped DBpedia property, and no subjects as EAT, just two generic HAREM categories) dictates that the best strategy is to query DBpedia for properties associated to the entity `http://dbpedia.org/resource/Astrid_Lindgren`, with the following SPARQL query:

```
SELECT ?place, ?time where {
  <http://dbpedia.org/resource/Astrid_Lindgren>
  <http://dbpedia.org/ontology/deathPlace> ?place .
  <http://dbpedia.org/resource/Astrid_Lindgren>
  <http://dbpedia.org/ontology/deathDate> ?time .
}
```

With the DBpedia 3.5.1 dataset, this SPARQL query returns two answers: `http://dbpedia.org/resource/Stockholm` and `2002-01-28^^Date`. As the answers match the expected answer type, they are considered correct answers. In the automatic run, 21 out of 25 topics had reformulated queries that are different from the baseline queries.

## 3.3 Supervised run

The supervised run is generated from revised queries of the automatic run. Some of the question interpretations and entity groundings were manually corrected, some geographic expansions (listing African countries in the topic #14) where completed and minor adjustments were executed. Changes were made to a total of 6 queries.

## 3.4 Extended run

The extended run is an experiment on an alternative plan when Renoir can't generate answers for the question. For example, topic #2 asks "When and where did hurricane Katrina make landfall in the United States?;" while Renoir is able to ground the entities "United States" and "hurricane Katrina" and define the EAT, it can not map the expression "make landfall in the" to a DBpedia property.

To compensate for this problem, we experimented with a simple heuristic of adding NEs that are strongly related to the entities found in the topic title, and that match the EAT. We use Rembrandt to tag the `http://dbpedia.org/property/abstract` property. This property contains an abstract for each DBpedia resource, taken from the first paragraphs of the corresponding Wikipedia document. We collect the NEs that match the EAT, in this case all places and temporal expressions, and reformulate the query by including NEs targeted to the geoscope, temporal and NE indexes, but not to the term index, such as "Atlantic Ocean" or "Bahamas." Changes were made to a total of 14 queries.

The reformulated query for topic #2 used in the extended run is the following:

```
 hurricane katrina make landfall united states ne-
LOCAL-HUMAN-COUNTRY-index:"United States" ne-LOCAL-
PHYSICAL-WATERMASS-index:"Atlantic Ocean" ne-LOCAL-
HUMAN-COUNTRY-index:"Bahamas" ne-LOCAL-PHYSICAL-ISLAND-
index:"Bahamas" ne-LOCAL-HUMAN-DIVISION-index:Florida
ne-LOCAL-HUMAN-DIVISION-index:Louisiana ne-LOCAL-
PHYSICAL-REGION-index:Gulf ne-LOCAL-HUMAN-DIVISION-
index:Texas ne-LOCAL-HUMAN-DIVISION-index:"New Orleans"
ne-EVENT-index:"Hurricane Katrina" woeid-index:23424977
woeid-index:55959709 woeid-index:23424758 woeid-
index:55959686 woeid-index:2347577 woeid-index:2347602
woeid-index:615134 tg-index:20050830 tg-index:20050823
```

# 4. ANALYSIS OF THE RESULTS

Table 2a) summarizes the official results for the submitted runs, and Table 2b) details the MAP values per topic of each run. In Table 2b), the superscript in front of the topic number signals the runs that were generated with a different query compared to the previous run; for instance, for the topic nr. 1, there was a different query from run #1 to run #2, and from run #3 to run # 4.

We observe that our best run was the automatic run (#2) and the run with the lowest MAP score was the extended run (#4). The overall MAP values between runs #1, #2 and #3 don't look significantly different, which lead us to conclude that the reformulated queries did not had a significative impact on retrieval process.

**Table 2: a) mean average precision (AP), Q-measure and normalized discounted cumulative gain (nDCG), and b) MAP values for each topic**

| a) Run | mean AP | mean Q | nDCG |
|---|---|---|---|
| Run 2 | **0.3354** | 0.3584 | **0.5705** |
| Run 1 | 0.3301 | **0.5701** | 0.5701 |
| Run 3 | 0.3255 | 0.3482 | 0.5593 |
| Run 4 | 0.2978 | 0.3205 | 0.5325 |

| b) Topic | Run 1 | Run 2 | Run 3 | Run 4 |
|---|---|---|---|---|
| $\mathbf{1}^{24}$ | 0.6791 | 0.6821 | 0.6821 | 0.6821 |
| $\mathbf{2}^{234}$ | 0.3627 | 0.3671 | 0.3671 | 0.1720 |
| $\mathbf{3}^{24}$ | 0.7833 | 0.7833 | 0.7833 | 0.8242 |
| $\mathbf{4}^{24}$ | 0.2281 | 0.2281 | 0.2281 | 0.2926 |
| $\mathbf{5}^{234}$ | 0.1965 | 0.1821 | 0.1821 | 0.1821 |
| $\mathbf{6}^{2}$ | 0.4895 | 0.4837 | 0.4837 | 0.4837 |
| $\mathbf{7}^{24}$ | 0.7291 | 0.8041 | 0.8041 | 0.7761 |
| $\mathbf{8}^{234}$ | 0.5372 | 0.4892 | 0.4642 | 0.3954 |
| $\mathbf{9}^{24}$ | 0.3824 | 0.4379 | 0.4379 | 0.4328 |
| $\mathbf{10}^{2}$ | 0.1400 | 0.1500 | 0.1500 | 0.1500 |
| $\mathbf{11}^{2}$ | 0.1178 | 0.1057 | 0.1057 | 0.1057 |
| $\mathbf{12}^{234}$ | 0.0877 | 0.1259 | 0.1330 | 0.1330 |
| $\mathbf{13}^{24}$ | 0.4290 | 0.2400 | 0.2400 | 0.2596 |
| $\mathbf{14}^{23}$ | 0.3066 | 0.2931 | 0.0630 | 0.0630 |
| **15** | 0.0199 | 0.0199 | 0.0199 | 0.0199 |
| $\mathbf{16}^{34}$ | 0.4877 | 0.4877 | 0.4877 | 0.2224 |
| $\mathbf{17}^{34}$ | 0.2006 | 0.2006 | 0.2006 | 0.0432 |
| $\mathbf{18}^{2}$ | 0.0249 | 0.0166 | 0.0166 | 0.0166 |
| $\mathbf{19}^{24}$ | 0.8084 | 0.8084 | 0.8084 | 0.7172 |
| $\mathbf{20}^{2}$ | 0.3255 | 0.3255 | 0.3255 | 0.3255 |
| $\mathbf{21}^{234}$ | 0.0122 | 0.0122 | 0.0122 | 0.0056 |
| $\mathbf{22}^{2}$ | 0.0648 | 0.0684 | 0.0684 | 0.0684 |
| $\mathbf{23}^{2}$ | 0.3469 | 0.6239 | 0.6239 | 0.6239 |
| **24** | 0.2222 | 0.2222 | 0.2222 | 0.2222 |
| $\mathbf{25}^{24}$ | 0.2712 | 0.2269 | 0.2269 | 0.2269 |

One possible explanation of this narrow difference is in the used English newspaper collection, which does not have enough subject diversity for certain entities to challenge retrieval systems. Take for instance topic #1, which addresses the death of Astrid Lindgren in January 2002; in the collection there are 10 documents containing the expression "Astrid Lindgren," and 6 of then are from January 2002, which are precisely the 6 documents found relevant by the assessors. In other words, for this topic there is no improvement to be gained by reformulated queries compared to the baseline query.

Conversely, in similar topic that address the death of Yasser Arafat (#12), we observe lower MAP values because there are more documents about Arafat on different contexts and scattered through the years, making it more challenging than topic #1. In this case, there is an improvement from the baseline (0.0877) to the supervised run (0.1330) because the reformulated query included references to where and when Arafat's death occurred, which helped to promote relevant documents in the result list.

Another result worth of mention is in topic #23, "When did the largest expansion of the European Union take place, and which countries became members?" While the baseline included only non-stop terms from the initial query, RENOIR expanded "European Union" to its 27 countries, adding them to the NE and geographic index, which was the main cause of a MAP increase from 0.3469 to 0.6239. While the correct answer (1st of May, 2004) was not found by RENOIR, the multiple references to EU countries contributed to the retrieval performance in this topic. Note that, since geographic signatures include information of place's ancestors up to a country level, a query geoscope only needs to be expanded if it is a continent or a geo-political entity that encompasses several countries.

Curiously, in topic #14, "When and where did a volcano erupt in Africa during 2002?," the expansion of "Africa" to its countries significantly lowered the MAP value. We believe that this is the result of disproportional partial BM25 scores in LGTE, which did not had a control mechanism for index weights at the time. In other words, the scores from the term index may have been eclipsed by NE or signature index scores, resulting in an unbalanced retrieval biased towards the geographic and temporal part of the topic rather than its thematic part.

# 5. CONCLUSIONS AND FUTURE WORK

The NTCIR Geo-Temporal evaluation track presented a challenging ad-hoc retrieval task, encouraging researchers to focus on the reasoning part of their systems. Our participation results in NTCIR were according to our expectations: other participants chose to use the description and narrative fields of the topic, achieving higher MAP values, but our MAP results, obtained using only the topic titles, were also high.

In our GikiCLEF participation [4], we already observed that RENOIR depends on the DBpedia datasets and their ontology mappings, and DBpedia information depends on the coverage of infoboxes in English Wikipedia articles. Also, RENOIR's reasoning step is fragile, as it may easily fail when questions involve specific actions which are difficult or even impossible to map to DBpedia properties (for instance, the GikiCLEF topic "Name places where Goethe fell in love"). As GeoTime presented event-biased topics, this problem was not so significant, but nonetheless it requires updated sets of matching rules to cope with the topics.

We conclude that the lack of significative improvement on the overall retrieval results using RENOIR's reformulated queries show that our GIR system still has limited reasoning capabilities and the bridge between the reasoning and retrieval step needs further improvements. However, there are some improvements observed for a specific set of complex questions, which are encouraging. Future work will focus on widening the coverage of queries that RENOIR can reason over, knowing that, at the same time, resources like Wikipedia and DBpedia will also evolve in quality and size, and on the reasoning/retrieval bridge, by experimenting with more query reformulation strategies, and having a better control on the LGTE setup for each retrieval, such as index weights and BM25 parameters.

# 6. ACKNOWLEDGMENTS

We would like to thank Jorge Machado, for the support and development of LGTE, Diana Santos for comments in the paper, and SINTEF ICT. This work is supported by FCT for its LASIGE Multi-annual support, GREASE-II project (grant PTDC/EIA/73614/2006) and a PhD scholarship grant SFRH/BD/45480/2008, and by the Portuguese Government, the European Union (FEDER and FSE) through the Linguateca project, under contract ref. POSC/339/1.3/C/NAC, UMIC and FCCN.

# 7. REFERENCES

[1] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives. DBpedia: A Nucleus for a Web of Open Data. In *6th International Semantic Web Conference, 2nd Asian Semantic Web Conference, ISWC 2007 + ASWC 2007, Busan, Korea, November 11–15, 2007, Proceedings*, number 4825 in LNCS, pages 722–735. Springer, 2007.

[2] N. Cardoso. REMBRANDT - Reconhecimento de Entidades Mencionadas Baseado em Relações e Análise Detalhada do Texto. In C. Mota and D. Santos, editors, *Desafios na avaliação conjunta do reconhecimento de entidades mencionadas: O Segundo HAREM*. Linguateca, 2009. in Portuguese.

[3] N. Cardoso. GikiCLEF topics and Wikipedia articles: did it blend? In C. Peters, G. D. Nunzio, M. Kurimo, T. Mandl, D. Mostefa, A. Peñas, and G. Roda, editors, *Multilingual Information Access Evaluation Vol. I: Text Retrieval Experiments*. Springer, 2010. to appear.

[4] N. Cardoso, D. Baptista, F. J. Lopez-Pellicer, and M. J. Silva. Where in the Wikipedia is that answer? the XLDB at the GikiCLEF 2009 task. In C. Peters, G. D. Nunzio, M. Kurimo, T. Mandl, D. Mostefa, A. Peñas, and G. Roda, editors, *Multilingual Information Access Evaluation Vol. I: Text Retrieval Experiments*. Springer, 2010. to appear.

[5] F. Gey, R. Larson, N. Kando, J. Machado, and T. Sakai. NTCIR-GeoTime Overview: Evaluating Geographic and Temporal Search. In *Working Notes of the 8th NTCIR Workshop*, Tokyo, Japan, June 15–18 2010.

[6] Y. Li, A. Moffat, N. Stokes, and L. Cavedon. Exploring Probabilistic Toponym Resolution for Geographical Information Retrieval. In R. Purves and C. Jones, editors, *Proceedings of the 3rd ACM Workshop On Geographic Information Retrieval, GIR 2006, Seattle, WA, USA, August 10, 2006*, pages 17–22. Department of Geography, University of Zurich, 2006.

[7] F. J. Lopez-Pellicer, M. Chaves, C. Rodrigues, and M. J. Silva. Geographic Ontologies Production in Grease-II. Technical Report DI/FCUL TR 09-18, Faculty of Sciences, University of Lisbon, November 2009. DOI 10455/3256.

[8] J. Machado. LGTE: Lucene Extensions for Geo-Temporal Information Retrieval. In *Workshop on Geographic Information on the Internet Workshop*

*(GIIW), held at ECIR 2009*, Toulouse, France, 9 April 2009.

[9] A. Miles, B. Matthews, M. Wilson, and D. Brickley. SKOS Core: Simple Knowledge Organisation for the Web. In *DCMI '05: Proceedings of the 2005 International Conference on Dublin Core and Metadata Applications*, pages 1–9. Dublin Core Metadata Initiative, 2005.

[10] S. E. Robertson, S. Walker, M. Hancock-Beaulieu, A. Gull, and M. Lau. Okapi at TREC-3. In *Proceedings of TREC-3*, pages 21–30, Gaithersburg, MD, USA, 1992.

[11] D. Santos and L. M. Cabral. GikiCLEF: Crosscultural issues in an international setting: asking non-English-centered questions to Wikipedia. In F. Borri, A. Nardi, and C. Peters, editors, *Cross Language Evaluation Forum: Working notes for CLEF 2009*, 30 September - 2 October 2009.

[12] D. Santos, N. Cardoso, P. Carvalho, I. Dornescu, S. H. an d Johannes Leveling, and Y. Skalban. GikiP at GeoCLEF 2008: Joining GIR and QA forces for querying Wikipedia. In C. Peters, T. Deselaers, N. Ferro, J. Gonzalo, G. J.F.Jones, M. Kurimo, T. Mandl, A. Peñas, and V. Petras, editors, *Evaluating Systems for Multilingual and Multimodal Information Access: 9th Workshop of the Cross-Language Evaluation Forum, CLEF 2008, Aarhus, Denmark, Sept ember 17-19, 2008, Revised Selected Papers*, pages 894–905. Springer, 2009.

[13] D. Santos, P. Carvalho, H. Oliveira, and C. Freitas. Second HAREM: new challenges and old wisdom. In A. Teixeira, V. L. S. de Lima, L. C. de Oliveira, and P. Quaresma, editors, *Computational Processing of Portuguese Language, 8th International Conference (PROPOR'2008), September 8-10, Aveiro, Portugal. Proceedings*, number 5190 in LNCS, pages 212–215. Springer, 2008.

[14] Yahoo! Geoplanet$^{TM}$. `http://developer.yahoo.com/geo/geoplanet/`, 2010. Accessed on May 2010.

[15] Yahoo! GeoPlanet$^{TM}$Key Concepts. `http://developer.yahoo.com/geo/geoplanet/guide/concepts.html`, 2010. Accessed on May 2010.