Создание и использование многоязычного корпуса объектно-ориентированных топонимических текстов для оптимизации задачи автоматического генерирования описания изображений

Development and implementation of multilingual object type toponym-referenced text corpora for optimizing automatic image description generation

Gornostay T. (tatjana.gornostaja@tilde.lv)
Tilde, Riga (www.tilde.com)

Aker A. (a.aker@dcs.shef.ac.uk)
Department of Computer Science, University of Sheffield

С точки зрения обработки стремительного роста объема графической информации в сети Интернет целесообразна разработка автоматических методов генерирования ее описаний. В последнее время используется метод автоматического реферирования набора документов определенной тематики. В настоящей статье описывается подход к созданию и использованию многоязычного корпуса объектно-ориентированных топонимических текстов для четырех языков (английского, немецкого, итальянского и латышского) в контексте оптимизации задачи автоматического реферирования для генерирования описаний графических изображений топонимов.

1. Introduction

In recent years the number of images on the Web has experienced an immense growth facilitated by the development of affordable digital hardware and the availability of online image sharing social sites. For successful indexing and retrieving the available images, their content has to be correctly identified and annotated. One way of annotating images is by tagging them with image descriptions.

Image description, or *summary*, is a more general term for *image index* (Hollink et al., 2004) which is used broadly in both image indexing and retrieval disciplines. Image descriptions can contain miscellaneous information about an image, including generic, specific, and interpretative explanation of what is shown in the image (Shatford, 1986). Apart from being essential for automatic indexing and retrieval of images, image descriptions are also useful for human users to get information about the content of the image. For example, descriptions of images showing locations could help a user who seeks information about a certain place, or a journalist, who writes an article about a location, or a tourist who looks for interesting places to visit (Aker and Gaizauskas, 2008).

Image descriptions can be written manually by individuals who are specially hired for this purpose or describe images for their private usage. However, manual generation of image descriptions is a tedious, expensive and inaccurate task (Pan et al., 2004; Jamieson et al., 2007). Therefore, methods of automatizing image description generation have been developed recently.

There are different approaches to automatic generation of image descriptions (Mori et al., 2000; Barnard and Forsyth, 2001; Duygulu et al., 2002; Barnard et al., 2003; Pan et al., 2004; Deschacht and Moens, 2007; Feng and Lapata, 2008). All these approaches generate image captions based on texts associated with the image in combination with or without analysis of image features (color, shape, texture). The resulting image descriptions contain named entities (e.g. person names), and/or a set of open class words (nouns, verbs, adjectives and adverbs) describing the image.

Another approach of image captioning is that of Aker and Gaizauskas (2008) who apply generic and

query-based multi-document summarization techniques to generate image descriptions for toponym-referenced images. *Toponyms* are terms describing places, such as *Westminster Abbey*, *University of Sheffield*, etc. *Toponym-referenced images* are images tagged with toponyms. Image descriptions generated automatically for toponym-referenced images can be called a *toponym-referenced description*, or *summary*.

In contrast to other researchers, who assume that the image has an associated text with it, Aker and Gaizauskas (2008) generate toponym-referenced descriptions from multiple web documents containing toponym-referenced texts which describe the places reflected in the image. The web documents are retrieved using the toponyms of the image.

Aker and Gaizauskas (2008) have shown that query-based toponym-referenced descriptions outperform generic descriptions. Query-based summaries are generated by biasing the summarizer towards the query which is the set of toponyms associated with the image. This makes sure that sentences which contain the query (toponyms) are more highly scored than the ones which do not contain any query term. In contrast, for generation of generic summaries the query (toponym) is not used to bias the summarizer. Therefore, the sentences containing the toponym are not necessarily scored more highly than those which do not contain any query term.

Although toponym-referenced descriptions generated by query-based summarizer were better than those generated by the generic summarizer, the authors noted that the best image descriptions were still far from perfect. The evaluation showed that the agreement between query-based descriptions and human generated summaries was not satisfactory. In contrast, the agreement between descriptions generated by humans for each image was high. This observation allowed the authors to hypothesize that humans have some conceptual model of what is salient regarding a certain *object type* of a toponym (churches, bridges, etc.). The authors suggest collecting existing textual resources about object types as one way of capturing these conceptual models. More specifically, they propose to compile a corpus of toponym-referenced texts for each object type and use the information commonly associated with each object type in the corpus to bias the summarizer. For example, a collection of texts describing object type *church* would contain information about the age of the church, dates of construction, the architectural style, its height, etc. This object type specific information can be used by the summarizer to give higher score to sentences which also contain this information.

In this paper we describe a possible way to derive such object type text corpora from the Web for different languages. We use Wikipedia as a resource and demonstrate how each Wikipedia article can be categorized automatically by object type for English. We also show how such a corpus can be extended to further languages. We are in particular interested in four languages: English, German, Italian and Latvian, for which we collect object type corpora. Finally, we exemplify how such object type corpora can be used for automatic multi-document summarization.

In Section 2 we define our requirements, give an overview of existing text resources on the Web and argue for our choice of Wikipedia as the most suitable resource to derive an object type corpus. Wikipedia is described in Section 3. Section 4 focuses on the procedure of automatic categorization of Wikipedia articles by object types and reports on evaluation of our categorization procedure. We also describe how object type corpora of German, Italian and Latvian can be obtained from the English object type corpus. In Section 5 we explain how object type corpora can be used for automatic generation of image descriptions using multi-document summarization. Section 6 concludes the paper and outlines future directions.

2. Corpus development scenarios

Object type text corpora that we aim to build for our target languages are collections of toponym-referenced texts for single object types (churches, bridges, etc.) compiled from texts available on the Web. The Web is generally accepted to be the most suitable and helpful resource for corpus development (Kilgarriff and Grefenstette, 2003; Cheng et al., 2004; Liu and Curran, 2006; Bernardini et al., 2006; Kilgarriff, 2007). Its main advantage is that a large amount of textual data about locations is available in electronic form and multiple languages and can be efficiently searched and processed. A particularly important aspect behind our decision to use the Web is that one of our target languages, Latvian, is considered to be an under-resourced language with few corpus resources available, e.g. the JRC-Acquis corpus (Steinberger et al., 2006) and Contemporary Latvian language text corpus[1]. However, none of the currently existing resources is rich in toponym-referenced descriptions of places. Using the Web as a resource for corpus development is therefore a way to avoid data scarcity problems that would arise from lack of toponym-referenced texts in existing Latvian corpora.

Collecting a corpus from the Web is a challenging task since the Web is unstructured, not systematically organized, and does not have any definite directory (Kuo and Yang, 2004; Liu and Curran, 2006: 234). Therefore, searching domain-specific text resources on the Web often appears to be a low precision activity (much invaluable and unnecessary information can be obtained) (Kuo and Yang, 2004; Hughes, 2006). In addition, the following observation is relevant in our case with respect to Latvian: "*finding relevant*

[1] www.korpuss.lv

materials for low density <…> languages on the Web is in general an increasingly inefficient exercise even for experienced searchers" (Hughes, 2006). To alleviate these disadvantages, we use Wikipedia as a single resource for derivation of toponym-referenced text corpora from the Web. Wikipedia is a well structured Web resource, rich in location descriptions and available in multiple languages, including our four target languages. We therefore consider it particularly suitable for our purposes.

3. Wikipedia as a corpus

Wikipedia is a free multilingual encyclopedia project by the non-profit Wikimedia Foundation[2]. With 11 million articles written in different languages it is currently the largest and most popular general reference work on the Web. Altogether 265 languages are represented in Wikipedia. Different languages have different growth rates. In 2006 there were 2 languages with no article other than the main page, 5 languages have about 100 articles, and 25 languages less than 100 articles (Adafre and Rijke, 2006). The remaining languages are now represented by more than 100 articles. For example:

- English 2.7 Million
- German, Spanish, French, Italian, Polish: 300,000
- Esperanto, Catalan, Ukrainian: 100,000
- Bulgarian, Estonian, Lithuanian: 50,000;
- Latin, Macedonian: 20,000.

As these numbers show, the overwhelming part of Wikipedia is more or less developed and constitutes a rich resource for study of language. It has recently been successfully used for a number of natural language processing tasks like deriving a large scale taxonomy (Ponzetto and Strube, 2007), named entity recognition and translation in question answering (Bouma et al., 2006), named entity disambiguation, translation and transliteration (Wentland et al., 2008), domain specific query translation in multilingual information access (Jones et al., 2008) among others.

Wikipedia has at least three features that can be of use in corpus development: redirection pages, disambiguation pages, and internal links.

Redirection pages are used to normalize different variants or synonyms of a given concept (Bouma et al., 2006; Wentland et al., 2008), e.g. *Rīga* and *Rīgas pilsēta* — the capital of Latvia. Figure 1 shows an example of a redirection page. The redirection to the main page *Rīga* is given as: #REDIRECT [[Rīga]].

Disambiguation pages illustrated in Figure 2 are used to disambiguate homonyms (Wentland et al., 2008), e.g. *Rīga* as Latvia's capital, sport team, airport, cinema, etc.

[2] http://en.wikipedia.org/wiki/Wikipedia

```
- <page>
    <title>Rīgas pilsēta</title>
    <id>66353</id>
  - <revision>
      <id>416001</id>
      <timestamp>2008-10-07T09:21:13Z</timestamp>
    - <contributor>
        <username>Juzeris</username>
        <id>23</id>
      </contributor>
      <comment>Pāradresē uz [[Rīga]]</comment>
      <text xml:space="preserve">#REDIRECT [[Rīga]]</text>
```

Figure 1. Wikipedia redirection page from the article about Rīgas pilsēta to the article about Rīga in XML format

Rīga (nozīmju atdalīšana)

Vikipēdijas raksts

Rīga var būt:

- Rīga, Latvijas galvaspilsēta
- sporta komandas:
  - ASK Rīga, dažādas komandas
  - FK Rīga, futbola komanda
  - HK Rīga 2000, hokeja komanda
- Arēna Rīga
- Starptautiskā lidosta "Rīga"
- Kinoteātris "Rīga"
- Rīga, mazā planēta, atklāta 1966
- Rīga, mopēds

Skatīt arī

apdzīvotas vietas ASV:

- Riga, pilsēta Ņujorkas štatā, ASV
- Riga Township, ciemats Mičiganā, ASV

personvārds:

- *Riga Mustapha*, Ganā dzimis nīderlandiešu futbolists

Figure 2. Wikipedia disambiguation page about Rīga

Wikipedia is a hypertext document and has an internal link structure. Wikipedia's internal links can be of two types: *cross-article* links and *cross-language* links. Usually the text of a Wikipedia article contains links to other articles, cross-article links. For example, the article about the river *Abava* has 17 links to other Wikipedia articles. These articles are topically associated with the main article and refer to more detailed or more general information related to it (Adafre and Rijke, 2005). Cross-language links are links from any page describing an entity in one Wikipedia language to a page describing the same entity in another language (Wentland et al., 2008). For the article describing *Riga*, there are 97 links to Wikipedia pages in other languages with the same entry. The format of cross-language links is ***[[language code: Title]]***, e.g. ***[[ru: Рига]]*** the link to the article about *Riga* in Russian.

Since Wikipedia articles in different languages on the same topic are closely related, this relation can be of use in deriving comparable corpora for different languages. Several research studies use this feature of Wikipedia, demonstrating that Wikipedia has reached a level where it can support multilingual research (Adafre and Rijke, 2005, 2006; Bouma et al., 2006; Declerck et al., 2006; Jones et al., 2008). In our work cross-language links were used to develop object type corpora for languages other than English (cf. Section 4.2).

4. Wikipedia's content categorization procedure

To build the object type corpus for English we categorized each article in the entire Wikipedia dump from 24/07/2008 by *object type* by applying *Is-A patterns*. Our patterns are similar to the ones described in Hearst (1992) and Mann (2002) who used manually written patterns to extract hyponyms from large text corpora. Is-A patterns are described in the automaton shown in Figure 3.

We took a Wikipedia article, split it into sentences and POS tagged each sentence using shallow text analysis tools (OpenNLP tools[3]). Then each sentence was checked for the occurrence of an Is-A pattern. If multiple Is-A patterns were found, the first one was taken. In case it matched, the part of the sentence before the pattern was deleted and only the subsequent text was kept. Example 1 demonstrates the first sentence of the article about *Westminster Abbey*:

(22) *The Collegiate Church of St Peter at Westminster, which is almost always referred to by its original name of Westminster Abbey, is a large, mainly Gothic church, in Westminster, London, just to the west of the Palace of Westminster.*

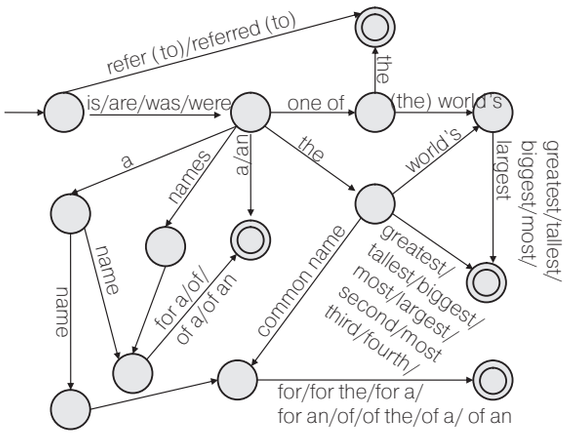

Figure 3. Is-A Patterns

[3] http://opennlp.sourceforge.net/

Is-A pattern was matched in the sentence and everything until the first occurrence of "*large*" was deleted, so that the following short version of the sentence was kept (Example 2).

(23) large/ADJ mainly/ADV Gothic/ADJ church/N ,/, in/P Westminster/PN ,/, London/PN ,/, just/ADV to/P the/DET west/N of/P the/DET Palace/PN of/P Westminster/NP

In order to find the category of the image, noun phrase (NP) search was applied to the short version of the sentence. For NP search, an approach similar to Bennett et al. (1999) rule-based NP identification for medical texts was implemented. The texts are pre-processed by tokenization and POS tagging and NPs are identified on basis of rules composed of different sequences of POS tags as shown in the automaton in Figure 4.

The noun input in the automaton is the starting POS tag for any sequence of rules. Therefore, the first occurrence of a noun in the shortened version of the sentence is searched in the beginning. Then, further POS tags which can be combined with a noun, e.g. Noun + Noun, Noun + Possessive, Noun + Possessive + Adjective + Noun, etc. are checked for. There is a possibility in each state of the automaton to terminate if the new input (*something else*) is not allowed to follow the term or POS tag found before. After an NP has been found, the nouns which occur at the end of the NP as the object type of the image are extracted. In the example with the above shortened sentence for *Westminster Abbey*: ***church/N*** was found as the first noun, and nothing described by the automaton was found after it. Therefore the automaton terminates and ***church*** is returned as the object type for the article about *Westminster Abbey*.

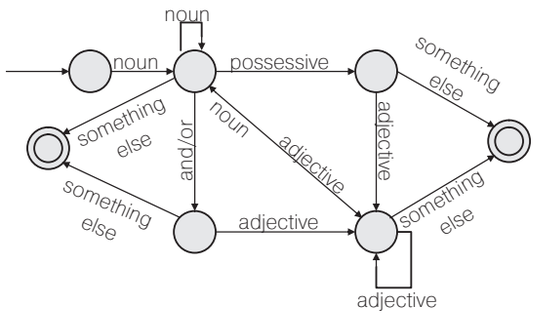

Figure 4. Noun Phrases

In this way about a half of Wikipedia articles (2.1 of the 5.4 millions) were automatically categorized. All together 40648 categories have been identified by our procedure. However, not all of these categories are about places, e.g. there are categories such as politician, leader, person, etc. which don't describe places and are not useful for our purposes. We manually filtered all identified categories in order to retain the ones describing places. That resulted in a set of 734 categories. From that set we retained 175 categories which are associated

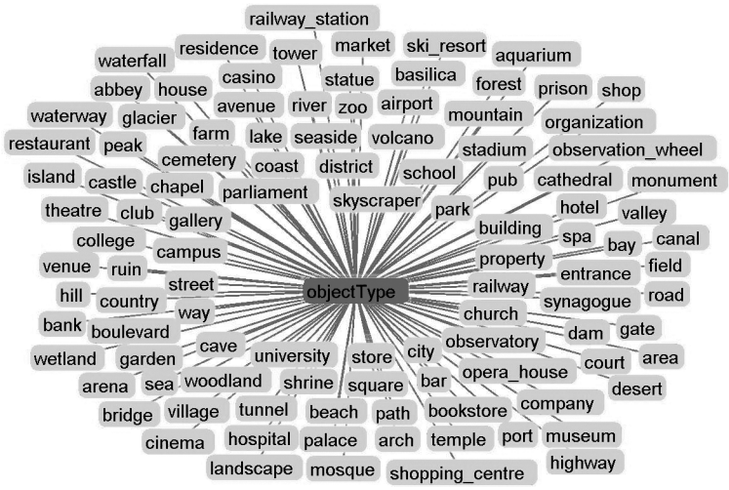

Figure 5. Object Types

with at least 50 Wikipedia articles. Finally, we manually assigned specific categories such as ***catholic church*** to a more general category ***church***. In this way we collected 107 categories containing articles about places around the world (Figure 5).

4.1. *Is-A pattern* evaluation

To assess the accuracy of the object type categorization we randomly selected 35 object type corpora and 50 articles from each corpus. Then we checked for each of these articles whether it is correctly or wrongly assigned to its object type. Finally, we calculate an accuracy value for each object type by dividing the number of correctly assigned articles by 50 (cf. Table 1). We observed an average accuracy of 80% for all 35 object types.

Table 1. Object types and the accuracy of the categorization

shopping center	0.9	bank	0.74
ski resort	1.0	monument	0.62
mountain	0.92	university	0.98
highway	0.82	building	0.52
railway station	1.0	park	0.96
mosque	0.66	gallery	0.725
waterfall	0.88	museum	0.7
street	0.58	canal	0.82
landscape	0.5	temple	0.74
restaurant	0.86	tower	0.52
island	0.92	prison	0.83
airport	1.0	residence	0.8
area	0.64	aquarium	0.62
volcano	0.92	castle	0.86
village	0.96	bridge	0.72
zoo	0.96	waterway	0.83
arena	0.96	river	0.94
wetland	0.79	average accuracy	0.80

A similar categorization has already been conducted for Wikipedia articles. It is stored in an ontology DBpedia[4]. However, the categorization of places by object types in the current version is not precise enough because it does not cover all object types identified by our procedure, e.g. churches, volcano, valley, pub, etc.

4.2. Multi-lingual object-type corpora collection

To obtain multi-lingual object type corpora we used cross-language links (cf. Section 3) to find Wikipedia articles in other languages for each article in the English corpus. For instance, the article about *Eiffel Tower* from the English Wikipedia was added to the object type ***tower***. Using cross-language links from this article we were able to obtain articles in those languages for which an article about the *Eiffel Tower* exists. Following this strategy we have collected object type text corpora for German, Italian and Latvian. Table 2 below shows the top-20 object types and the number of articles found for each object type for the four languages: English (EN), German (DE), Italian (IT), and Latvian (LV).

Table 2. Top-20 object types with the number of articles for English, German, Italian, and Latvian

Language / Object type	EN	DE	IT	LV
Airport	6493	560	201	9
Area	6934	731	382	31
Church	3005	392	371	7
City	14233	3788	2815	316
Company	5734	625	250	7
Country	3186	349	267	100

[4] http://www4.wiwiss.fu-berlin.de/dbpedia/dev/ontology.htm

Table 2. *Ending*

Language / Object type	EN	DE	IT	LV
District	6565	858	860	26
Island	6400	1292	689	73
Lake	3649	500	190	151
Mountain	5290	952	534	11
Museum	2320	277	121	0
Organization	9393	673	332	17
Park	3754	425	154	11
River	5851	1102	604	64
Road	3421	980	455	14
School	15794	292	128	5
Stadium	3665	509	274	11
University	7101	901	211	14
Village	39970	3550	3042	150
Way	2508	284	172	20

5. Example use of the object type corpus: Deriving language models for automatic summarization

Object type corpora for different languages described in the previous section can be used to derive language models to improve the results of multi-lingual image summary generation.

In multi-document summarization *language models* are used to improve the sentence selection procedure. Jagadeesh et al. (2005), for instance, generate language models using the query terms and a large text corpus to approximate the probability of a word occurring in a sentence relevant to the query, i.e., a sentence is generated on basis of the terms occurring in the language model. The more language model terms a sentence contains, the higher is the probability that the output sentence can be created on its basis. This probability is calculated and all the sentences are scored according to this calculation. As a result, the summary consists of sentences with highest ranks.

We implemented a similar approach for the task of toponym-referenced image description generation. Object type corpora are used to derive language models which in turn are used to bias the sentence selection during the summarization process. For instance, if an image summary is to be generated for the object *Westminster Abbey* then its object type ***church*** is automatically identified first with the help of Is-A patterns. Then uni-gram and bi-gram language models are derived from the corpus for the object type ***church***. Finally, while generating the summary for the object *Westminster Abbey* the ***church*** language model is used to rank the sentences of the documents to be summarized. More precisely, sentence generation process described in Jagadeesh et al. (2005) is applied to calculate the probability that the sentence is generated based on the ***church*** language model. Sentences with high probabilities are used to generate a summary to an image. Example 4 demonstrates the description automatically generated for the object *Westminster Abbey* in English.

(24) The Westminster abbey museum is located in the 11th century vaulted undercroft of St Peter beneath the former monks' dormitory in Westminster Abbey. The church is one of the most famous in Britain and is one of London's most visited tourist attractions. Westminster Abbey's long history can be traced back to the community of Benedictine monks established here c. 960 by Dunstan, bishop of London. It was most probably designed for the High Altar of the Abbey, although it has been damaged in past centuries. The Westminster Abbey's a magnifical monument, full of history and meaning. Westminster Abbey was originally a Benedictine monastery, refounded as the Collegiate Church of St. Peter in Westminster (today one of the boroughs constituting Greater London) by Queen Elizabeth I in 1560. It is the traditional place of coronation and burial site for English monarchs. The Westminster Abbey is certain that in about AD 785 there was a small community of monks on the island and that the monastery was enlarged and remodelled by St. Dunstan in about AD 960.

6. Conclusions and future work

In this paper we proposed a method for developing object type text corpora for four languages: English, German, Italian and Latvian with the aim to optimize multi-document summarization for generating image descriptions. We argued that Wikipedia, being a well structured web resource, suits our goals best since it is a rich source of location descriptions and offers features that facilitate multi-lingual cross-referencing. As such it is a valuable source for such languages as Latvian which is considered to have an under-resourced status. We described and evaluated categorization of Wikipedia's content according to object types using Is-A patterns . The evaluation results have shown that Is-A patterns are one simple way to identify the object type corpora which nevertheless renders satisfactory results. Finally, we illustrated how such object type corpora can be used to enhance multi-document summarization on the example of building automatic image captions by summarizing multiple web documents.

In spite of the fact that object type corpora were developed for all the four languages, most experiments

on automatic generating toponym-referenced descriptions of images have been performed for the English. The work on our remaining target languages is ongoing. We plan to carry out evaluation experiments for all the four languages using our object type corpora. It will allow us to investigate whether image summaries that incorporate conceptual models about objects derived from object type corpora are better in quality than those generated without such models and whether the contribution of conceptual models can be observed across languages.

Acknowledgment

The research reported was partly funded by the TRIPOD project (TRI-Partite multimedia Object Description)[5] supported by the European Commission under the contract No. 045335. We would like to thank Lars Borin, Robert Gaizauskas, Emina Kurtic, Andrew Salway, Inguna Skadiņa and Raivis Skadiņš for discussions and comments.

[5] http://tripod.shef.ac.uk/

References

1. *Adafre S. F. and Rijke M.* (2005) Discovering Missing Links in Wikipedia // Proceedings of the Workshop on Link Discovery: Issues, Approaches and Applications — LinkKDD, 2005: 90–97.
2. *Adafre S. F. and Rijke M.* (2006) Finding Similar Sentences across Multiple Languages in Wikipedia // Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics, 2006: 62–69.
3. *Aker A. and Gaizauskas R.* (2008) Evaluating automatically generated user-focused multi-document summaries for geo-referenced images // Proceedings of the 22nd International Conference on Computational Linguistics — COLING 2008, Manchester.
4. *Barnard K. and Forsyth D.* (2001) Learning the semantics of words and pictures // Proceedings of International Conference on Computer Vision, Vol. 2, Vancouver: IEEE, pp. 408–415.
5. *Barnard K., Duygulu P., Forsyth D., de Freitas N., Blei D., Jordan M.* (2003) Matching words and pictures // The Journal of Machine Learning Research, 3: 1107–1135.
6. *Bennett N., He Q., Powell K., Schatz B.* (1999) Extracting noun phrases for all of MEDLINE // Proceedings of American Medical Informatics Association.
7. *Bernardini S., Baroni M., Evert S.* (2006) A WaCky introduction // Wacky! Working papers on the Web as Corpus, Bologna: GEDIT, 2006: 9–40.
8. *Bouma G., Fahmi I., Mur J., Noord G., Plas L., Tiedermann J.* (2006) Using Syntactic Knowledge for QA // Proceedings of Cross Language Evaluation Forum Workshop, Aarhus, Denmark, 2008.
9. *Cheng P.-J., Pan Y.-C., Lu W.-H., Chein L.-F.* (2004) Creating Multilingual Translation Lexicons with Regional Variations // Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics, Barcelona, Spain, Article No. 534, 2004.
10. *Declerck T., Gómez-Pérez A., Vela O., Gantner Z., Manzano-Macho D.* (2006) Multilingual Lexical Semantic Resources for Ontology Translation // Proceedings of the 5th International Conference on Language Resources and Evaluation: LREC 2006, Genoa, Italy, 2006: 1492–1495.
11. *Deschacht K. and Moens M.* (2007) Text Analysis for Automatic Image Annotation // Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics. East Stroudsburg: ACL.
12. *Duygulu P., Barnard K., de Freitas J., Forsyth D.* (2002) Object Recognition as Machine Translation: Learning a Lexicon for a Fixed Image Vocabulary // Proceedings of the 7th European Conference on Computer Vision (ECCV), 4: 97–112.
13. *Feng Y. and Lapata M.* (2008) Automatic Image Annotation Using Auxiliary Text Information // Proceedings of Association for Computational Linguistics (ACL) 2008, Columbus, Ohio, USA.
14. *Hearst M.* (1992) Automatic acquisition of hyponyms from large text corpora // Proceedings of the 14th conference on Computational linguistics, Vol. 2: 539–545.
15. *Hollink L., Schreiber A., Wielinga B., Worring M.* (2004) Classification of user image descriptions // International Journal of Human-Computer Studies, 61(5): 601–626.
16. *Hughes B.* (2006) A web search service for minority language communities // Proceedings of Open Road 2006 Conference: Challenges and Possibilities, 2006.
17. *Jagadeesh J., Pingali P., Varma V.* (2005) A relevance-based language modeling approach to DUC 2005 // Proceedings of Document Understanding Conferences (along with HLT-EMNLP 2005), Vancouver, Canada.
18. *Jamieson M., Fazly A., Dickinson S., Stevenson S., Wachsmuth S.* (2007) Learning Structured Appearance Models from Captioned Images of Cluttered Scenes // Computer Vision, 2007, ICCV 2007, IEEE 11th International Conference: 1–8.

19. *Jones G., Fantino F., Newman E., Zhang Y.* (2008) Domain-Specific Query Translation for Multilingual Information Access using Machine Translation Augmented With Dictionaries Mined from Wikipedia // Proceedings of the 2nd International Workshop on "Cross Lingual Information Access" Addressing the Information Need of Multilingual Societies, 2008: 34–41.
20. *Kilgarriff A.* (2007) Googleology is bad science // Computational Linguistics, Vol. 33(1): 147–151.
21. *Kilgarriff A. and Grefenstette G.* (2003) Introduction to the special issue on the web as corpus // Computational Linguistics, Vol. 29: 333–347.
22. *Kuo J.-S. and Yang Y.-K.* (2004) Constructing Transliteration Lexicons from Web Corpora // Proceedings of the Association for Computational Linguistics 2004 on Interactive poster and demonstration sessions, Barcelona, Spain, Article No. 3, 2004.
23. *Liu V. and Curran J. R.* (2006) Web Text Corpus for Natural Language Processing // Proceedings of the European Chapter of the Association for Computational Linguistics, 2006: 233–240.
24. *Mann G.* (2002) Fine-grained proper noun ontologies for question answering // Proceedings of International Conference On Computational Linguistics, 2002: 1–7.
25. *Mori Y., Takahashi H., Oka R.* (2000) Automatic word assignment to images based on image division and vector quantization // Proceedings of RIAO 2000: Content-Based Multimedia Information Access.
26. *Pan J., Yang H., Duygulu P., Faloutsos C.* (2004), Automatic image captioning // Multimedia and Expo, 2004. ICME'04, Vol. 3.
27. *Ponzetto S. and Strube M.* (2007) Deriving a Large Scale Taxonomy from Wikipedia // Proceedings of the 22nd Conference on the Advancement of Artificial Intelligence, Vancouver, B.C., Canada, 2007: 1440–1445.
28. *Shatford S.* (1986) Analyzing the Subject of a Picture: A Theoretical Approach // Cataloging and Classification Quarterly, 6(3): 39–61.
29. *Steinberger R., Pouliquen B., Widiger A., Ignat C., Erjavec T., Tufiş D., Varga D.* (2006) The JRC-Acquis: A multilingual aligned parallel corpus with 20+ languages // Proceedings of the 5th International Conference on Language Resources and Evaluation: LREC 2006, Genoa, Italy, 2006.
30. *Wentland W., Knopp J., Silberer C., Hartung M.* (2008) Building a Multilingual Lexical Resource for Named Entity Disambiguation, Translation and Transliteration // The sixth Language Resources and Evaluation Conference: LREC'08, Marrakech, Morocco, 2008.