INEX Tweet Contextualization Track at CLEF 2012: Query Reformulation using Terminological Patterns and Automatic Summarization

Jorge Vivaldi and Iria da Cunha

Universitat Pompeu Fabra
Institut Universitari de Lingüística Aplicada
Barcelona
{iria.dacunha,jorge.vivaldi}@upf.edu
http://www.iula.upf.edu

Abstract. The tweet contextualization INEX task at CLEF 2012 consists of the developing of a system that, given a tweet, can provide some context about the subject of the tweet, in order to help the reader to understand it. This context should take the form of a readable summary, not exceeding 500 words, composed of passages from a provided Wikipedia corpus. Our general approach to get this objective is the following: we perform some automatic reformulations of the initial tweets provided for the task (obtaining a list of terms related with the main topic of all them using terminological patterns). Then, using these reformulated tweets, we obtain related documents with the search engine Indri. Finally, we use REG, an automatic extractive summarization system based on graphs, to summarize these documents and provide the summary associated to each tweet.

Key words: INEX, CLEF, Tweets, Terms, Named Entities, Wikipedia, Automatic Summarization, REG.

1 Introduction

The tweet contextualization INEX (Initiative for the Evaluation of XML Retrieval) task at CLEF 2012 (Conference and Labs of the Evaluation Forum) consists of the developing of a system that, given a tweet, can provide some context about the subject of the tweet, in order to help the reader to understand it. This context should take the form of a readable summary, not exceeding 500 words, composed of passages from a provided Wikipedia corpus. Like in the Question-Answering (QA) of INEX 2011, the task to be performed by the participating groups is contextualizing tweets, that is answering questions of the form "what is this tweet about?" using a recent cleaned dump of the Wikipedia. The general process involves: tweet analysis, passage and/or XML elements retrieval and construction of the answer. Relevant passages would be segments containing relevant information and also containing as little non-relevant information as possible (the result is specific to the question).

The test data are about 1000 tweets in English collected by the organizers of the task from Twitter. They were selected among informative accounts (for example, @CNN, @TennisTweets, @PeopleMag, @science...), in order to avoid purely personal tweets that could not be contextualized. Information such as the user name, tags or URLs is provided. The document collection for all the participants, that is the corpus, has been rebuilt based on a dump of the English Wikipedia from November 2011. Resulting documents are made of a title, an abstract and sections with sub-titles.

We consider that automatic extractive summarization systems could be useful in this QA task, taking into account that a summary can be defined as "a condensed version of a source document having a recognizable genre and a very specific purpose: to give the reader an exact and concise idea of the contents of the source" [1]. Summaries can be divided into "extracts", if they contain the most important sentences extracted from the original text (ex. [2], [3], [4], [5], [6], [7]), and "abstracts", if these sentences are re-written or paraphrased, generating a new text (ex. [8], [9], [10]). Most of the current automatic summarization systems are extractive.

Our general approach is the following: we perform some automatic reformulations of the initial queries provided for the task (obtaining a list of terms related with the main topic of all the tweets using terminological patterns). Then, using these reformulated queries, we obtain related documents with the search engine Indri[1]. Finally, we use REG ([11], [12]), an automatic extractive summarization system based on graphs, to summarize these documents and provide the final summary associated to each query.

This approach is similar to the one used at QA@INEX track 2010 (see [13]) and 2011 (see [14]), since the same summarization system is employed. Nevertheless, in our past participations, the system was semi-automatic, while in this work the system is totally automatic, from the reformulation of the queries using terminological patterns, until the multi-document summarization of all the retrieved documents.

The evaluation of the participant systems involves two aspects: informativeness and readability. Informativeness evaluation is automatic, using the automatic evaluation system FRESA (FRamework for Evaluating Summaries Automatically) ([15], [16], [17]), and readability evaluation is carried out manually (evaluating syntactic incoherence, unsolved anaphora, redundancy, etc.).

Following this introduction, the paper is organized as follows. In Section 2, the summarization system REG is shown. In Section 3, some information about terminology and terminological patterns is given. In Section 4, the methodol-

[1] Indri is a search engine from the Lemur project, a cooperative work between the University of Massachusetts and Carnegie Mellon University in order to build language modelling information retrieval tools: `http://www.lemurproject.org/indri/`

ogy is explained. In Section 5, experimental settings and results are presented. Finally, in Section 6, conclusions are exposed.

2 State-of-the-art and Resources

2.1 Term Extraction

The notion of term that we have adopted in this work is based on the "Communicative Theory of Terminology" [18]: a term is a lexical unit (single/multiple word) that activates a specialized meaning in a thematically restricted domain. Terms detection implies the distinction between domain-specific terms and general vocabulary. Its results are useful for any NLP task containing a domain specific component such as: ontology and (terminological) dictionary building, text indexing, automatic translation and summarization systems, among others. In spite of its large application field, its reliable and practical recognition still constitutes a bottleneck for many applications.

As shown in [19], [20] and [21] among others, there are several methods to obtain the terms from a corpus. On the one hand, there are methods based on linguistic knowledge, like Ecode [22]. On the other hand, there are methods based on single statistical measures, such as ANA [23] or a combination of them, such as EXTERMINATOR [24]. Some tools combine both linguistic knowledge and statistically based methods, such as TermoStat [25], the algorithm shown in [26] or the bilingual extractors by [27] and [28]. However, none of these tools uses any kind of semantic knowledge. Notable exceptions are Metamap [29], Trucks [30] and YATE [31], among others. Also Wikipedia must be considered, since it is a very promising resource that is increasingly being used for both monolingual ([32], [33]) and multilingual term extraction [34].

Most of the tools, in particular those including an important linguistic component, takes into consideration the fact that terms usually follow a small number of POS patterns. In [35] it was shown that three patterns (noun, noun-adjective and noun-preposition-noun) cover more that 90% of the entries found in medical terminological dictionaries. Many of the above mentioned tools make some use of this fact. Nevertheless, some researchers like in [36] dynamically calculate the list of patterns found in terminological resources.

2.2 Named Entities Extraction

Named Entity Recognition (NER) may be defined as the task to identify names referring to persons, organizations and locations in free text; later this task has been expanded to obtain other entities like dates and numeric expressions. This task was originally introduced as possible types of fillers in Information Extraction systems at the 6th Message Understanding Conference [37]. Although initially this task was limited to identify such expressions, later it has been expanded to their labeling with one entity type label ("person", "organization", etc.). Note that an entity (such as "Stanford", the American university at the

U.S.) can be referenced using several surface forms (e.g., "Stanford University" and "Stanford") and a single surface form (e.g., "Stanford") can refer to several entities (the university but also an American financer, several places in the UK or a financial group). See [38] for an interesting review.

NER has proved to be a task useful for a number of NLP tasks as question answering, textual entailment and coreference resolution, among others. The recent interest in emerging areas like bioinformatics allows to expand this recognition task to proteins, drugs and chemical names. While early studies were mostly based on handcrafted rules, most recent ones use supervised machine learning as a way to automatically induce rule-based systems or sequence labeling algorithms starting from a collection of training examples.

Often, corpus processing tools include some text handling facilities to perform simple NER detection for facilitating later processing. Some of them are based in language specific peculiarities such as initial upper case letters together with some heuristics for name entities placed at the beginning of the sentence. This is the case of the tool used for this experiment (see a description in [39]).

2.3 The REG System

REG ([11], [12]) is an Enhanced Graph summarizer (REG) for extract summarization, using a graph approach. The strategy of this system has two main stages: a) to carry out an adequate representation of the document and b) to give a weight to each sentence of the document. In the first stage, the system makes a vectorial representation of the document. In the second stage, the system uses a greedy optimization algorithm. The summary generation is done with the concatenation of the most relevant sentences (previously scored in the optimization stage).

REG algorithm contains three modules. The first one carries out the vectorial transformation of the text with filtering, lemmatization/stemming and normalization processes. The second one applies the greedy algorithm and calculates the adjacency matrix. We obtain the score of the sentences directly from the algorithm. Therefore, sentences with a higher score are selected as the most relevant. Finally, the third module generates the summary, selecting and concatenating the relevant sentences. The first and second modules use CORTEX [6], a system that carries out an unsupervised extraction of the relevant sentences of a document using several numerical measures and a decision algorithm.

3 Methodology

A main point in this research is to consider that named entities as well as words sequences that agree with the typical terminological patterns (see section 2.1) are representative of the tweets' topic. To test this assertion, we design a methodology to automatically retrieve all significant sequences from the tweets that satisfy the above mentioned criteria.

The first step is to POS tag the tweets file. As a matter of fact, and in order to keep the process fully automatic, a minimal manipulation of the tweets file has been done. It includes only a minor modification to allow the text handling tool to keep the tweet id connected to the tweet itself.

The next step, terminological patterns extraction, has been done using an already existent module of the YATE term extraction tool [31]. This information, together with the POS tagged tweet (to obtain proper nouns info) is used to build the query string for Indri.

Some care has been taken to keep track of multiword sequences as indicated by the Indri query language specification (see examples below).

In order to enrich the queries, we use a local installation of a Wikipedia dump[2] to expand the terms with redirection information from such Wikipedia info. In this way, a query term like "Falklands" may be searched in the Wikipedia to find that it can be also referenced as "Falkland Islands"; therefore, the final query term is rewritten as:

```
#syn(Falklands #1(Falkland Islands))
```

This strategy is also useful to find acronyms expansion as "USGS" and "United States Geological Survey" resulting in the following query:

```
#syn("USGS" #1(United States Geological Survey))
```

Moreover, it allows to find words with different spellings as:

```
#syn(#1(Christine de Pisan) #1(Christine de Pizan))
```

The resulting query has been delivered to Indri, using track organizer's script, to obtain the Wikipedia pages relevant to every query. The following is an example of a full tweet:

```
Increasingly, central banks, especially in emerging markets,
 have been the marginal  buyers of gold http://t.co/9mftD5ju
 via WSJ.
```

and its corresponding query string:

```
#1(marginal buyers of gold),#1(emerging markets),
#1(central banks),#syn("WSJ" #1(The Wall Street Journal))
```

The resulting set of Wikipedia pages has been split in several documents. Each document contains the pages relevant to the query. Such document is the input to the REG summarization system (see section 2.3), which builds a summary with the significant passages.

[2] This resource has been otained using [40].

4 Experiments Settings and Results

As mentioned in section 3, the process is fully automatic. No human intervention has taken place; therefore, errors and/or mistakes in the process may have a multiplicative effect. Most of such issues are exemplified as follows:

1. Tweet itself. The tweets file (including 1000 tweets) prepared by the organization includes several errors like: mispelling, joined words, foreign language, etc. Consider the following examples:
   - 169657757870456833: "Lakers now 17-12 on the season & 12-2 at home. @paugasol 20pts 13rebs 4blks. Bynum 15pts 15rebs. @0goudelock 10pts, two 3 PTers."
   - 169904294642978816: "@ranaoboy @Utcheychy @Jhpiego Thx for the #wiwchat RTs! Great conversation!"
   - 169655717538701312: "METTA. WORLD. PEACE."
   - 170175722449670145: "http://t.co/amQ6IShA"
   - 170207412366745600: "RT @MexicanProblms: #41. When you're eating junk food y tu mom te dice que no comas "chucherias." #MexicanProblems".

   Please note that, in some cases, it results in an empty query string or the resulting sentence is too short, causing POS tagging errors due to lack of context.
2. POS tagging. The output of most of the tools used for tagging (TreeTagger in this case) has some error rate. Unfortunately, errors mentioned above as well as extremely short sentences have a negative influence in the tagger performance.
3. Wikipedia expansion. It may happen that information added through Wikipedia expansion is not fully useful. This may be the case the only added information is the change of the case of some letters of the query term.
4. Indri query system. As shown in [41], this retrieval system has its own limits.
5. REG summarization system. The retrieval system issues a number of Wikipedia pages; therefore, it would be necessary to use a multidocument summarization system. As a matter of fact, REG is a single document summarizer, so some redundance may appear in the summaries.

Some of the above issues may cause unusual results in the terminological patterns extraction tool. Therefore, in such cases, the pages retrieved by Indri may not correspond to the information available in Wikipedia about tweets' topics.

The evaluation of all the participant systems in the tweet contextualization INEX task at CLEF 2012 involves two aspects: informativeness and readability. On the one hand, as mentioned, to evaluate the informativeness the automatic FRESA package is used. This evaluation framework includes document-based summary evaluation measures based on probabilities distribution, specifically, the Kullback-Leibler (KL) divergence and the Jensen-Shannon (JS) divergence.

As in the ROUGE package [42], FRESA supports different n-grams and skip n-grams probability distributions. FRESA environment has been used in the evaluation of summaries produced in several European languages (English, French, Spanish and Catalan), and it integrates filtering and lemmatization in the treatment of summaries and documents.

Table 1 includes the official results of the informativeness evaluation in the the tweet contextualization INEX task at CLEF 2012. This table presents the scores of the 33 participant runs.

Table 1. Final results of informativeness in the tweet contextualization INEX task at CLEF 2012.

Rank	Run	Uni	Bi	Skip
1	178	0.7734	0.8616	0.8623
2	152	0.7827	0.8713	0.8748
3	170	0.7901	0.8825	0.8848
4	194	0.7864	0.8868	0.8887
5	169	0.7959	0.8881	0.8904
6	168	0.7972	0.8917	0.8930
7	193	0.7909	0.8920	0.8938
8	185	0.8265	0.9129	0.9135
9	171	0.8380	0.9168	0.9187
10	186	0.8347	0.9210	0.9208
11	187	0.8360	0.9235	0.9237
12	154	0.8233	0.9254	0.9251
13	162	0.8236	0.9257	0.9254
14	155	0.8253	0.9280	0.9274
15	153	0.8266	0.9291	0.9290
16	196b	0.8484	0.9294	0.9324
17	196c	0.8513	0.9305	0.9332
18	196a	0.8502	0.9316	0.9345
19	164	0.8249	0.9365	0.9368
20	197	0.8565	0.9415	0.9441
21	163	0.8664	0.9628	0.9629
22	165	0.8818	0.9630	0.9634
23	150	0.9052	0.9871	0.9868
24	188	0.9541	0.9882	0.9888
25	176	0.8684	0.9879	0.9903
26	149	0.9059	0.9916	0.9916
27	156	0.9366	0.9913	0.9916
28	157	0.9715	0.9931	0.9937
29	191	0.9590	0.9947	0.9947
30	192	0.9590	0.9947	0.9947
31	161	0.9757	0.9949	0.9950
32	177	0.9541	0.9981	0.9984
33	151	0.9223	0.9985	0.9988

As shown in Table 1, our run (165) obtains the position 22 in the rank. Exactly, it obtains 0.8818 using unigrams, 0.9630 using bigrams and 0.9634 using skip bigrams. The best run in the ranking (178) obtains 0.7734, 0.8616 and 0.8623, respectively.

On the other hand, readability is evaluated manually. Evaluators are asked to evaluate several aspects related to syntactic incoherence, unsolved anaphora, redundancy, etc. The specific orders given to evaluators are:

– Syntax S: "Tick the box is the passage contains a syntactic problem (bad segmentation for example)".
– Anaphora A: "Tick the box if the passage contains an unsolved anaphora".
– Redundancy R: "Tick the box if the passage contains a redundant information, i.e. an information that have already been given in a previous passage".
– Trash T: "Tick the box if the passage does not make any sense in its context (i.e. after reading the previous passages). These passages must then be considered as trashed, and readability of following passages must be assessed as if these passages were not present".

The score is the average normalized number of words in valid passages, and participants are ranked according to this score. Summary word numbers are normalized to 500 words each.

Table 2 includes the final results of readability evaluation in the tweet contextualization INEX task at CLEF 2012. Estimated average scores are available for:

– Relevance: proportion of text that makes sense in context.
– Syntax: proportion of text without syntax problems.
– Structure: proportion of text without broken anaphora and avoiding redundancy.

These measures were estimated on the same pool of tweets as for previously released informativeness evaluation by organizers.

Runs that failed to provide at least 6 consistent summaries in this pool have been kept apart because the estimates were too uncertain for inclusion in the official results. Because of this reason, in Table 2 only 27 runs are shown.

As shown in Table 2, our run (165) obtains the position 7 in the rank. Exactly, it obtains 0.5936 using unigrams, 0.6049 using bigrams and 0.5442 using skip bigrams. The best run in the ranking (185) obtains 0.7728, 0.7452 and 0.6446, respectively.

These results show that the performance of our system is not so good regarding informativeness, but it is much better regarding readability. This difference between informativeness and readability is also shown by other systems (see for example the best runs in both categories, 178 and 185). In our case, we consider that the mentioned mistakes in the tweets and the fact that the terminology extraction is totally automatic can cause that the pages retrieved by Indri are not as relevant as expected. Nevertheless, using an automatic summarization system, we can guarantee that the quality of readability is acceptable.

Table 2. Final results of readability in the tweet contextualization INEX task at CLEF 2012.

Run	Relevance	Syntax	Structure
185	0.7728	0.7452	0.6446
171	0.631	0.606	0.6076
168	0.6927	0.6723	0.5721
194	0.6975	0.6342	0.5703
186	0.7008	0.6676	0.5636
170	0.676	0.6529	0.5611
165	0.5936	0.6049	0.5442
152	0.5966	0.5793	0.5433
155	0.6968	0.6161	0.5315
178	0.6336	0.6087	0.5289
169	0.5369	0.5208	0.5181
193	0.6208	0.6115	0.5145
163	0.5597	0.555	0.4983
187	0.6093	0.5252	0.4847
154	0.5352	0.5305	0.4748
196b	0.4964	0.4705	0.4204
153	0.4984	0.4576	0.3784
164	0.4759	0.4317	0.3772
162	0.4582	0.4335	0.3726
197	0.5487	0.4264	0.3477
196c	0.449	0.4203	0.3441
196a	0.4911	0.3813	0.3134
176	0.2832	0.2623	0.2388
156	0.2933	0.2716	0.2278
188	0.1542	0.1542	0.1502
157	0.1017	0.1045	0.1045
161	0.0867	0.0723	0.0584

5 Conclusions and Future Work

In this paper, our strategy and results for the tweet contextualization INEX task at CLEF 2012 are presented. The task consists of the developing of a system that, given a tweet, can provide some context about the subject of the tweet, in order to help the reader to understand it. This context should take the form of a readable summary, not exceeding 500 words, composed of passages from a provided Wikipedia corpus. The test data are about 1000 tweets in English collected by the organizers of the task from Twitter.

Our system performs some automatic reformulations of the initial tweets provided for the task (obtaining a list of terms related with their main topic using terminological patterns). Then, using these reformulated tweets, we obtain related documents with the search engine Indri. Finally, we use REG to summarize these documents and provide the final summary associated to each tweet.

The results show that, comparing to the other participants, the performance of our system is not so good regarding informativeness (probably due to mistakes in the tweets and problems in the terminology extraction process), but it is much better regarding readability (probably due to the fact of using a summarization system).

In the future we plan to follow several parallel lines: i) to improve term selection and its expansion to refine the queries and therefore to improve the pertinence of the Wikipedia pages retrieved by Indri; ii) to further investigate the actual pertinence of the Wikipedia retrieved pages to the query; and iii) to check the actual weight of summarization process in the full task by testing other summarization systems.

References

1. Saggion, H.; Lapalme, G. (2002). *Generating Indicative-Informative Summaries with SumUM*. Computational Linguistics 28(4). 497-526.
2. Edmunson, H. P. (1969). *New Methods in Automatic Extraction*. Journal of the Association for Computing Machinery 16. 264-285.
3. Nanba, H.; Okumura, M. (2000). *Producing More Readable Extracts by Revising Them*. In Proceedings of the 18th Int. Conference on Computational Linguistics (COLING-2000). Saarbrucken. 1071-1075.
4. Gaizauskas, R.; Herring, P.; Oakes, M.; Beaulieu, M.; Willett, P.; Fowkes, H.; Jonsson, A. (2001). *Intelligent access to text: Integrating information extraction technology into text browsers*. In Proceedings of the Human Language Technology Conference. San Diego. 189-193.
5. Lal, P.; Reger, S. (2002). *Extract-based Summarization with Simplication*. In Proceedings of the 2nd Document Understanding Conference at the 40th Meeting of the Association for Computational Linguistics. 90-96.
6. Torres-Moreno, J-M.; Velázquez-Morales, P.; Meunier, J. G. (2002). *Condensés de textes par des méthodes numériques*. In Proceedings of the 6th Int. Conference on the Statistical Analysis of Textual Data (JADT). St. Malo. 723-734.
7. da Cunha, I.; Fernández, S.; Velázquez, P.; Vivaldi, J.; SanJuan, E.; Torres-Moreno, J-M. (2007). *A new hybrid summarizer based on Vector Space Model, Statistical Physics and Linguistics*. Lecture Notes in Computer Science 4827. 872-882.
8. Ono, K.; Sumita, K.; Miike, S. (1994). *Abstract generation based on rhetorical structure extraction*. In Proceedings of the Int. Conference on Computational Linguistics. Kyoto. 344-348.
9. Paice, C. D. (1990). *Constructing literature abstracts by computer: Techniques and prospects*. Information Processing and Management 26. 171-186.
10. Radev, D. (1999). Language Reuse and Regeneration: Generating Natural Language Summaries from Multiple On-Line Sources. PhD Thesis. New York, Columbia University.
11. Torres-Moreno, J-M.; Ramírez, J. (2010). *REG : un algorithme glouton appliqué au résumé automatique de texte*. Proceedings of the 10th Int. Conference on the Statistical Analysis of Textual. Roma, Italia.
12. Torres-Moreno, J-M.; Ramírez, J.; da Cunha, I. (2010). *Un resumeur a base de graphes, indépendant de la langue*. In Proceedings of the Int. Workshop African HLT 2010. Djibouti.

13. Vivaldi, J.; da Cunha, I.; Ramírez, J. (2011). *The REG summarization system with question reformulation at QA@INEX track 2010*. In Geva, S. et al. (eds.). INEX 2010, Lecture Notes in Computer Science 6932. 295-302. Berlín: Springer.
14. Vivaldi, J.; da Cunha, I. (2012). *QA@INEX Track 2011: Question Expansion and Reformulation Using the REG Summarization System*. Lecture Notes in Computer Science (LNCS) 7424. 257-268. Berlin: Springer.
15. Saggion, H.; Torres-Moreno, J-M.; da Cunha, I.; SanJuan, E.; Velázquez-Morales, P.; SanJuan, E. (2010). *Multilingual Summarization Evaluation without Human Models*. In Proceedings of the 23rd Int. Conference on Computational Linguistics (COLING 2010). Pekin.
16. Torres-Moreno, J-M.; Saggion, H.; da Cunha, I.; SanJuan, E.; Velázquez-Morales, P. (2010). *Summary Evaluation With and Without References*. Polibitis: Research journal on Computer science and computer engineering with applications 42.
17. Torres-Moreno, J-M.; Saggion, H.; da Cunha, I.; Velázquez-Morales, P.; SanJuan, E. (2010). *Ealuation automatique de résumés avec et sans référence*. In Proceedings of the 17e Conférence sur le Traitement Automatique des Langues Naturelles (TALN). Montreal: Univ. de Montréal et Ecole Polytechnique de Montréal.
18. Cabré, M. T. (1999). *La terminología. Representación y comunicación*. Barcelona: IULA.
19. Cabré, M. T.; Estopà, R.; Vivaldi, J. (2001). *Automatic term detection. A review of current systems*. Recent Advances in Computational Terminology 2. 53-87.
20. Pazienza, M. T.; Pennacchiotti, M.; Zanzotto, F.M. (2005). *Terminology Extraction: An Analysis of Linguistic and Statistical Approaches*. Studies in Fuzziness and Soft Computing 185. 255-279.
21. Ahrenberg, L. (2009). *Term Extraction: A Review*. (Unpublished draft).
22. Alarcón, R.; Sierra, G.; Bach, C. (2008). *ECODE: A Pattern Based Approach for Definitional Knowledge Extraction*. In Proceedings of the XIII EURALEX Int. Congress. Barcelona: IULA, UPF, Documenta Universitaria. 923-928.
23. Enguehard, C.; Pantera, L. (1994). *Automatic Natural Acquisition of a Terminology*. Journal of Quantitative Linguistics 2(1). 27-32.
24. Patry, A.; Langlais, P. (2005). *Corpus-based terminology extraction*. In Proceedings of 7th Int. Conference on Terminology and Knowledge Engineering. Copenhagen.
25. Drouin, P. (2002). *Acquisition automatique des termes: l'utilisation des pivots lexicaux spécialisés*. Ph.D. Thesis. Montreal (Canada): Université de Montréal.
26. Frantzi, K. T.; Ananiadou, S.; Tsujii, J. (2009). Erdmann, M.; Nakayama, K.; Hara, T.; Nishio, S. (2009). *The C-value/NC-value Method of Automatic Recognition for Multi-word Terms*. Lecture Notes in Computer Science 1513. 585-604
27. Vintar, S. (2010). *Bilingual term recognition revisited: The bag-of-equivalents term alignment approach and its evaluation*. Terminology 16(2). 141-158.
28. Gómez Guinovart, X. (2012). *A Hybrid Corpus-Based Approach to Bilingual Terminology Extraction*. In I. Moskowich and B. Crespo (eds.). Encoding the Past, Decoding The Future: Corpora in the 21st Century. Cambridge Scholar Publishing: Newcastle upon Tyne. 147-175.
29. Aronson, A.; Lang, F. (2010). *An overview of MetaMap: historical perspective* and recent advances. Journal of the American Medical Informatics Association 17(3). 229-236.
30. Maynard, D. (1999). *Term Recognition Using Combined Knowledge Sources*. Ph.D. Thesis. Manchester Metropolitan University. Manchester (UK).
31. Vivaldi, J. (2001). *Extracción de candidatos a término mediante combinación de estrategias heterogéneas*. Ph.D. thesis. Universitat Politècnica de Catalunya. Barcelona (Spain).

32. Vivaldi, J.; Rodríguez, H. (2010). *Using Wikipedia for term extraction in the biomedical domain: first experiences*. Procesamiento del Lenguaje Natural 45. 251-254.
33. Cabrera-Diego, L.; Sierra, G.; Vivaldi, J.; Pozzi, M. (2011). *Using Wikipedia to Validate Term Candidates for the Mexican Basic Scientific Vocabulary*. In Proceedings of LaRC 2011: First Int. Conference on Terminology, Languages, and Content Resources. Seoul. 76-85.
34. Erdmann, M.; Nakayama, K.; Hara, T.; Nishio, S. (2009). *Improving the Extraction of Bilingual Terminology from Wikipedia*. ACM Transactions on Multimedia Computing, Communications and Applications 5(4). 31.1-31.16.
35. Estopà, R. (1999). *Extracting Lexical Semantic Knowledge from Wikipedia and Wiktionary*. Ph.D. Thesis. Pompeu Fabra University. Barcelona (Spain).
36. Nazar, R.; Cabré, M. T. (2012). *Supervised Learning Algorithms Applied to Terminoloy Extraction*. In 10th Terminology and Knowledge Engineering Conference.
37. Grishman, R.; Sundheim, B. (1996). *Message Understanding Conference - 6: A Brief History*. In Proceedings of the 16th Int. Conference on Computational Linguistics. 466-471.
38. Nadeau, D.; Sekine, S. (2007). *A survey of named entity recognition and classification*. Journal of Linguisticae Investigationes 30(1). 3-26.
39. Martínez, H.; Vivaldi, J.; Villegas, M. (2010). *Text handling as a Web Service for the IULA processing pipeline*. In Proceedings of the 7th conference on International Language Resources and Evaluation (LREC'10). 22-29.
40. Zesch, T.; Müller, C., Gurevych, I. (2008). *Extracting Lexical Semantic Knowledge from Wikipedia and Wiktionary*. In 6th LREC Conference Proceedings. 1646-1652.
41. Strohman, Trevor; Metzler, Donald; Turtle, Howard; Croft, Bruce (2005). *Indri: A language-model based search engine for complex queries*. University of Massachusetts Amherst. CIIR Technical Report IR-407.
42. Lin, C-Y. (2004). *ROUGE: A Package for Automatic Evaluation of Summaries*. In Proceedings of Text Summarization Branches Out: ACL-04 Workshop. 74-81.