A Resource-light Approach to Phrase Extraction for English and German Documents from the Patent Domain and User Generated Content

Julia Maria Schulz, Daniela Becks, Christa Womser-Hacker, Thomas Mandl

University of Hildesheim
Marienburger Platz 22
31141 Hildesheim
{Julia-Maria.Schulz, Daniela.Becks, Womser, Mandl} @ uni-hildesheim.de

Abstract

In order to extract meaningful phrases from corpora (e. g. in an information retrieval context) intensive knowledge of the domain in question and the respective documents is generally needed. When moving to a new domain or language the underlying knowledge bases and models need to be adapted, which is often time-consuming and labor-intensive. This paper adresses the described challenge of phrase extraction from documents in different domains and languages and proposes an approach, which does not use comprehensive lexica and therefore can be easily transferred to new domains and languages. The effectiveness of the proposed approach is evaluated on user generated content and documents from the patent domain in English and German.

Keywords: multilingual phrase extraction, shallow parsing, cross-language information retrieval, opinion mining

1. Introduction

In our globalizing world, information is often found in documents which are not written in the native language of a user. To enable users to find these documents, methods are needed that overcome language borders. This kind of challenges, where the query language and the languages of the document collection may differ, are dealt with in *Cross-Language Information Retrieval* (CLIR) and explored in evaluation initiatives like CLEF[1] and NTCIR[2] and will also be adressed in this paper (Becks et al., 2011, 157).

Commonly used approaches in mono- and multilingual information retrieval are the bag-of-words approaches. Those approaches, which on the one hand have been extensively used during the indexing process and on the other hand during the formulation of queries, are being replaced recently and benefits of using phrases in opposition to simple terms in the information retrieval process are being discussed (Tseng et al., 2007, 1222) (Becks and Schulz, 2011, 389). Therefore this paper adresses the extraction of phrases in a multilingual context.

This can also be demonstrated by a simple example: A query for *big ben* does not only deliver documents on the well known sight in London, but also documents mentioning someone named *Ben* and containing the adjective *big*, like in the following example sentence: "*Ben* mastered the *big* challenge very well". If one consideres the above terms in the query as a phrase, the ambiguity is resolved and only documents containing a combination of the terms are beeing retrieved (Becks et al., 2011, 157).

As domains and document types in a lot of retrieval scenarios (e. g. web retrieval) are very diverse, methods are needed, that can cope with the different requirements this diversity of domains and documents entail. The extraction of applicable phrases in a multilingual context is a challenging problem, since each language and corpus has different characteristics.

The proposed approach for the extraction of meaningful phrases combines shallow and deep parsing and can be adapted to different languages and domains with only small modifications. To evaluate the domain independence and transferability of the approach to different languages, English and German documents of two very different domains are being used: patents and customer reviews which is considered as user generated content (Becks et al., 2011, 157).

The paper is organized as follows: In the next section the scope of the application and the challenges in multilingual and multidomain phrase extraction are being described, followed by an overview of related work. In section 4. our methodology is outlined and in section 5. the evaluation is described including a brief outline of the data and an overview of the evaluation results. The paper closes with some conclusions and an outlook.

2. Context of the research

The proposed approach is used in two different application areas. The first one – a cooperation between the University of Hildesheim and FIZ Karlsruhe – focuses on the patent domain. The aim in this context is to investigate the additional benefit of phrases for (interactive) query expansion in patent retrieval. Therefore, the effect of different types of phrases is evaluated within a series of controlled experiments (Becks, 2010, 423).

Opinion mining – the second area of application – is a recent disciplin, that adresses the identification and classification of opinions as well as the identification of the objects, opinions are beeing expressed about (*opinion target*) (e. g. Hu and Liu (2004), Popescu (2007) or Fang and Chen (2011)), and persons or institutions expressing the opinion in question (*opinion holder*) (e. g. Kim and Hovy (2006) or Sayeed et al. (2010)) (see figure 1). The application adresses the extraction of phrases in this context aiming at

[1] Cross-Language Evaluation Forum: http://clef2011.org, http://www.clef-campaign.org

[2] National Institute of Informatics Test Collection for IR Systems: http://research.nii.ac.jp/ntcir/index-en.html

identifying and extracting phrases containing opinions and the respective opinion target (e. g. product features) from a multilingual document collection. As an opinion target is modified by an opinion, there exists a head-modifier-relation between the target and the opinion.

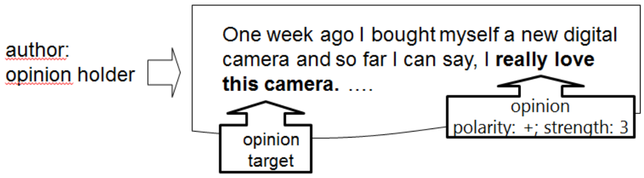
author:
opinion holder
One week ago I bought myself a new digital camera and so far I can say, I really love this camera. ….
opinion
polarity: +; strength: 3
opinion
target

Figure 1: Example sentence demonstrating the different areas of interest in opinion mining

With respect to the application areas (information retrieval and opinion mining) a phrase is defined as a combination of terms showing a head-modifier relation. This relation may have different characteristics (e. g. adjective-noun-relation, verb-adverb-relation). These phrases are different from chunks, which consist of a single content word surrounded by function words and pre-modifiers and follow a predefined template (Abney, 1991, 257). The example in fig. 2 demonstrates, that a phrase can exceed the borders of a chunk following the above definition, which differs from the classical linguistic phrase definition due to the areas of application. Multi-word terms (e. g. information retrieval system) are as well considered as combinations of subject and predicate (Becks et al., 2011, 158). A list of the considered phrase types is given in table 1.

$$\underbrace{[\text{a system}]}_{\text{chunk}}\ \underbrace{[\text{for information retrieval}]}_{\text{chunk}}$$

$$\text{vs}$$

$$\text{a}\ \underbrace{[\text{system for information retrieval}]}_{\text{phrase}}$$

Figure 2: Example of a chunk and a phrase in the described meaning (Becks et al., 2011, 158)

The development of an applicable extraction component is determined by the following aims within the areas of application: The phrase extraction should be adaptable to multiple European languages with only small efforts and as linguistic resources are not available comprehensively, it should work without extensive knowledge bases (resource-light approach). It should also be capable of handling the *unknown words problem* – handle words, that are neither present in a used dictionary nor in any training corpus (Uchimoto et al., 2001, 91) – which is particularly important within the patent domain.

Within the research contexts the primary focus is on the precision of the extracted phrases, as they are further incorporated in different applications (patent retrieval system and opinion mining system). Within the patent domain usually the recall is the primary focus, as it is often important to find all the relevant documents, but since this paper adresses the extraction of phrases, which will later be used for query expansion in this area of application, the focus is different here.

phrase type	**example**
subject-predicate	high electrical insulating properties are needed
predicate-object	prevent dust
verb-adverb	biasing forwardly
multi-word terms	cryosurgery procedure
adjectiv-noun	electrical insulating material
noun-prepositional phrase	device for uniform fluid distribution
noun-genitiv	center region of the cylinder
noun-relative clause, noun-participle	core pipe having a porous wall

Table 1: Annotated and extracted phrase types

3. Related work

Among the traditional phrase extraction methods there are on the one hand rule-based approaches and on the other hand dependency based methods. One very early and until today commonly used rule-based approach is the delimiter-based approach (original German term: *Begrenzerverfahren*) of Jaene and Seelbach (1975), who use phrases in terms of multi-word terms forming a syntactic-semantic entity (Jaene and Seelbach, 1975, 9) for the indexing process of English technical documents. They define pairs of words functioning as delimiters enclosing the noun phrases to be extracted.

A similar approach for the extraction of maximal-length noun phrases from French documents has been proposed by Bourigault and Jacquemin (1999). During the extraction process they use pairs of delimiters as well as shallow grammar of the noun phrases and in a second step, split the phrases in their constituent parts (head and modifier).

A comparable approach for the patent domain is described by Tseng et al. (2007), who use a list of stop words to extract descriptors.

In the opinion mining domain Guo et al. (2009) also use stop words accomplished by opinion words (e. g. adjectives) for the extraction of product features from the semi-structured part of reviews, where a short summary of the opinions of a user is given by listing the features in *pro* and *con* categories.

As this short overview demonstrates most of the phrase extraction techniques focus on the extraction of noun phrases. As Koster (2004) argues, verbal phrases are often used to describe processes and thus capture important information. In the context of the two mentioned application areas this fact is very important. Especially in the patent domain objects are often not mentioned directly but described in a vague way in order to broaden the claim of a patent as far as possible Schamlu (1985, 124). e. g. a *personal computer* could be described as a “machine processing data and hav-

ing a graphical user interface in order to enable the user to manipulate the executed processes". In an opinion mining context verbal phrases also play an important role, as they are often used, to express opinions about an object or service (see fig. 1). As a consequence, verbal phrases are considered in the extraction process as well.
On the other hands there are also methods relying on external knowledge bases as dependency parsing approaches do. In a retrieval context dependency relations are often used in terms of head-modifier-pairs, with the modifier specifying the head (Koster, 2004, 423).
The benefit of head-modifier-pairs is that they provide not only syntactic but also semantic information (e. g. see Ruge (1989, 9)). Therefore they are used mainly during the indexing process (Koster, 2004; Ruge, 1995) and can also be used beneficially in the context of classification tasks in terms of triples (term-relation-term) (Koster and Beney, 2009). (Becks and Schulz, 2011, 390) Dependency relations have also been used for identifying opinions and the corresponding opinion target by incorporating a dependency parser (Wu et al., 2009; Popescu, 2007; Popescu and Etzioni, 2005).

4. Methodology

This paper describes a new method for the extraction of phrases combining the previously mentioned categories. The main target of the extraction approach is to develop a tool for the identification of phrases that can be easily adapted to new domains and languages (e. g. adaption of the delimiter pairs or valid prepositions for noun-genitive-phrases (NG)) without using domain specific knowledge bases in order to maintain the domain independence. As the semantic of the extracted phrases needs to be kept in mind, a hybrid technique is used, combining the functionality of a shallow parser and the flat semantic classification based on linguistic rules (Becks and Schulz, 2011, 390).
Within the extraction component the rule-based approach of Jaene and Seelbach (1975) and Bourigault and Jacquemin (1999) is used and combined with the basics of dependency parsing (Ruge, 1995). The approach uses part-of-speech tags for the delimiter pairs instead of words and therefore only needs a POS-Tagger and the implementation of the delimiter rules instead of extensive word lists for each language. Hence it is a resource-light approach.
The implemented rules vary for each phrase type (see table 1). For example an adjective-noun-relation (AN-R) is often enclosed by an article and a punctuation mark or a preposition (see fig. 3). Additionally, the phrase needs to contain at least one adjective and one noun and is not allowed to contain for example a personal pronoun. Because the category *article* for example includes the German articles as well as the English ones, the rule can be used for different languages. This abstracted version of the delimiter approach can hence be generalized and does not need complex word lists for the extraction process.
As mentioned earlier the basics of dependency parsing are also taken into consideration, as each phrase should consist of a head and a modifier, whose identification is also done rule-based. The following rules are for example used to identify the head of an English phrase:

- predicate-object: head is positioned at the first position
- subject-predicate: head is positioned at the end of the phrase
- verb-adverb: head is located at the end of the phrase
- adjective-noun: head is located at the last position
- noun-genitiv: head is positioned in front of the preposition "'of"'
- multi-word terms: head is located at the last position
- noun-relative clause: head is located at the first position
- noun-prepositional phrase: head is in front of the preposition (e. g. "for")

In the examples in figure 3 the head is positioned at the end of the phrase ("stud", "front panel button layout"), the modifier precedes the head. The mentioned example shows that the proposed method cannot only handle phrases consisting of a single head-modifier-pair, but can also handle longer phrases consisting of more complex head-modifier-relations. (Becks et al., 2011, 159f.)

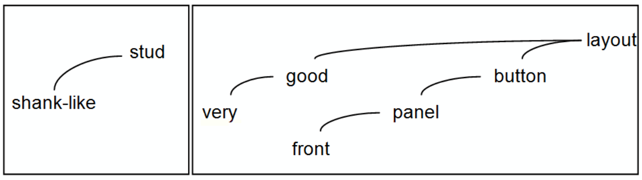

Figure 4: Dependency relations of the phrases in figure 3

5. Evaluation

5.1. Data

Usually the quality of the generated output is evaluated against a gold standard, see for example Verbene et al. (2010). They used a manually annotated sample of 100 sentences. The calculation of the precision is based on a comparison of the extracted phrases and the manually annotated sample (Becks and Schulz, 2011, 391).
The corpus used in the first area of application consists of approximately 105,000 documents of the CLEF-IP[3] test collection 2009, containing about 1.6 million patents in English as well as in German and French (Roda et al., 2010, 388).
In the opinion mining context a corpus is used consisting of customer reviews in English (Hu and Liu, 2004; Ding et al., 2008), German and Spanish (Schulz et al., 2010).
Looking at the characteristics of the two corpora particularly the length of the documents differs significantly, as patents are very extensive and complex documents (e. g. Iwayama et al. (2003)) and the customer reviews in

[3] Cross-Language Evaluation Forum, Intellectual Property Track

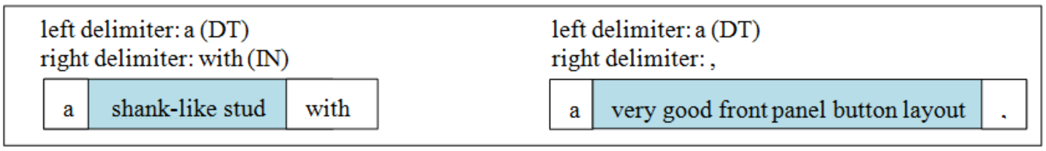

Figure 3: Example of an extracted adjective-noun phrase (Becks and Schulz, 2011, 391); left: patent document(EP-1120530-B1), right: customer review (Hu and Liu, 2004)

contrast are rather short documents, which sometimes only consist of a few sentences. They also differ in respect to the structure at a sentence level. Patents often consist of very long and strongly nested sentences whereas customer reviews tend to have shorter sentences, which are at most only slightly nested. These two domains are chosen in order to demonstrate that the proposed approach can be used for the extraction of meaningful phrases in very different domains with only small adaptions.

A manually annotated gold standard is used for the evaluation. It is based on a random sample of sentences composed of about 2000 tokens[4] per language and domain of the two corpora described above. These sentences have been annotated by two independent annotators (the first and the second author of the paper), with regard to the phrase types listed int table 1.

5.2. Inter-Annotator-Agreement

Overall 2594 phrases (2431 unique phrases) have been annotated. Table 2 gives an overview of the number of phrases for the different domain and language samples.

48% of the 2592 phrases in the gold standard are uncontroversial meaning the annotated phrases coincide with respect to the annotated phrase borders and relations. For 25% of the unique phrases the annotators agreed regarding the annotated relation, but the annotated phrase borders differ. This means 74% of the phrases were annotated with the same relation and had at least partially matching phrase boundaries. One of the most common reasons for the differing phrase boundaries are coordinated phrases (e. g. *above-mentioned heat-conductive and electrical insulating tapes* vs *above-mentioned heat-conductive tapes* and *above-mentioned electrical insulating tapes*). Another common reason where phrases ending with a digit, which was enclosed by one annotator and excluded by the other (e. g. *predetermined track (target track) of the optical disc* vs *predetermined track (target track) of the optical disc 2*). Other differences in the annotations consider adverbial phrases, indirect objects, or reflexive pronouns, which belong to a verb, if it is a real reflexive verb, and are an object of the verb otherwise. In case of differing annotations an agreement was found by discussion or by consulting an independent expert. These numbers show that the identified and classified phrases match in most of the cases. A Cohen's cappa (Cohen, 1960) value of $\kappa = 0.61$ was reached, which is a good number considering the diverse domains of the documents and the discontinuity of some of the phrases (Becks et al., 2011, 161).

	Patent Sample		Review Sample		
	English	German	English	German	
unique phrases	616	498	682	635	2431
all phrases	726	521	702	645	2594

Table 2: Number of annotated phrases in the gold standard

5.3. Results

For evaluation purposes the automatically generated output of the phrase extraction component was matched against the gold standard. A phrase was considered to be an exact match, if the relation as well as the phrase boundaries matched. For the partial matches two different alternatives are considered: a lenient and a strict one. With respect to the lenient alternative, the relation of the phrase and one border of the phrase had to coincide. Additionally, the phrase was only allowed to differ by one term for the strict alternative. Besides the relation and the phrase boundaries the frequency of a phrase is also considered during the evaluation. This is done, because the frequency can be used as indicator of relevance. In the opinion mining context this is very important, as opinions are always subjective and the frequency can therefore be an indicator of how likely a customer him- or herself might have this opinon too. Thus, helping the customer to make a decision in the purchasing process.

For evaluation purposes the classical recall and precision measures are chosen with respect to the areas of application. They are calculated as follows:

$$precision = \frac{\#correctly\_retrieved\_phrases}{\#retrieved\_phrases}$$

$$recall = \frac{\#correctly\_retrieved\_phrases}{\#correct\_phrases}$$

Following the above definitions a precision of about 60% for the strict evaluation (Reviews: 65%; Patents: 58%) and 75% (Reviews: 73%; Patents: 76%) for the lenient evaluation was reached regarding the different noun phrases. The best results could be achieved for adjective-noun phrases with a precision of 87% (strict) and 91% (lenient) respectively, followed by noun-genitive phrases with a precision of 68% (strict) and 84% (lenient). While for the first phrase type the precision for English and German differ only between 4 and 5%-points, for the latter there is a difference of more than 18%-points for the strict and still about 7%-points for the lenient evaluation. The inferior results in German can be explained by the stronger nested noun-genitive-constructions – especially in the German patent documents

[4] The number of tokens is based on the output of the pos-tagger counting punctuation marks as well.

– making the detection of the correct phrase boundaries more difficult.

Regarding the verb phrases the results for the review sample are considerably better than those for the patent domain, which emphasizes the complexity of the task and leads to the assumption that some adaptions need to be made for the patent domain, e. g. enlarging the maximal phrase length and taking gerunds into consideration for English documents. Even this small adaptions lead to improvements of over 20%-points for the extracted verb phrases in the English patent sample resulting in a precision of about 63%.

Alltogether the recall in the conducted experiments is rather low (strict: 34%, lenient: 42%), which can be explained by the focus on precision for the phrase extraction in the considered areas of application (see section 2.).

6. Conclusion

The paper shows that one can achieve a good precision for the extraction of phrases from documents in different domains and languages with a resource-light approach, which is of primary focus in a retrieval context. However, the results also show that some slight modifications (e. g. for the patent domain) can clearly improve the results.

In the future we want to evaluate the effectiveness of the approach for additional languages (French, Spanish) and examine the influence of the pos tagging model in order to further improve the precision of the approach.

7. References

Stephen P. Abney. 1991. Parsing by Chunks. In Robert C. Berwick, Stephen P. Abney, and Carol Tenny, editors, *Principle-Based Parsing*, volume 44 of *Studies in linguistics and philosophy*, pages 257–278. Kluwer, Dordrecht.

Daniela Becks and Julia M. Schulz. 2011. Domänenübergreifende Phrasenextraktion mithilfe einer lexikonAnalysekomponente. In Joachim Griesbaum, Thomas Mandl, and Christa Womser-Hacker, editors, *Information und Wissen: global, sozial und frei?*, volume 58 of *Schriften zur Informationswissenschaft*, pages 388–392. Werner Hülsbusch, Boizenburg.

Daniela Becks, Julia M. Schulz, Christa Womser-Hacker, and Thomas Mandl. 2011. Multilinguale Phrasenextraktion mit Hilfe einer lexikonunabhängigen Analysekomponente am Beispiel von Patentschriften und nutzergenerierten Inhalten. In Hanna Hedeland, Thomas Schmidt, and Kai Wörner, editors, *Multilingual Resources and Multilingual Applications*, volume 96 of *Arbeiten zur Mehrsprachigkeit - Folge B*, pages 157–162, Hamburg. Universität Hamburg.

Daniela Becks. 2010. Begriffliche Optimierung von Patentanfragen. *Information - Wissenschaft & Praxis*, 61(6-7):423.

Didier Bourigault and Christian Jacquemin. 1999. Term Extraction + Term Clustering: an Integrated Platform for Computer-aided Terminology. In *Proceedings of the ninth conference on European chapter of the Association for Computational Linguistics*, EACL99, pages 15–22. Association for Computational Linguistics.

Jacob Cohen. 1960. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1):37–46.

Xiaowen Ding, Bing Liu, and Philip S. Yu. 2008. A Holistic Lexicon-based Approach to Opinion Mining. In *Proceedings of the International Conference on Web Search and Web Data Mining*, pages 231–240, New York and NY and USA. ACM.

Ji Fang and Bi Chen. 2011. Incorporating Lexicon Knowledge into SVM Learning to Improve Sentiment Classication. In *Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology (SAAIP)*, pages 94–100.

Honglei Guo, Huijia Zhu, Zhili Guo, XiaoXun Zhang, and Zhong Su. 2009. Product Feature Categorization with Multilevel Latent Semantic Association. In *Proceeding of the 18th ACM Conference on Information and Knowledge Management*, pages 1087–1096, New York and NY and USA. ACM.

Minqing Hu and Bing Liu. 2004. Mining Opinion Features in Customer Reviews. In Deborah L. Mcguinness and George Ferguson, editors, *Proceedings of the Nineteenth National Conference on Artificial Intelligence, Sixteenth Conference on Innovative Applications of Artificial Intelligence*, pages 755–760. AAAI Press / The MIT Press.

Makoto Iwayama, Atsushi Fujii, Noriko Kando, and Yuzo Marukawa. 2003. An Empirical Study on Retrieval Models for Different Document Genres: Patents and Newspaper Articles. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, SIGIR03, pages 251–258. ACM.

Hartmut Jaene and Dieter Seelbach. 1975. *Maschinelle Extraktion von zusammengesetzten Ausdrücken aus englischen Fachtexten*. Beuth, Berlin and Köln and Frankfurt(Main).

Soo M. Kim and Eduard Hovy. 2006. Extracting Opinions, Opinion Holders, and Topics Expressed in Online News Media Text. In *Proceedings of ACL/COLING Workshop on Sentiment and Subjectivity in Text*, pages 1–8, Stroudsburg and PA and USA. Association for Computational Linguistics.

Cornelis H. A. Koster and Jean G. Beney. 2009. Phrase-based Document Categorization Revisited. In *Proceeding of the 2nd international workshop on Patent information retrieval (PaIR09)*, pages 49–56. ACM.

Cornelis H. A. Koster. 2004. Head/Modifier Frames for Information Retrieval. In Alexander Gelbukh, editor, *Computational Linguistics and Intelligent Text Processing*, LNCS 2945, pages 420–432. Springer, Berlin and Heidelberg.

Ana-Maria Popescu and Oren Etzioni. 2005. Extracting product features and opinions from reviews. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 339–346, Morristown and NJ and USA. Association for Computational Linguistics.

Ana-Maria Popescu. 2007. *Information extraction from unstructured web text*. Ph.D. thesis, University of Washington, Seattle and WA and USA.

Giovanna Roda, John Tait, Florina Piroi, and Veronika Zenz. 2010. CLEF-IP 2009: Retrieval Experiments in the Intellectual Property Domain. In Carol Peters, Giorgio Maria Di Nunzio, Mikko Kurimo, Thomas Mandl, Djamel Mostefa, Anselmo Peñas, and Giovanna Roda, editors, *Multilingual Information Access Evaluation I*, volume 6241 of *Lecture Notes in Computer Science*, pages 385–409. Springer, Berlin and Heidelberg.

Gerda Ruge. 1989. Generierung semantischer Felder auf der Basis von Frei-Texten. *LDV Forum*, 6(2):3–17.

Gerda Ruge. 1995. *Wortbedeutung und Termassoziation. Methoden zur automatischen semantischen Klassifikation*. Olms, Hildesheim and Zürich and New York.

Asad B. Sayeed, Hieu C. Nguyen, Timothy J. Meyer, and Amy Weinberg. 2010. Expresses-an-opinion-about: using corpus statistics in an information extraction approach to opinion mining. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, pages 1095–1103, Beijing and China. Chinese Information Processing Society of China.

Mariam Schamlu. 1985. *Patentschriften - Patentwesen: Eine argumentationstheoretische Analyse der Textsorte Patentschrift am Beispiel der Patentschriften zu Lehrmitteln*. Iudicium-Verlag, München.

Julia M. Schulz, Christa Womser-Hacker, and Thomas Mandl. 2010. Multilingual Corpus Development for Opinion Mining. In Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner, and Daniel Tapias, editors, *Proceedings of the Seventh conference on International Language Resources and Evaluation*, pages 3409–3412. European Language Resources Association (ELRA).

Yuen-Hsien Tseng, Chi-Jen Lin, and Yu-I Lin. 2007. Text Mining Techniques for Patent Analysis. *Information Processing and Management*, 43(5):1216–1247.

Kiyotaka Uchimoto, Satoshi Sekinez, and Hitoshi Isahara. 2001. The Unknown Word Problem: a Morphological Analysis of Japanese Using Maximum Entropy Aided by a Dictionary. In *Proceedings of the 2001 Conference on Empirical Methods in Natural Language Processing (EMNLP 01)*, pages 91–99. ACL.

Suzan Verbene, Eva D'hondt, and Nelleke Oostdijk. 2010. Quantifying the Challenges in Parsing Patent Claims. In *Proceedings of the 1st International Workshop on Advances in Patent Information Retrieval (AsPIRe'10)*, pages 14–21. Milton Keynes.

Yuanbin Wu, Qi Zhang, Xuanjing Huang, and Lide Wu. 2009. Phrase dependency parsing for opinion mining. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, volume Volume 3, pages 1533–1541, Morristown and NJ and USA. Association for Computational Linguistics.