GikiP at GeoCLEF 2008: Joining GIR and QA forces for querying Wikipedia

Diana Santos[1], Nuno Cardoso[1,2], Paula Carvalho[2], Iustin Dornescu[3], Sven Hartrumpf[4], Johannes Leveling[5], Yvonne Skalban[3]

[1]Linguateca, SINTEF ICT (Norway)
[2]University of Lisbon, DI, LasiGE, XLDB, Linguateca (Portugal)
[3]Research Group in Computational Linguistics (CLG) at the University of Wolverhampton (UK)
[4]Intelligent Information and Communication Systems (IICS), University of Hagen (FernUniversität in Hagen) (Germany),
[5]Centre for Next Generation Localisation (CNGL), Dublin City University (Ireland)
Diana.Santos@sintef.no, ncardoso@xldb.di.fc.ul.pt, pqfcarvalho@gmail.com, i.dornescu2@wlv.ac.uk, Sven.Hartrumpf@fernuni-hagen.de, Johannes.Leveling@computing.dcu.ie, yvonne.skalban@wlv.ac.uk

Abstract. This paper reports on the GikiP pilot that took place in 2008 in GeoCLEF. This pilot task requires a combination of methods from geographical information retrieval and question answering to answer queries to the Wikipedia. We start by the task description, providing details on topic choice and evaluation measures. Then we offer a brief motivation from several perspectives, and we present results in detail. A comparison of participants' approaches is then presented, and the paper concludes with improvements for the next edition.

1 Introduction

This paper introduces GikiP, an evaluation contest on retrieving geographically-related information from Wikipedia in the form of a list of answers (corresponding to articles). Or, as stated on the website[1]: *Find Wikipedia entries (i.e. articles) that answer a particular information need which requires geographical reasoning of some sort.*

To guarantee a common evaluation ground, participants were requested to use the Wikipedia collection(s) already used in the QA@CLEF main track (2007 and 2008), dating from the end of 2006. Fifteen topics (see Table 1) were released on the 2nd of June 2008 in English, German, and Portuguese (eight example topics had already been published).

In order to conform to expectations of both the question answering (QA) and the geographical information retrieval (GIR) practitioners, topic titles were

[1] http://www.linguateca.pt/GikiP/

Table 1: Topic titles in GikiP 2008. "Lang." stands for the language biases in topic choice. There were three English, three Portuguese, four German and five other topics.

ID	English topic title	Lang.
GP1	Which waterfalls are used in the film "The Last of the Mohicans"?	EN
GP2	Which Vienna circle members or visitors were born outside the Austria-Hungarian empire or Germany?	DE
GP3	Portuguese rivers that flow through cities with more than 150,000 inhabitants	PT
GP4	Which Swiss cantons border Germany?	DE
GP5	Name all wars that occurred on Greek soil.	other
GP6	Which Australian mountains are higher than 2000 m?	EN
GP7	African capitals with a population of two million inhabitants or more	other
GP8	Suspension bridges in Brazil	PT
GP9	Composers of Renaissance music born in Germany	DE
GP10	Polynesian islands with more than 5,000 inhabitants	other
GP11	Which plays of Shakespeare take place in an Italian setting?	EN
GP12	Places where Goethe lived	DE
GP13	Which navigable rivers in Afghanistan are longer than 1000 km?	other
GP14	Brazilian architects who designed buildings in Europe	PT
GP15	French bridges which were in construction between 1980 and 1990	other

in a QA format, while the topic description was generally a less condensed and more verbose version of the topic title, but would not add crucial information. See for example topic GP5 in the English version:

```
<top lang="en">
<num>GP5</num>
<title>Name all wars that occurred on Greek soil.</title>
<description>Wars that took place in (ancient or modern) Greece
are relevant.</description> </top>
```

Participants had ten days to return the results as a list of titles of Wikipedia pages. The maximum number of documents returned per topic was set to 100, but the topics chosen by the organizers had typically considerably fewer hits.

Only answers / documents of the right type were considered correct. In other words, if a topic concerned painters, the result should be a list of names of painters, and not names of boats or countries. That is, it was not enough that the answer were found in a particular Wikipedia document: it had to be its title. GikiP expected only precise and concise answers that must have been mapped to the correct Wikipedia article name (no homographs).

Evaluation was devised in order to emphasize diversity and multilinguality. Systems able to retrieve a higher number of answers and in more languages should be considered better, so we introduced a simple bonus in order to reward multilinguality, *mult*, being 1, 2 or 3 depending on the number of languages tried out by the systems. More precisely, the score for each topic was calculated according to the following formula: $mult \cdot N \cdot N/total$, where N is the number of correct answers found, and $N/total$ is the precision. The system's final score was defined as the average of its scores per topic.

For further discussion on topic choice, as well as interesting problems on topic translation and assessment, see our longer paper in the CLEF 2008 Working Notes [1], where for example, we point out that different questions are easier to answer (and more natural to pose) in different languages.

2 Motivation

A number of different motivations led us to organize or participate in GikiP:

- the wish to innovate and add difficulty to both QA and GIR (in its GeoCLEF variety), given that both tasks have been quite stable in the last 3–4 years,
- the fact that Wikipedia has established itself as the main/largest multilingual resource for NLP in the last years,
- the belief that asking list questions to an encyclopaedia is very useful for a wide variety of people,
- the need to devise truly crosslingual and multilingual evaluation set-ups so that it makes sense to harvest information in more than one language (as previously pointed out e.g. by [2]),
- the hope that merging QA and IR is a fruitful path for information access in general, and
- the interest in geographical information in natural language, which led us to explore different and more complex ways of encoding place in texts than those ordinarily coped for by GeoCLEF (see e.g. [3, 4]).

GikiP is closely related to WiQA [5], a challenging task devised to assist creators of new Wikipedia pages with multilingual data. GikiP has a more general user model, and a more modest requirement: we simply expect systems to answer open list questions with a list of factoids.

From an IR perspective, we believe that IR research has been well aligned with the real demand for bigger and better retrieval models. Nowadays, another trend on the user's demand on search tools can be observed: users start to expect an IR system to understand better the topics addressed and to reason over the answers, instead of just computing and ranking results according to simple term similarity approaches. Therefore, new IR approaches require a good understanding of the context of the user's query to capture the real information need behind it, and must therefore turn to knowledge extraction and use what traditionally is considered an NLP approach to improve their retrieval and ranking based on more than bag-of-words approaches. Geographic IR is a good place to start working on these new semantic IR approaches, as it focuses on a specific angle (geographic) and has as obvious application to endorse IR systems with geographic reasoning capabilities.

One of the bottlenecks that limit the performance of QA systems is the fact that they focus on extracting the answer from plain text, using several specific

NLP tools that may offer extra information (e.g. named entity recognition, parsing, semantic role labeling) at the cost of adding their own biases and limitations.

One property of GikiP that may put QA systems on the right track is that it encourages systems to exploit hyperdata, not just plain text, and also process semi-structured contexts. In fact, to answer GikiP questions, information had to be extracted from several articles, links between articles had to be analyzed, and often the category hierarchy and the infoboxes of Wikipedia had to be used. This is a property that we believe realistic QA systems should have: process several kinds of information sources and strategies and merge that information in a coherent and more informative answer.

Another issue in which we believe GikiP represents considerable progress compared to usual QA contests (at least QA@CLEF) is the fact that these restrict list questions to what we consider the rarest and least interesting kind of lists, dubbed closed lists, such as "Name the seven hills of Rome" or "Name the four Beatles members". In fact, it seems that most participating systems in QA@CLEF even ignore this kind of questions altogether (probably because of the rigid evaluation and their little number – 10 out of 200 in 2008). Our opinion is that a realistic QA system should not be restricted to closed list questions. From a user point of view, any list question makes sense and, in fact, the user may not even know the right number of answers from the start.

The strongest point for GikiP is that it definitely calls for an integration perspective between QA and IR. In fact, there are several arguments for combining approaches from QA and IR for a successful solution to the GikiP task.

Taking a look at the topics, one finds that several require an interpretation of geographic relations (e.g. GP2, GP4), some include measurable properties of locations (e.g. GP3, GP6, GP7), others aim at resolving temporal constraints (e.g. GP2, GP9, GP15), and still others include words with an irregular morphology which are derived from location names (e.g. GP3, GP4, GP5, GP11, GP15).

While IR methods will be useful to provide an initial result set for these topics, methods like inferences or semantic processing will be required to ensure high precision. Interpreting the task as a kind of QA, on the other hand, QA often provides high precision originating from methods that can deal successfully with these kinds of problems. However, some QA methods lack robustness to provide each question with a correct answer, as already pointed out.

On the document level, information required to find matching Wikipedia articles can be contained in either the textual part or in other parts like tables. Measurable properties of entities are typically listed as attribute-value pairs in tables. Thus, processing the GikiP topics benefits from approaches analyzing the textual information in the Wikipedia articles (e.g. methods from QA) and approaches employing structured information (e.g. methods from IR or information extraction). Furthermore, the Wikipedia corpus offers a wealth of additional resources which are useful for both QA and IR approaches (e.g. Wikipedia categories, Wikipedia links between articles, and inter-language links).

Table 2: Topic size of GikiP 2008, only automatic runs. "Unique correct" stands for number of correct hits, removing duplicates in other languages.

Topic	**Results**	**Correct**	**Accuracy (%)**	**Unique correct**
GP1	5	1	20.00	1
GP2	31	7	22.58	4
GP3	28	8	28.57	5
GP4	79	19	24.05	6
GP5	69	19	27.54	15
GP6	36	7	19.44	4
GP7	90	33	36.67	14
GP8	49	2	4.08	1
GP9	49	15	30.61	15
GP10	53	2	3.77	1
GP11	35	24	68.57	12
GP12	51	25	49.02	10
GP13	9	4	44.44	2
GP14	60	6	10.00	4
GP15	18	2	11.11	1
Total	662	174	26.28	95

Finally, a result for a GikiP topic is a small set of answers to an open list question. The size of result sets for GikiP lies between fixed-size result sets of 1000 ranked documents found in IR tasks and the single-answer result set for QA tasks. In contrast to QA@CLEF, no redundancy in candidate answers can be exploited (because there are no duplicate Wikipedia articles); furthermore, no answer extraction from the articles is necessary for GikiP.

In summary, a successful solution to the GikiP task calls for a combination of approaches from IR and QA, either in sequence (e.g. filtering candidates and applying semantic filters) or in parallel (e.g. using IR as a fall-back to brittle QA methods).

3 Global Results

Before comparing the performance of the actual participating systems, we wanted to assess the task's feasibility, and investigate whether the results were of interest from a crosslingual point of view. We therefore pooled all results obtained, listing, for each topic, the total number of answers, as well as the number of correct answers found, displayed in Table 2. The answers themselves may not be different, they just need to correspond to a different Wikipedia article, so we had to also compute the number of **distinct** correct answers (Unique correct). If we group the answers by closest language (the one described in Table 1), no group seems to stand out: for *correct/total* we get 33.2% for German, 34.2% for Portuguese, 35.0% for English and 25.2% for the remaining topics.

A different way of investigating language weight is presented in Table 3(a) which counts the cases where answers were found in all three languages as well

Table 3: An investigation of the weight of languages in GikiP 2008.

Kind	DE	EN	PT
Total	233	255	174
Correct (174)	32	84	58
Unique correct	0	32	11

(a) Number of answers per language in GikiP 2008.

From/To	DE	EN	PT
DE	-	32C	29C, 3M
EN	71C, 13M	-	45C, 35M, 1W
PT	52C, 6M	51C, 7M	-

(b) Crosslingual boost by simply following the direct links.

as in only one of them. This is not, however, a reliable measure of Wikipedia contents, due to the very different approaches (and success rate) of the different participants, to which we turn in the following section.

Finally, we investigated the issue of, departing from each language in turn, how many other language hits were possible to recover using only direct translation links. In Table 3(b) we present the results by manually following the links present in the correct answers of each language, and classifying them into C(orrect): the answer arrived at by following the translation link is correct; M(issing): there is no translation link, so one would not arrive at this answer by simply following the translation link; or W(rong): the answer arrived at by following the translation link is wrong.

4 Overview of Participation and System Results

These are the participating systems: one participant per country where one of the three languages is spoken, and participation was divided equally between GeoCLEFers and QA@CLEFers (given that the IICS group is known to participate in both):

- **GIRSA-WP**, represented by Sven Hartrumpf and Johannes Leveling, Intelligent Information and Communication Systems (IICS) at the FernUniversität in Hagen (Germany), submitted six fully automatic runs with results in the three languages
- **RENOIR** (acronym for REMBRANDT's Extended NER On Interactive Retrievals), represented by Nuno Cardoso, University of Lisbon, Faculty of Sciences, LaSIGE, XLDB (Portugal), submitted one semi-automatic run only in English and Portuguese
- **WikipediaListQA@wlv**, represented by Iustin Dornescu, Research Group in Computational Linguistics (CLG) at the University of Wolverhampton (UK), submitted a fully automatic run in the three languages

We have also requested manual runs (based on the current on-line Wikipedia), for two reasons: we wanted to compare human performance to automatic answers (after the evaluation), and we also wanted to assess how much the information in Wikipedia had changed regarding the particular topics, from the official collections dating from November 2006 to the June 2008 date.

The results obtained by the systems, together with a fully manual run based on the on-line Wikipedia, can be found in Table 4. It is interesting to note that the human participant was not able to find any results for topics 2 and 5, contrary to the automatic systems, which together managed to find 7 and 5 correct hits, respectively.

Turning now to a short description of the participating systems, as can be seen in further detail in [1], they took a wide range of approaches, as well as different starting collections: GIRSA-WP used the German corpus, WikipediaListQA@wlv the English corpus, RENOIR the English and the Portuguese corpus. RENOIR is currently only semi-automatic, the other two systems are fully automatic.

4.1 GIRSA-WP

GIRSA-WP (GIRSA for Wikipedia) is a fully automatic, hybrid system combining methods from QA and GIR. In particular, it merges results from InSicht, an open-domain QA system [6], and GIRSA, a system for textual GIR [7]. In comparison with the two underlying systems, GIRSA-WP applies a semantic filter on the article titles (which are encoded in the answers in GikiP) to increase precision. This semantic filter ensures that the expected answer type (EAT) of the topic and the title of a Wikipedia article are compatible. This technique is widely known from QA for typical answer types such as PERSON, ORGANIZATION, or LOCATION. In the GIRSA-WP system, a concept (a disambiguated word) corresponding to the EAT is extracted from the topic title or description. Then, this concept and the title of a candidate article are parsed by WOCADI [8], a syntactico-semantic parser for German text. The semantic representations (more specifically, the ontological sort and the semantic features, see [9] for details) of the semantic heads are unified. If this unification succeeds, the candidate article is kept; otherwise it is discarded.

The major differences to InSicht and GIRSA are that GIRSA-WP does not merge streams of answers and does not include a logical answer validation. In contrast to GIRSA, the retrieval is based on documents indexed on a per-sentence basis of Wikipedia articles. In addition, the documents from Wikipedia had not been geographically annotated at all.

For the GikiP experiments, the topic title and description were analyzed and sent to GIRSA and InSicht. In GIRSA, the top 1000 results were retrieved and scores were normalized in the interval from 0 to 1. For results returned

Table 4: GikiP results in 2008.

Run	**Answers**	**Correct**	**Avg. Prec.**	**Score**
GIRSA-WP (best)	79	9	0.107	0.704
GIRSA-WP (all runs merged)	372	11	0.038	0.286
RENOIR	218	120	0.551	10.706
WikipediaListQA@wlv	123	94	0.634	16.143

by both GIRSA and InSicht, the maximum score was chosen. Results whose score was below a given threshold were discarded and the semantic filter was applied to the remaining results. To obtain multilingual results, the German article names were translated to English and Portuguese using the Wikipedia linking between languages. Note that this linking was the only non-textual information GIRSA-WP used from Wikipedia; for example, categories and inter-wiki links were completely ignored.

In InSicht, the semantic representation of the query and the semantic representations of document sentences are compared. To go beyond perfect matches, InSicht uses many techniques, for example intratextual coreference resolution, query expansion by inference rules and lexico-semantic relations, and splitting the query semantic network at certain semantic relations. InSicht employed a special technique called *query decomposition* (first tried in GeoCLEF in 2007 [7]) or *question decomposition* in the context of QA [10]. Among the different decomposition classes described in the latter paper, only meronymy decomposition and description decomposition are promising for current queries in GikiP.

The results during the evaluation were somewhat disappointing. There are several reasons for this. Due to time constraints, the Wikipedia articles had not been fully processed for GIRSA and some methods have been applied to the topics only although they should have been applied to the documents, too. For InSicht, the main problems were (1) that important information is given in tables (like inhabitant numbers), but the syntactico-semantic parser ignores these parts of articles and (2) that the semantic matching approach forming the basis of QA is still too strict for the IR-oriented parts of GikiP queries (similar problems occurred for GeoCLEF experiments).

Future work will include enabling the annotation of geographic entities and geo-inferences, and preferring special regions of Wikipedia articles (for example, the introductory sentences).

4.2 RENOIR

The goal of RENOIR's participation in GikiP was to explore new ways of doing GIR, specially for those kinds of geographic queries that cannot be correctly handled by just naïvely expanding the query terms and hoping that an IR system with some sort of geographic reasoning capabilities would capture the full meaning of the topic at stake, as XLDB does for GeoCLEF [11].

As such, XLDB chose to participate with one semi-automatic run using *query procedures* as retrieval input, instead of query terms, defining query procedures as a group of pipelined actions that express each GikiP topic. The selection of query procedures for a given topic was entirely manual, and the execution varied between automatic, semi-automatic and manual.

RENOIR is an interactive tool where query procedures are executed, generating partial and final results for each GikiP topic. RENOIR makes extensive use of REMBRANDT [12], a named entity recognition module which explores the Wikipedia document structure, links and categories, to identify and classify named entities (NEs) in Portuguese and English texts.

Table 5: Query procedures in RENOIR (A: automatic, M: manual, A/M: semi-automatic).

`SEARCH TERM`	A	Performs a simple term query search in the GikiP 2008 collection, and returns a list of Wikipedia documents.
`SEARCH CATEGORY`	A	Searches the Wikipedia dumps for documents with the given Wikipedia category, and returns a list of Wikipedia documents.
`SEARCH INLINKS`	A	Searches the Wikipedia dumps for documents that link to a given Wikipedia document.
`MAP DOC`	A/M	Maps a document from the Wikipedia dump to its counterpart in the GikiP 2008 collection.
`MAP NE`	A/M	Maps a NE to its corresponding document in the GikiP 2008 collection.
`REMBRANDT`	A	Annotates selected Wikipedia document(s) with REMBRANDT, generating lists of NEs for each document.
`REMB. DOC TO NE`	A	Invokes REMBRANDT to classify the title of a given Wikipedia document, generating the respective NE.
`FILT. NE BY TYPE`	A	Filters a list of NEs of a given classification category, generating a subset of NEs.
`FILT. DOC BY TERM`	A	Filters a list of Wikipedia documents by having (or not) a given term/pattern
`FILT. DOC BY EVAL`	M	Filters a list of Wikipedia document by evaluating a condition for a given subset of NEs. For instance, if the document has a number NE greater than 1000, or if it has a place name NE within Europe.

The GikiP 2008 collection was indexed with MG4J [13], which was used for basic document retrieval. For retrievals involving Wikipedia categories and links, different snapshots of Wikipedia (namely the Portuguese and English static SQL dumps from April 2008) were used, since the information regarding Wikipedia categories, redirections and page links was already available in SQL databases.

The RENOIR actions used for the query procedures are described in Table 5. The query procedures were formulated in a simple modular and pipelined approach, to “divide and conquer” the complex task of translating the GikiP topics into a machine-understandable way. So, the actions that could be made automatically were therefore implemented, while the more complex actions performed in GikiP with human intervention (so far) were also kept simple in order to be possible to extend RENOIR to perform them automatically in the future.

The next obvious step is to implement the automatic generator of query procedures, dealing with the problems that were mitigated by using human reasoning. At the same time, future work in RENOIR includes the improvement of the Wikipedia mining approaches, namely extracting information from infoboxes.

4.3 WikipediaListQA@wlv

The participation in this pilot task was motivated by CLG's interest in using Wikipedia as a backbone in QA. In addition, the task required the system to rely on the information inherent in the Wikipedia article link graph and the relation between entities, rather than developing accurate textual answer extractors. For example, in GP4, the fact that a river flows through a city with a high population is not stated in any Wikipedia article. However, given an article that describes a river (e.g. Douro), all the out links can be extracted and by just examining the category assignments and the infoboxes of the corresponding articles, the list of cities that have a population higher than the given threshold can be obtained.

WikipediaListQA@wlv proposes a simple model for topic interpretation that uses few language dependent resources. It exploits relationships between entities that may not be expressed in the article text, but are implied by the links between the articles. In order to navigate the Wikipedia link graph, the Wikipedia SQL dump was used and the articles' text was indexed with Lucene [14]. The system starts by identifying a domain category that comprises candidate articles, and then removes the ones that do not correspond to the topic filter. Thus two parts are identified in each topic: a) the *domain* of the candidate answers, and b) the *filters* to apply in order to select the correct ones.

In order to identify a Wikipedia category that would describe the domain of candidate articles, a parser [15] was used to extract the first noun phrase of the topic which was then matched to a category, by querying the Lucene index. For the purpose of this pilot only very simple filters were implemented: entity filters and factoid filters. The entity filters match documents that mention or have a link to a given entity (8 of the 15 topics). The factoid filters try to extract a particular fact, and the value is then compared to the selection criterion. The facts were: *population* (GP3, GP7, GP10), *nationality* (GP2, GP9), *height* (GP6) and *length* (GP13). Articles from which the fact could not be extracted were dismissed.

The accuracy of the system is limited due to the ambiguity of links. Category relations are not classified: hypernymy vs. meronymy vs. similarity, thus very large article sets might be extracted (the system did not return any results for GP5,GP9 and GP10 because of ambiguity). This might be avoided, in further work, by using resources that map Wikipedia articles to WordNet and disambiguate the type of the entity described in each article (e.g. Yago [16] and DBpedia [17]).

The main advantage of the system is that – using a small set of filters – very complex data can be extracted from Wikipedia. Its disadvantage lies in the complexity of correctly identifying (combined) filters in natural language questions.

5 Concluding Remarks

The results presented by this pilot are encouraging. We believe to have demonstrated the possibility of automating the particular task at hand, and that there are interesting kinds of non-trivial questions that have a retrievable answer in Wikipedia. Furthermore, answer correctness can be quickly assessed by the users, often without even having to visit the page.

But before these systems can reach the general public, much work still remains to be done: for example, dealing with issues of redundancy removal, choice of which language / answer to present first (or only), how to present a compound set of pages to justify a particular answer, and so on.

In particular, we found that mixing answers of different granularity should be properly dealt with. (A list of places containing cities, castles, and countries is not a pleasing outcome for most users.) More often than expected, answers to (apparently simple) questions required clarification: Temporal aspects are important and cannot be overlooked, and even definitional aspects pop up: if you are looking for wars, do you want also battles?

A new edition of this task with more languages (Bulgarian, Dutch, Italian, Norwegian, Romanian and Spanish), more (50) topics, and with a stronger focus on cross-cultural issues, is currently being organized, named GikiCLEF.[2] Some of the improvements are (i) a more appropriate scoring function rewarding multilinguality, which distinguishes, on a per topic basis, whether there were answers at all in a particular language, (ii) reducing the value of (trivial) direct translation links compared to radically different items in another language, and (iii) allowing for more complex justifications if needed.

Acknowledgements The GikiP organizers are grateful to Ross Purves for checking the English topics, to Sven Hartrumpf's WOCADI parser for debugging the German and English topics, to Anselmo Peñas for supplying the Wikipedia collections and to Paulo Rocha for providing a manual run in two days.

Organization was done in the scope of the Linguateca project, jointly funded by the Portuguese Government and the European Union (FEDER and FSE) under contract ref. POSC/339/1.3/C/NAC. Development of the WikipediaListQA@wlv system was partly supported by the EU funded project QALL-ME (FP6 IST-033860). Grant SFRH/BD/29817/2006 from FCT (Portugal) supported the development of RENOIR.

References

1. Santos, D., Cardoso, N., Carvalho, P., Dornescu, I., Hartrumpf, S., Leveling, J., Skalban, Y.: Getting geographical answers from Wikipedia: the GikiP pilot at CLEF. In Borri, F., Nardi, A., Peters, C., eds.: Working Notes for the CLEF 2008 Workshop, 17-19 September, Aarhus, Denmark. (2008)
2. Santos, D., Rocha, P.: The key to the first CLEF in Portuguese: Topics, questions and answers in CHAVE. [18] 821–832
3. Santos, D., Chaves, M.: The place of place in geographical IR. In: Proceedings of GIR06, the 3rd Workshop on Geographic Information Retrieval (GIR 2006), Seattle, 10 August 2006. (2006) 5–8

[2] http://www.linguateca.pt/GikiCLEF/

4. Gey, F., Larson, R., Sanderson, M., Bishoff, K., Mandl, T., Womser-Hacker, C., Santos, D., Rocha, P., Nunzio, G.D., Ferro, N.: GeoCLEF 2006: the CLEF 2006 Cross-Language Geographic Information Retrieval Track Overview. [19] 852–876
5. Jijkoun, V., de Rijke, M.: Overview of the WiQA Task at CLEF 2006. [19] 265–274
6. Hartrumpf, S.: Question answering using sentence parsing and semantic network matching. [18] 512–521
7. Leveling, J., Hartrumpf, S.: Inferring location names for geographic information retrieval. In Peters, C., Jijkoun, V., Mandl, T., Müller, H., Oard, D.W., Peñas, A., Petras, V., Santos, D., eds.: Advances in Multilingual and Multimodal Information Retrieval: 8th Workshop of the Cross-Language Evaluation Forum, CLEF 2007. Volume 5152 of LNCS., Berlin, Springer (2008) 773–780
8. Hartrumpf, S.: Hybrid Disambiguation in Natural Language Analysis. Der Andere Verlag, Osnabrück, Germany (2003)
9. Helbig, H.: Knowledge Representation and the Semantics of Natural Language. Springer, Berlin (2006)
10. Hartrumpf, S.: Semantic decomposition for question answering. In Ghallab, M., Spyropoulos, C.D., Fakotakis, N., Avouris, N., eds.: Proceedings of the 18th European Conference on Artificial Intelligence (ECAI), Patras, Greece (July 2008) 313–317
11. Cardoso, N., Sousa, P., Silva, M.J.: The University of Lisbon at GeoCLEF 2008. In Borri, F., Nardi, A., Peters, C., eds.: Working Notes for the CLEF 2008 Workshop, 17-19 September, Aarhus, Denmark. (2008)
12. Cardoso, N.: REMBRANDT - Reconhecimento de Entidades Mencionadas Baseado em Relações e Análise Detalhada do Texto. In Mota, C., Santos, D., eds.: Desafios na avaliação conjunta do reconhecimento de entidades mencionadas, Linguateca (2008) 187–204
13. Boldi, P., Vigna, S.: MG4J at TREC 2005. In: Proceedings of the 14th Text REtrieval Conference (TREC 2005). (2005)
14. Hatcher, E., Gospodnetic, O.: Lucene in Action (In Action series). Manning, Greenwich, CT, USA (2004)
15. Tapanainen, P., Järvinen, T.: A non-projective dependency parser. In: Proceedings of the 5th Conference of Applied Natural Language Processing, Washington D.C., USA, ACL (1997) 64–71
16. Suchanek, F.M., Kasneci, G., Weikum, G.: Yago: a core of semantic knowledge. In: WWW '07: Proceedings of the 16th international conference on World Wide Web. ACM Press, New York, NY, USA (2007) 697–706
17. Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., Ives, Z.: DBpedia: A Nucleus for a Web of Open Data. In: The Semantic Web: 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference, ISWC 2007 + ASWC 2007, Busan, Korea, November 11–15, 2007. Proceedings. Springer (2008) 722–735
18. Peters, C., Clough, P., Gonzalo, J., Jones, G.J.F., Kluck, M., Magnini, B., eds.: Multilingual Information Access for Text, Speech and Images, 5th Workshop of the Cross-Language Evaluation Forum, CLEF 2004, Bath, UK, September 15-17, 2004, Revised Selected Papers. Springer (2005)
19. Peters, C., Clough, P., Gey, F.C., Karlgren, J., Magnini, B., Oard, D.W., de Rijke, M., Stempfhuber, M., eds.: Evaluation of Multilingual and Multi-modal Information Retrieval: 7th Workshop of the Cross-Language Evaluation Forum, CLEF 2006. Revised selected papers. Volume 4730 of LNCS. Springer, Berlin (2007)