Optimizing Geographical Entity and Scope Resolution in Texts using Non-Geographical Semantic Information

Panos Alexopoulos and Carlos Ruiz
iSOCO, Intelligent Software Components S.A.
Av. del Partenon, 16-18, 28042, Madrid, Spain,
Email: {palexopoulos, cruiz}@isoco.com

***Abstract*—Assigning geographical meta-information to textual pieces of information in an automatic way is a challenging semantic processing task that has been getting increasing attention from application and research areas that need to exploit this kind of information. With that in mind, in this paper, we propose a novel ontology-based framework for correctly identifying geographical entity references within texts and mapping them to corresponding ontological uris, as well as determining the geographical scope of texts, namely the areas and regions to which the texts are geographically relevant. Unlike other approaches which utilize only geographical information for performing these tasks, our approach allows the exploitation of any kind of semantic information that is explicitly or implicitly related to geographical entities in the given domain and application scenario. This exploitation, according to our experiments, manages to substantially improve the effectiveness of the geographical entity and scope resolution tasks, especially in scenarios where explicit geographical information is scarce.**

***Keywords*-Location Disambiguation; Geographical Scope Resolution; Ontologies.**

I. Introduction

With the rapidly increasing popularity of Social Media sites, a lot of user-generated content has been injected in the Web resulting in a large amount of both multimedia items and textual data (tags and other text-based documents) [1]. As a consequence, it has become increasingly difficult to find exactly the objects that best match the users' information needs. Besides, as more of those searches are performed from mobile applications, geographic intent and scope become indispensable as users expect a search system not only to know their current location, but to understand their entire geographic context. Therefore, it is crucial for the system to be able to infer what is the location (if any) implicit in their search and the user-generated content.

Thus, Geographical Intention Retrieval [2] concerns all kinds of techniques related to the retrieval of information involving some kind of spatial awareness. These methods can improve all kinds of services and applications that rely on geographical information, ranging from its quite straightforward use in map services, to more advanced techniques of personalization. For example, a user searching for cheap flights to Paris has the implicit intent of flying from his current location, although the latter was not stated. This implicit geographic nature of user queries is called geographic intent.

On the other hand, a text or a query has a geographic scope. For example, a query for cheap flights from London to Paris would include both London and Paris in the geographic scope, but not locations in between. Similarly, a text describing the Eiffel tower will have the geographic scope of Paris, rather than of France, although both locations could be mentioned in the tag set.

Geo-location services enable retrieval of likely geographical locations for given keywords or text [3]. Most of them apply data mining and statistical techniques on big-scale data sets in the Internet, nevertheless they rely only in syntactic analysis, missing the benefits of exploiting the real meaning of a piece of text. This leaves them suffering issues such as disambiguation problems with locations with the same name (Paris, France vs. Paris, Texas) or locations named somehow similar to non-geographic concepts (such as Reading, UK).

On the other hand, semantic analysis, either built on top of statistical analysis or as a standalone approach, can improve the previous approach by extracting not only geographical entities from a text, but also other types of entities (people, companies, etc.) that can, via reasoning or inference techniques, extract further geographic information.

Of course, the main limitation of semantic approaches is the need for geographical knowledge bases as input to the system, typically a bottleneck in the whole process. Previous approaches have tried to build geographic knowledge on top of different kind of resources, including ad hoc ontologies, geo-gazetteers or more generic knowledge hubs such as Wikipedia. However, the reuse of Open Data is a key element for improving this approach, avoiding or at least limiting the initial entry barriers for geographical semantic analysis. In particular, the Linked Data initiative [4] provides a crucial starting point for building a large and reliable geographical centered knowledge base, with enough information from other type of entities to allow for a comprehensive coverage of most domains.

Given the above, in this paper, we focus on geographical analysis of textual information and we propose a novel ontology-based framework for tackling two problems:

1) The problem of **geographical entity resolution**,

 ISBN: 978-1-61208-240-0

namely the detection within a text of geographical entity references and their correct mapping to ontological uris that represent them.
2) The problem of **geographical scope resolution**, namely the determination of areas and regions to which the text is geographically relevant.

The distinguishing characteristic of this framework is that, unlike other ontology-based approaches which utilize only geographical information for performing the above tasks, it allows the exploitation of any kind of semantic information that is explicitly or implicitly related to geographical entities in the given domain and application scenario. In that way, it manages to significantly improve the accuracy of the the above tasks, especially in domains and scenarios where explicit geographical information is scarce.

The rest of the paper is as follows. Section II presents related works. Section III presents our proposed framework and its components while Section IV presents and discusses experimental results regarding the evaluation of the framework's effectiveness. Finally, conclusions are drawn in Section V.

II. Related Work

Most related approaches to our work originate from the area of geographical information retrieval [2], where several approaches based on information retrieval, machine learning or semantic techniques are proposed to resolve geographic entities and scope.

Andogah et al. [5] describe an approach to place ambiguity resolution in text consisting of three components; a geographical tagger, a geographical scope resolver, and a placename referent resolver. The same authors, in [6], also propose determining the geographical scope as means to improve the accuracy in relevance ranking and query expansion in search applications. However these processes only rely on limited geographical information rather than using some other data available.

More related to the process of attempting to discern whether a texts topic is location-related, Mei et al. [7] present methods for finding latent semantic topics over locations (states or countries) and Wang et al. [8] propose a Location-Aware Topic Model based on Latent Dirichlet Allocation [9].

Besides, some other general approaches related to location disambiguation and inference are based on a query expansion process that augments a user's query with additional terms in order to improve the results, plus a filtering process for determining the relevance of results to the original query. For that, different dimensions can be taken into account in terms of how the relevance should be measured, ranging from its accuracy in a particular context to the inner meaning between terms. There are two primary query expansion approaches [10], [11]: on the one hand, probabilistic approaches sample from terms that co-occur with the original query as the basis for the expansion in a local or global context, and, on the other hand, the use of ontologies by semantic approaches for query expansion relies on the formal and strongly defined structure they introduce, exploiting the existent relations between different concepts and entities.

Following a strict semantic approach, Kauppinen et al. [12] present an approach using two ontologies (SUO - a large Finnish place ontology, and SAPO - a historical and geographical ontology) and logic rules to deal with heritage information where modern and historical information is available (e.g., new name for a place, new borders in a country). This method is combined with some faceted search functionalities, but they do not propose any method for disambiguating texts.

More related to the fact that the disambiguation of a location depends on the context (such as in "London, England" vs. "London, Ontario"), Peng et al. [13] propose an ontology-based method based on local context and sense profiles combining evidence (location sense context in training documents, local neighbor context, and the popularity of individual location sense) for such disambiguation.

III. Proposed Geographical Scope Resolution Framework

Our proposed framework targets the two tasks of geographical entity and scope resolution based on a common assumption: that the existence of both geographical and non-geographical entities within a text may be used as **evidence** that indicate which is the most probable meaning of an ambiguous location term as well as which locations constitute the geographical scope of the whole text.

To see why this assumption is valid, consider a historical text containing the term "Tripoli". If this term is collocated with terms like *"Siege of Tripolitsa"* and *"Theodoros Kolokotronis"* (the commander of the Greeks in this siege) then it is fair to assume that this term refers to the city of Tripoli in Greece rather than the capital of Libya. Also, in a historical text like *"The victory of Greece in the Siege of Tripolitsa under the command of Kolokotronis was decisive for the liberation from Turkey"*, the evidence provided by *"Siege of Tripolitsa"* and *"Kolokotronis"* and *"Greece"* indicates that Tripoli is more likely to be the location the text is about rather than Turkey.

Of course, which entities and to what extent may serve as evidence in a given application scenario depends on the domain and expected content of the texts that are to be analyzed. For example, in the case of historical texts we expect to use as evidence historical events and persons that have participated in them. For that reason, our approach is based on the a priori determination and acquisition of the optimal evidential knowledge for the scenario in hand. This knowledge is expected to be available in the form of an ontology and it's used within the framework in order to perform geographical entity and scope resolution. In

 ISBN: 978-1-61208-240-0

particular, our proposed framework comprises the following components:

- A **Geographical Resolution Evidence Model** that contains both geographical and non-geographical semantic entities that may serve as location-related evidence for the application scenario and domain at hand. Each entity is assigned evidential power degrees which denote its usefulness as evidence for the two resolution tasks.
- A **Geographical Entity Resolution Process** that uses the evidence model to detect and extract from a given text terms that refer to locations. Each term is linked to one or more possible location uris along with a confidence score calculated for each of them. The uri with the highest confidence should be the correct location the term refers to.
- A **Geographical Scope Resolution Process** that uses the evidence model to determine, for a given text, the location uris that potentially fall within its geographical scope. A confidence score for each uri is used to denote the most probable locations.

In the following paragraphs, we elaborate on each of the above components.

A. Geographical Resolution Evidence Model

For the purpose of this paper, we define an ontology as a tuple $O = \{C, R, I, i_C, i_R\}$ where

- C is a set of concepts.
- I is a set of instances.
- R is a set of binary relations that may link pairs of concept instances.
- i_C is a concept instantiation function $C \rightarrow I$.
- i_R is a relation instantiation function $R \rightarrow I \times I$.

Given an ontology, the **Geographical Resolution Evidence Model** defines which ontological instances and to what extent should be used as evidence towards i) the correct meaning interpretation of a location term to be found within the text and ii) the correct geographical scope resolution of the whole text. More formally, given a domain ontology O and a set of locations $L \subseteq I$, a geographical resolution evidence model consists of two functions:

- A **location meaning evidence function** $lmef : L \times I \rightarrow [0, 1]$. If $l \in L$ and $i \in I$ then $lmef(l, i)$ is the degree to which the existence, within the text, of i should be considered an indication that l is the correct meaning of any text term that has l within its possible interpretations.
- A **geographical scope evidence function** $gsef : L \times I \rightarrow [0, 1]$. If $l \in L$ and $i \in I$ then $gsef(l, i)$ is the degree to which the existence, within the text, of i should be considered an indication that l represents the geographical scope of the text.

In order to determine the above functions for a given domain and scenario we need to consider the concepts whose instances are directly or indirectly related to locations and which are expected to be present in the text to be analyzed. This, in turn, means that some a priori knowledge about the domain and content of the text(s) should be available. The more domain specific the texts are, the smaller the ontology needs to be and the more effective and efficient the whole resolution process is expected to be. In fact, it might be that using a larger ontology than necessary could reduce the effectiveness of the resolution process.

To illustrate this point assume that the texts to be analyzed are about American History. This would mean that the locations mentioned within these texts are normally related to events that are part of this history and, consequently, locations that had nothing to do with these events need not be considered. In that way, the range of possible meanings for location terms within the texts as well as the latter's potential scope is considerably reduced.

Thus, a strategy for selecting the minimum required instances that should be included in the location evidence model would be the following:

- First, identify the concepts whose instances may act as location evidence in the given domain and texts.
- Then, identify the subset of these concepts which constitute the central meaning of the texts and thus "determine" mostly their location scope.
- Finally, use these concepts in order to limit the number of possible locations that may appear within the text as well as the number of instances of the other evidential concepts.

For example, when building a location evidence model for texts that describe historical events, some concepts whose instances may act as evidence for locations expected to be found in these texts are related locations, historical events, and historical groups and persons that participated in these events. The most location-determining concept would be the Historical Event, so from all the possible locations, groups and persons we consider only those that are, directly or indirectly, related to some event. Indirectly means, for example, that while "Siege of Tripolitsa" is directly related to "Tripoli", it is indirectly related to Greece as well.

The result of the above process should be a location evidence mapping function $lem : C \rightarrow R^n$ which given an evidential concept $c \in C$ returns the relations $\{r_1, r_2, ..., r_n\} \in R^n$ whose composition links c's instances to locations. Table I shows such a mapping for the history domain and in particular about that of military conflicts.

Using this mapping function, we can then calculate the location meaning evidence function $lmef$ as follows. Given a location $l \in L$ and an instance $i \in I$, which belongs to some concept $c \in C$ and is related to l through the composition of relations $\{r_1, r_2, ..., r_n\} \in lem(c)$, we derive i) the set of instances $I_{amb} \subseteq I$ which share common names with i and ii) the set of locations $L_{amb} \subseteq L$ which share common names with l and are also related to i through

Table I
LOCATION EVIDENCE MAPPING FUNCTION FOR MILITARY CONFLICTS DOMAIN

Evidence Concept	Location Linking Relation(s)
Military Conflict	*tookPlaceAtLocation*
Military Conflict	*tookPlaceAtLocation, isPartOfLocation*
Military Person	*participatedInConflict, tookPlaceAtLocation*
Combatant	*participatedInConflict, tookPlaceAtLocation*
Location	*isPartOfLocation*

$\{r_1, r_2, ..., r_n\} \in lem(c)$. Then the value of the function $lmef$ for this location and this instance is:

$$lmef(l, i) = \frac{1}{|L_{amb}| \cdot |I_{amb}|} \tag{1}$$

The intuition behind this formula is that the evidential power of a given instance is inversely proportional to its own ambiguity as well as to the number of different target locations it provides evidence for. If, for example, a given military person has fought in many different locations with the same name, then its evidential power for this name is low. Similarly, if a given military person's name is very ambiguous (i.e., there are many persons with the same name) then its evidential power is also low.

Using the same equation we can also calculate the geographical scope evidence function $gsef$, the only difference being that we consider the set L'_{amb} that contains all the locations related to i, not just the ones with the same name as l:

$$gsef(l, i) = \frac{1}{|L'_{amb}| \cdot |I_{amb}|} \tag{2}$$

The intuition here is that the geographical scope-related evidential power of a given instance is inversely proportional to the number of different locations it is related to.

B. Geographical Entity Resolution

The geographical entity resolution process for a given text document and a location meaning evidence function works as follows. First, we extract from the text the set of terms T that match to some $i \in I$ along with a term-meaning mapping function $m : T \rightarrow I$ that returns for a given term $t \in T$ the instances it may refer to. We also consider I_{text} to be the superset of these instances.

Then, we consider the set of potential locations found within the text $L_{text} \subseteq I_{text}$ and for each $l \in L_{text}$ we derive all the instances from I_{text} that belong to some concept $c \in C$ for which $lem(c) \neq \emptyset$. Subsequently, by combining the location evidence model function $lmef$ with the term meaning function m we are able to derive a location-term meaning support function $sup_m : L_{text} \times T \rightarrow [0, 1]$ that returns for a location $l \in L_{text}$ and a term $t \in T$ the degree to which t supports l. If $l \in L_{text}$, $t \in T$ then

$$sup_m(l, t) = \frac{1}{|m(t)|} \cdot \sum_{i \in m(t)} lmef(l, i) \tag{3}$$

Using this function, we are able to calculate for a given term $t \in T$ in the text the confidence that it refers to location $l \in m(t)$:

$$c_{ref}(l) = \frac{\sum_{t_j \in T} K(l, t_j)}{\sum_{l' \in m(t)} \sum_{t_j \in T} K(l', t_j)} \cdot \sum_{t_j \in T} sup_m(l, t_j) \tag{4}$$

where $K(l, t) = 1$ if $sup_m(l, t) > 0$ and 0 otherwise.

In other words, the overall support score for a given candidate location is equal to the sum of the location's partial supports (i.e., function sup_m) weighted by the relative number of terms that support it. It should be noted that in the above process, we adopt the one referent per discourse approach which assumes one and only one meaning for a location in a discourse.

C. Geographical Scope Resolution

The process of geographical scope resolution is similar to the entity resolution one, the difference being that we consider as candidate scope locations not only those found within the text but practically all those that are related to instances of the evidential concepts in the ontology. In that way, even if a location is not explicitly mentioned within the text, it still can be part of the latter's scope. More specifically, given a text document and a geographical scope evidence function $gsef$ we first consider as candidate locations all those for which there is evidence within the text, that is all those for which $gsef(l, i) > 0, l \in L, i \in I_{text}$. We call this set L_{cand}. Then, for a given $l \in L_{cand}$ we compute the scope related support it receives from the terms found within the text as follows:

$$sup_s(l, t) = \frac{1}{|m(t)|} \cdot \sum_{i \in m(t)} gsef(l, i) \tag{5}$$

Finally, we compute the confidence that l belongs to the geographical scope of the text in the same way as Equation 4 but with sup_s substituting sup_m:

$$c_{scope}(l) = \frac{\sum_{t_j \in T} K(l, t_j)}{\sum_{l' \in L_{cand}} \sum_{t_j \in T} K(l', t_j)} \cdot \sum_{t_j \in T} sup_s(l, t_j) \tag{6}$$

where $K(l, t) = 1$ if $sup_s(l, t) > 0$ and 0 otherwise.

IV. EXPERIMENTAL EVALUATION

To illustrate the effectiveness of our proposed framework we performed two experiments on historical texts describing military conflicts. In the first experiment, we focused on correctly resolving ambiguous location references within the texts while in the second, on correctly determining the texts' geographical scope. In both cases, we built a common location evidence model using an appropriate ontology, derived from DBPedia, comprising about 4120 military conflicts, 1660 military persons, 4270 locations, 890 combatants and, of course, the relations between them (conflicts with locations, conflicts with persons etc.). The model's location evidence mapping function was that of Table I and it was used to calculate the evidential functions $lmef$ and $gsef$ for all pairs of locations and evidential entities (other locations, conflicts, persons and combatants).

Table II shows a small sample of these pairs where, for example, James Montgomery acts as evidence for the disambiguation of Beaufort County, South Carolina because he's fought a battle there. Moreover, his evidential power for that location is 0.25, practically because there are 3 other military persons in the ontology also named Montgomery. Similarly, Pancho Villa acts as evidence for the consideration of Columbus, New Mexico as the scope of a text (because he's fought a battle there) and his evidential power for that is 0.2 since, according to the ontology, he's fought battles in 4 other locations as well.

Table II
EXAMPLES OF LOCATION EVIDENTIAL ENTITIES

Location	Evidential Entity	lmef	gsef
Columbus, Georgia	James H. Wilson	1.0	0.17
Columbus, New Mexico	Pancho Villa	1.0	0.2
Beaufort County, South Carolina	James Montgomery	0.25	0.25

Using this model, we first applied our proposed geographic entity resolution process in a dataset of 50 short texts describing military conflicts. All texts contained ambiguous location entities but little other geographical information and, in average, each ambiguous location reference had 2.5 possible interpretations. For each such reference, we determined its possible interpretations and ranked them using the confidence score derived from Equation 4. We then measured the effectiveness of the process by determining the number of correctly interpreted location references, namely references whose highest ranked interpretation was the correct one.

Table III shows results achieved by our approach compared to those achieved by some well-known publicly available semantic annotation and disambiguation services, namely DBPedia Spotlight [14], Wikimeta [15], Zemanta [16], AlchemyAPI [17] and Yahoo! [18]. As one can see, the consideration of non-geographical semantic information that our approach enables, manages to significantly improve the effectiveness of the geographical entity resolution task.

For the second experiment, we applied our proposed geographic scope resolution process in two different datasets, all comprising 50 short military conflict related texts but with different characteristics. The first dataset comprised texts whose geographical scope was not explicitly mentioned within them and which contained little other geographical information. The second dataset comprised texts whose geographical scope related locations were explicitly and unambiguously mentioned within them but along with other geographical entities that were not part of this scope.

Table III
GEOGRAPHICAL ENTITY RESOLUTION EVALUATION RESULTS

System/Approach	Effectiveness
Proposed Approach	72%
DBPedia Spotlight	54%
Wikimeta	33%
Zemanta	26%
AlchemyAPI	26%
Yahoo!	24%

In both cases, we determined for each text the possible locations that comprised its geographical scope and ranked them using the confidence score derived from equation 6. We then measured the effectiveness of the process by determining the number of correctly scope resolved texts, namely texts whose highest ranked scope locations were the correct ones. As a baseline, we compared our results to the ones derived from Yahoo! Placemaker [19] geoparsing web service.

The results of the above process are shown in Table IV. As one can see, the improvement our method achieves in the effectiveness of the scope resolution task is quite significant in both datasets and especially in the first one where the scope-related locations are not explicitly mentioned within the texts. This verifies the central idea of our approach that non-geographical semantic information can significantly improve the geographical scope resolution process and in particular the subtasks of:

1) Inferring relevant to the text's geographical scope locations even in the absence of explicit reference of them within the text (first dataset).
2) Distinguishing between relevant and non-relevant to the text's geographical scope locations, even in the presence of non-relevant location references within the text (second dataset).

Table IV
GEOGRAPHICAL SCOPE RESOLUTION EVALUATION RESULTS

System/Approach	Dataset 1	Dataset 2
Proposed Approach	70%	85%
Yahoo! Placemaker	18%	30%

V. CONCLUSION

In this paper, we proposed a novel framework for optimizing geographical entity and scope resolution in texts by means of domain and application scenario specific non-geographical semantic information. First, we described how, given a priori knowledge about the domain(s) and expected content of the texts that are to be analyzed, one can define a model that defines which and to what extent semantic entities (especially non-geographical ones) can be used as contextual evidence indicating two things:

- Which is the most probable meaning of an ambiguous location reference within a text (geographical entity resolution task).
- Which locations constitute the geographical scope of the whole text (geographical scope resolution task).

Then, we described how such a model can be used for the two tasks of geographical entity and scope resolution by providing corresponding processes. The effectiveness of these processes was experimentally evaluated in a comprehensive and comparative to other systems way. The evaluation results verified the ability of our framework to significantly improve the effectiveness of the two resolution tasks by exploiting non-geographical semantic information.

It should be noted that our proposed framework is not meant as a substitute or rival of other geographical resolution approaches (that operate in open domains, use geographical information and relevant heuristics and apply machine learning and statistical methods) but rather as a complement of them in application scenarios where text domain and content are a priori known and comprehensive domain ontological knowledge is available (as in the case of historical texts used in our experiments). In fact, given these two requirements for our approach's applicability, future work will focus on investigating how statistical and machine learning approaches may be used, in conjunction with our approach, in order to i) automatically build geographical resolution evidence models based on text corpora and ii) deal with cases where available domain semantic information is incomplete.

ACKNOWLEDGMENT

This work was supported by the European Commission under contract FP7- 248984 GLOCAL.

REFERENCES

[1] A. M. Kaplan and M. Haenlein, "Users of the world, unite! the challenges and opportunities of social media," *Business Horizons*, vol. 53, no. 1, pp. 59–68, 2010.

[2] C. B. Jones and R. S. Purves, "Geographical information retrieval," *Int. J. Geogr. Inf. Sci.*, vol. 22, no. 3, pp. 219–228, Jan. 2008.

[3] J. Raper, G. Gartner, H. Karimi, and C. Rizos, "Applications of location-based services: a selected review," *J. Locat. Based Serv.*, vol. 1, no. 2, pp. 89–111, Jun. 2007.

[4] C. Bizer, T. Heath, and T. Berners-Lee, "Linked data - the story so far," *Int. J. Semantic Web Inf. Syst.*, vol. 5, no. 3, p. 122, 2009.

[5] G. Andogah, G. Bouma, J. Nerbonne, and E. Koster, "Place-name ambiguity resolution," in *Methodologies and Resources for Processing Spatial Language (Workshop at LREC 2008)*, 2008.

[6] G. Andogah, G. Bouma, and J. Nerbonne, "Every document has a geographical scope," *Data and Knowledge Engineering*, 2012.

[7] Q. Mei, C. Liu, H. Su, and C. Zhai, "A probabilistic approach to spatiotemporal theme pattern mining on weblogs," in *Proceedings of the 15th international conference on World Wide Web*, ser. WWW '06. New York, NY, USA: ACM, 2006, pp. 533–542.

[8] C. Wang, J. Wang, X. Xie, and W.-Y. Ma, "Mining geographic knowledge using location aware topic model," in *Proceedings of the 4th ACM workshop on Geographical information retrieval*, ser. GIR '07. New York, NY, USA: ACM, 2007, pp. 65–70.

[9] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Mar. 2003. [Online]. Available: http://dl.acm.org/citation.cfm?id=944919.944937

[10] J. Xu and W. B. Croft, "Query expansion using local and global document analysis," in *Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval*, ser. SIGIR '96. New York, NY, USA: ACM, 1996, pp. 4–11.

[11] J. Bhogal, A. Macfarlane, and P. Smith, "A review of ontology based query expansion," *Inf. Process. Manage.*, vol. 43, no. 4, pp. 866–886, Jul. 2007.

[12] T. Kauppinen, R. Henriksson, R. Sinkkil, R. Lindroos, J. Vtinen, and E. Hyvnen, "Ontology-based disambiguation of spatiotemporal locations," in *Proceedings of the 1st international workshop on Identity and Reference on the Semantic Web (IRSW2008), 5th European Semantic Web Conference 2008 (ESWC 2008)*. Tenerife, Spain: CEUR Workshop Proceedings, ISSN 1613-0073, June 1-5 2008.

[13] Y. Peng, D. He, and M. Mao, "Geographic named entity disambiguation with automatic profile generation," in *Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence*, ser. WI '06. Washington, DC, USA: IEEE Computer Society, 2006, pp. 522–525.

[14] "Dbpedia spotlight," http://dbpedia.org/spotlight, accessed: 05/07/2012.

[15] "Dbpedia spotlight," http://www.wikimeta.com, accessed: 05/07/2012.

[16] "Zemanta," http://www.zemanta.com, accessed: 05/07/2012.

[17] "Alchemy api," http://www.alchemyapi.com, accessed: 05/07/2012.

[18] "Yahoo!" http://developer.yahoo.com/search/ content/V2/contentAnalysis.html, accessed: 05/07/2012.

[19] "Yahoo! placemaker," http://developer.yahoo.com/geo/placemaker/, accessed: 05/07/2012.