On Metonymy Recognition for Geographic IR

Johannes Leveling
FernUniversität in Hagen (University of Hagen)
Intelligent Information and Communication Systems (IICS)
Hagen, Germany
johannes.leveling@fernuni-hagen.de

Sven Hartrumpf
FernUniversität in Hagen (University of Hagen)
Intelligent Information and Communication Systems (IICS)
Hagen, Germany
sven.hartrumpf@fernuni-hagen.de

ABSTRACT
Metonymic location names refer to other, related entities and possess a meaning different from the literal, geographic sense. Metonymic names are to be treated differently to improve performance of geographic information retrieval (GIR). This paper presents a method for disambiguating location names in textual information to distinguish literal and metonymic senses, based on shallow features.

The evaluation of this method is two-fold: First, we use a memory based learner to train a classifier and determine standard evaluation measures such as F-score and accuracy. Second, we perform retrieval experiments based on the GeoCLEF data (newspaper article corpus and queries) from 2005. We compare searching for location names in an index containing their literal and metonymic sense with searching in an index containing literal senses only. Evaluation results indicate that, using a large annotated corpus of location names, a classifier based on shallow features achieves adequate performance, and removing metonymic senses from a database index yields a higher performance for GIR.

Categories and Subject Descriptors
H.3.1 [**Content Analysis and Indexing**]: Indexing methods, Linguistic Processing; H.3.3 [**Information Search and Retrieval**]: Query formulation, Selection Process

General Terms
Metonymy, GIR

1. INTRODUCTION
The identification of location names in textual information is an essential task for a geographic information retrieval (GIR) application. Metonymic location names refer to other, related entities. They inherently possess a meaning different from the literal, geographic sense and are to be treated differently for retrieval to improve precision and recall.

This paper presents a method for disambiguating literal and metonymic senses of location names in textual information. The disambiguation is based on a machine learning classifier that relies on shallow features. Our investigations focus on metonymy in German documents, but there is evidence that the general method is language-independent and applicable to queries as well.

The evaluation of this method is two-fold: First, we use a memory based learner to train the classifier and determine standard evaluation measures such as F-score and accuracy. Second, we perform retrieval experiments based on the GeoCLEF task in 2005. We present results for retrieval experiments searching for location names in an index containing their literal and metonymic senses with searching in an index containing location names classified as their literal sense. Evaluation results indicate that, using a large annotated corpus of location names, a classifier based on shallow features achieves adequate performance, and removing metonymic senses from a database index yields a higher performance for geographic queries.

Metonymy is typically defined as a figure of speech in which a speaker uses *one entity to refer to another that is related to it* [7]. In textual information, metonymic proper nouns reduce performance for information retrieval (IR) applications in general and for GIR in particular, because metonymic and literal senses are not distinguished. Most traditional approaches to IR do not even distinguish between proper nouns and other parts of speech, or between nouns and proper nouns. The identification of metonymy in textual information will help to improve precision for GIR. Literal and metonymic senses will have to be indexed and queried in a novel way by a IR system.

2. RELATED WORK
Most research on metonymic use is on English [17, 12, 14], often based on the BNC (British National Corpus). There has been comparably less research for German. Experiments for German involved a small corpus of German product reviews, for which senses of proper names (countries and companies) were manually annotated [9].

Markert and Nissim have introduced a classification of regular metonymic patterns for organizations and locations. They extracted a set of metonymic proper nouns from the BNC words from two categories: country names [11] and organization names [13].

For the country names, Markert and Nissim distinguished the main metonymy classes `place-for-event`, `place-for-people` (with subclasses), and `place-for-product`. In addition, classes for the `literal` sense, for `mixed` senses and for metonymies not covered by the regular patterns (`othermet`) were defined.

Peirsman presents results for experiments with a machine learning approach to identifying regular metonymy for organizations and countries. He argues that a large annotated corpus is not needed because the number of training instances can be drastically reduced [14].

Schilder, Versley, and Habel [16] describe experiments in tagging spatial expressions, based on the German CoNLL corpus. Their set of tags includes a single form of metaphoric usage (corresponding to the subclass `CapGov`, see Section 3.1).

3. TRAINING THE LOCATION CLASSIFIER

3.1 Metonymy Classes for Locations

For the application in GIR, we focus on the distinction of `literal` and `metonymic` senses of location names. The `literal` sense refers to a location as an unmovable entity. This includes descriptions of a location, its properties (area, inhabitants, etc.), and geographic aspects.

The `metonymic` sense of a location covers readings that follow a regular pattern as well as metonymies that do not. We follow the classification suggested by Markert and Nissim:

- `literal` (literal, geographic sense), e.g. *The finals of the FIFA championship 2006 takes place in Berlin./ Das Endspiel der FIFA Weltmeisterschaft 2006 findet in Berlin statt.*
- `metonymic`
  - `place-for-event` (location referring to an event), e.g. *Korea turned out to be a military catastrophe for the USA./ Korea stellte sich als militärische Katastrophe für die USA heraus.*
  - `place-for-people` (location for people at location),
    - `CapGov` (capital city for government), e.g. *Yesterday, Seoul and Peking agreed to start diplomatic relations./Gestern sind sich Seoul und Peking einig geworden, diplomatische Beziehungen aufzunehmen.*
    - `Off` (location referring to official administration), e.g. *Hamburg has decided to expand the harbor./Hamburg hat entschieden, den Hafen zu erweitern.*
    - `Org` (location for organization at location), e.g. *After tying the score, Bayern pushed for a win with ten players only./Nach dem Ausgleich drängte Bayern mit nur zehn Spielern auf den Sieg.*
    - `Pop` (location for population), e.g. *Germany is threatened with extinction by the declining birth-rate./Deutschland ist durch rückläufige Geburtenrate vom Aussterben bedroht.*
  - `place-for-product` (place for product from that place), e.g. *Wine connoisseurs know that Chianti has to be decanted before drinking./Weinkenner wissen, dass Chianti vor dem Trinken dekantiert werden muss.*
  - `othermet` (metonymy not covered by the regular patterns)
- `mixed` (reference to both literal and metonymic sense), e.g. *A major part of the food originates from Ukraine, which decided to increase wheat export even more./Ein großer Teil der Nahrungsmittel stammt aus der Ukraine, die entschieden hat, den Weizenexport weiter zu erhöhen.*

Note that naively assigning a default sense to a location name would be highly corpus-dependent. For instance, in newspaper articles city names are often used metonymically for sports events (e.g. *Barcelona*) or sports clubs (e.g. *Oakland*). Contrariwise, descriptions of tourist attractions for travel information services would primarily contain literal senses of location names.

3.2 Data and Annotation

After tokenizing and splitting a text into sentences, location names are identified by a lookup in several name lexicons for different types of names (e.g., cities, countries, islands, lakes, regions, rivers).

We started annotating a subset of the German CoNLL-2003 corpus (test set A) using all subclasses introduced in the previous subsection. Table 1 gives an overview over the distribution of literal and metonymic senses in the CoNLL subset for location and organization names. First experimental results for a classifier trained on this subset were discouraging. A statistics on metonymic classes revealed that for some metonymic classes there was too little training data to set up an automatic approach. For GIR, we broadened the classification into `literal` and `metonymic` senses. The latter covers all subclasses. `Mixed` senses were removed before training the classifier.

The CoNLL corpus contains tagging errors and textual encoding errors. Automatic tagging introduces errors which will decrease the classifier performance (see [10]). Therefore, instead of continuing the annotation of the CoNLL corpus, we employed a deep linguistic analysis based on WOCADI, a disambiguating parser [5]. We employed WOCADI to preselect sentences from the GeoCLEF news corpus (short: News) that contain a location name. Furthermore, the selectional constraints (of verbs and nouns) applied by the parser allowed to retrieve sentences that probably contain metonymic uses of location names.

WOCADI is able to spot possible conventional metonymy by one of two methods: identifying violations of selectional restrictions or identifying references to a meaning facet of a concept. In our context, a selectional restriction (of a verb or noun) is violated, if a location name is used as a complement (of the verb or noun) that must be realized by a person or institution. For example, *to plan something* requires a person or institution in the subject role. Thus, *Berlin* is used metonymically in the sentence *Berlin plans to reduce taxes.*

Many locations (such as cities or countries) have certain meaning facets: an institutional facet, a geographic facet, etc. (See [6] for details on meaning facets in the computer lexicon used by WOCADI.) If a single facet other than the expected geographic facet is referenced by a location name, it is a possible metonymy. For example, *Pekings* institutional facet is referenced in the sentence *Seoul started diplomatic relations with Peking.*

The training data for the classifier consists of two corpora: a subset of the German CoNLL-2003 Shared Task corpus for Language-Independent Named Entity Recognition (II) [15] and the sentences extracted from the news corpus for the GeoCLEF task [3].

Table 1: Statistics on German CoNLL-2003 data (test set A) for locations (LOC) and organizations (ORG, given for comparison only).

sense	LOC		ORG	
literal	894	(75.06%)	469	(43.11%)
metonymic	203	(17.05%)	419	(38.51%)
mixed	94	(7.89%)	200	(18.38%)
total	1191		1088	

3.3 Features and Model

Each occurrence of a location name is represented as a feature vector, assigning features automatically, to provide a method for annotating a large corpus with an automatic tool. Table 2 gives an overview over the features that were taken into account. Note that no grammatical or semantic information is employed. However, features for verbs, modal verbs, and auxiliary verbs are captured separately to ensure some correlation with subject/object position of locations and active/passive voice. Furthermore, part-of-speech information was determined for closed word categories by a list lookup to save time in the annotation of the newspaper corpus. (An obvious improvement would be to use a part-of-speech tagger.)

For each location name a feature vector consisting of features for tokens in a context window is computed and the classifier is trained on these data with TiMBL [2], an implementation of memory based learning, employing the IB1 algorithm [1]. Peirsman argues for using TiMBL [14] for two reasons. First, it models human learning by interpreting new phenomena based on representations of known phenomena. Second, the classifier is suitable for automatically extracted data and requires no manual interaction (e.g., the formulation of rules) or pre-computed probabilities.

3.4 Evaluation of the Location Classifier

The machine learning tool TiMBL produced a classifier that achieved the performance shown in Table 3 (evaluation was performed with *leave-one-out*). The table shows the number of instances (I), the number of metonyms in the set of instances (M), accuracy (A), and the F_1-score (F). It is important to stress that the results were achieved using shallow features, which allows for quick processing of large text corpora instead of relying on more time-consuming steps such as parsing.

Table 2: Subset of features for instances. Vectors contain features for up to 4 main verbs.

Feature	Description and values
f1	part-of-speech from STTS tagset for closed word categories (ART, APPO, KON, ...)
f2	word length (1,2,3, ...)
f3	word prefix of length 3 (*anh*, *bez*, ...)
f4	word suffix of length 3 (*ion*, *ung*, ...)
f5	case information (lowercase, uppercase, punctuation, numeric)
⋮	
f10	sentence length (1,2,3, ...)
f11-f14	lemma of verb (*planen*, *besuchen*, ...)
f15-f18	absolute position of verb (1,2,3, ...)
f19-f22	position of verb relative to current token (left, right)
⋮	

Table 3: Performance of the metonymy classifier for location names (LOC) using 113 features per instance for two classes.

corpus	I	M	A	F
CoNLL set A	1097	203	0.846	0.842
News	2154	1270	0.810	0.810
CoNLL set A, News	3251	1473	0.817	0.817

Table 4: Confusion matrix for metonymy classification.

	predicted sense	
actual sense	LIT	MET
LIT	1409	289
MET	303	1170

4. APPLYING THE LOCATION CLASSIFIER TO GIR

To our knowledge, there has been no exhaustive research on how metonymic uses of location names influence performance in GIR, yet. We performed retrieval experiments following our approach for GeoCLEF 2005, adapting traditional information retrieval to GIR, but keeping a separate index for location names. The experimental setup is described in more detail in [8].

Experiments were performed with the Zebra database [4], employing a standard IR model (tf-idf). Indexes for location names were created in three ways: location names with literal and metonymic sense (LOCALL), location names in literal sense (LOCLIT), and location names with metonymic sense only (LOCMET). All other words were indexed and queried for as normal text.

Table 5: Performance for retrieval experiments with queries from the GeoCLEF retrieval tasks 2005.

queries	MAP	recall
GeoCLEF 2005 (LOCALL)	0.0857	0.645
GeoCLEF 2005 (LOCLIT)	0.0917	0.642

Documents in the GeoCLEF collection [3] consist of several sentences. Excluding a single metonymic name from the index means that the in-term-frequency is lowered and not necessarily zero.

The results shown in Table 5 indicate that performance is improved when querying the index containing only the literal senses for location names. Computation of recall values is based on official relevance assessments of documents. Results differ from the official results published because the textual description of GeoCLEF topics (narrative) was employed for constructing a query in addition to topic title and topic description.

A significance test was supposed to show whether the improvement in mean average precision (MAP) was due to chance. We conducted the sign test for small samples ($N = 18$, $H_0 = 0.5$) and calculated the *p-value* (0.0471). Improvements in performance seem marginally significant for several reasons so far:

- The set of queries examined is too small for significance testing, yet. Including queries from GeoCLEF 2006 for testing will return more useful results. (Relevance assessments were not available at the time of writing.)
- Relevance assessments from GeoCLEF may not include judgements for newly found relevant documents.
- The approach to GIR was an ad-hoc adaptation of IR methods and the underlying database management system employs a standard IR model (tf-idf).

5. CONCLUSION AND FUTURE WORK

In this paper we have explored the performance of a classifier which disambiguates literal and metonymic senses of location names. The classifier achieved an adequate performance using a set of shallow features only. It does not rely on any kind of semantic preprocessing and is therefore applicable to large text corpora. Furthermore, we performed different geographic IR experiments. One type of experiment used a traditional approach in which an index with location names with both literal and metonymic senses was queried. In another type of experiment, an index containing location names classified in their literal sense was queried. Results indicate that disambiguating literal and metonymic senses slightly improves precision.

There will be a growing need to successfully identify senses of location names. The described location classifier could be compared to a deeper method which employs WOCADI to parse document sentences. While our retrieval experiment already showed a slight improvement in performance by excluding all metonymic senses for querying, more complex approaches should be investigated. For example, term weights may have to be adjusted on the basis of the metonymic subclass and metonymy in queries could be treated differently.

6. ACKNOWLEDGMENTS

Research described in this paper was in part funded by the DFG (Deutsche Forschungsgemeinschaft) in the project IRSAW (Intelligent Information Retrieval on the Basis of a Semantically Annotated Web). Additional thanks to Tom Hombergs and Dirk Veiel.

7. REFERENCES

[1] D. W. Aha, D. Kibler, and M. K. Albert. Instance-based learning algorithms. *Machine Learning*, 6(1):37–66, 1991.

[2] W. Daelemans, J. Zavrel, K. van der Sloot, and A. van den Bosch. TiMBL: Tilburg memory based learner, version 5.1, reference guide. Technical Report 04-02, ILK, 2004.

[3] F. Gey, R. Larson, M. Sanderson, H. Joho, and P. Clough. GeoCLEF: The CLEF 2005 cross-language geographic information retrieval track. In C. Peters, F. C. Gey, J. Gonzalo, G. J. F. Jones, M. Kluck, B. Magnini, H. Müller, and M. de Rijke, editors, *Accessing Multilingual Information Repositories: 6th Workshop of the Cross-Language Evaluation Forum, CLEF 2005, Vienna, Austria*, LNCS. Springer, Berlin, 2006. In print.

[4] S. Hammer, A. Dickmeiss, H. Levanto, and M. Taylor. *Zebra – User's Guide and Reference*. Copenhagen, Denmark, 2005.

[5] S. Hartrumpf. *Hybrid Disambiguation in Natural Language Analysis*. Der Andere Verlag, Osnabrück, Germany, 2003.

[6] S. Hartrumpf, H. Helbig, and R. Osswald. The semantically based computer lexicon HaGenLex – Structure and technological environment. *Traitement automatique des langues*, 44(2):81–105, 2003.

[7] G. Lakoff and M. Johnson. *Metaphors we live by.* Chicago University Press, 1980.

[8] J. Leveling, S. Hartrumpf, and D. Veiel. Using semantic networks for geographic information retrieval. In C. Peters, F. C. Gey, J. Gonzalo, G. J. F. Jones, M. Kluck, B. Magnini, H. Müller, and M. de Rijke, editors, *Accessing Multilingual Information Repositories: 6th Workshop of the Cross-Language Evaluation Forum, CLEF 2005, Vienna, Austria*, LNCS. Springer, Berlin, 2006. In print.

[9] K. Markert and U. Hahn. Understanding metonymy in discourse. *Artificial Intelligence*, 135(1–2):145–198, 2002.

[10] K. Markert and M. Nissim. Metonymy resolution as a classification task. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2002)*, Philadelphia, USA, 2002.

[11] K. Markert and M. Nissim. Towards a corpus for annotated metonymies: the case of location names. In *Proceedings of the 3rd international Conference on Language Resources and Evaluation (LREC 2002)*, Las Palmas, Spain, 2002.

[12] D. Maynard, K. Bontcheva, and H. Cunningham. Towards a semantic extraction of named entities. In *Recent Advances in Natural Language Processing (RANLP 2003)*, Bulgaria, 2003.

[13] M. Nissim and K. Markert. Learning to buy a Renault and talk to BMW: a supervised approach to conventional metonymy. In *International Workshop on Computational Semantics (IWCS2005)*, Tilburg, The Netherlands, 2005.

[14] Y. Peirsman. Example-based metonymy recognition for proper nouns. In *Proceedings of the Student Research Workshop of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL-2006)*, pages 71–78, Trento, Italy, 2006.

[15] E. T. K. Sang and F. D. Meulder. Introduction to the CoNLL-2003 shared task: Language independent named entity recognition. In W. Daelemans and M. Osborne, editors, *Proceedings of CoNLL-2003*, pages 142–147, Edmonton, Canada, 2003.

[16] F. Schilder, Y. Versley, and C. Habel. Extracting spatial information: grounding, classifying and linking spatial expressions. In *Proceedings of the ACM SIGIR Workshop on Geographic Information Retrieval*, Sheffield, UK, 2004.

[17] M. Utiyama, M. Murata, and H. Isahara. A statistical approach to the processing of metonymy. In *COLING-2000*, pages 885–891, 2000.