Extracting Aspects and Polarity from Patents

Peter Anick, Marc Verhagen and James Pustejovsky
Computer Science Department
Brandeis University
Waltham, MA, United States
Peter_anick@yahoo.com, marc@cs.brandeis.edu,
jamesp@cs.brandeis.edu

Abstract

We describe an approach to terminology extraction from patent corpora that follows from a view of patents as "positive reviews" of inventions. As in aspect-based sentiment analysis, we focus on identifying not only the components of products but also the attributes and tasks which, in the case of patents, serve to justify an invention's utility. These semantic roles (component, task, attribute) can serve as a high level ontology for categorizing domain terminology, within which the positive/negative polarity of attributes serves to identify technical goals and obstacles. We show that bootstrapping using a very small set of domain-independent lexico-syntactic features may be sufficient for constructing domain-specific classifiers capable of assigning semantic roles and polarity to terms in domains as diverse as computer science and health.

1 Introduction

Automated data mining of patents has had a long history of research, driven by the large volume of patents produced each year and the many tasks to which they are put to use, including prior art investigation, competitive analysis, and trend detection and forecasting (Tseng, 2007). Much of this work has concentrated on bibliographic methods such as citation analysis, but text mining has also been widely explored as a way to assist analysts to characterize patents, discover relationships, and facilitate patent searches. One of the indicators of new technology emergence is the coinage, adoption and spread of new terms; hence the identification and tracking of technical terminology over time is of particular interest to researchers designing tools to support analysts engaged in technology forecasting (e.g., Woon, 2009; deMiranda, 2006)

For the most part, research into terminology extraction has either (1) focused on the identification of keywords within individual patents or corpora without regard to the roles played by the keywords within the text (e.g., Sheremetyeva, 2009) or, (2) engaged in fine-grained analysis of the semantics of narrow domains (e.g., Yang, 2008). In this paper we strive towards a middle ground, using a high-level classification suitable for all domains, inspired in part by recent work on sentiment analysis (Liu, 2012). In aspect-based sentiment analysis, natural language reviews of specific target entities, such as restaurants or cameras, are analyzed to extract *aspects,* i.e., features of the target entities, along with the sentiment expressed toward those features. In the restaurant domain, for example, aspects might include the breadth of the menu, quality of the service, preparation of the food, and cost. Aspects thus tend to capture the tasks that the entity is expected to perform and various dimensions and components related to those tasks. Sentiment reflects the reviewer's assessment of these aspects on a scale from negative to positive.

A patent application is required by definition to do three things: describe an invention, argue for its novelty, and justify its utility. The utility of a patent is typically defined by the accomplishment of a new task or an improvement to some existing task along one or more dimensions. Thus, a patent can be thought of as a *positive* review of a product with respect to specific aspects of its task(s). Indeed, the most commonly occurring verbs in patents include those indicative of *components* ("comprise", "include"), *attributes* ("increase", "reduce"), and *tasks* ("achieve", "perform"). Organizing keywords along these high-level distinctions, then, would allow patent analysts to explore terminological infor-

Proceedings of SADAATL 2014, pages 31–39,
Dublin, Ireland, August 24, 2014.

mation from several different relevant perspectives. Furthermore, given the interpretation of a patent as a positive review, it should be possible to identify the default polarity of measurable aspects in the context of a domain. For example, if a patent makes a reference to *increasing network bandwidth*, then this should lend support to the notion that network bandwidth is not only a relevant attribute within the patent's domain but also a positive one. Likewise, if a patent refers to *reducing power consumption*, then we might interpret power consumption as an aspect with negative polarity. For analysts trying to assess trends within a technology domain, tracking the occurrences of terms signifying tasks and attributes, along with their polarity, could help them characterize the changing goals and obstacles for inventors over time.

The US patent office receives over half a million patent applications a year.[1] These are classified by subject matter within several standardized hierarchical schemes, which permits dividing up the corpus of patents both by application date and subfield (e.g., computer science, health, chemistry). Since our goal is to support analysts across all domains, it is highly desirable to extract domain-specific aspects through semi-supervised machine learning rather than incur the cost of domain-specific knowledge engineering. To this end, we employed a bootstrapping approach in which a small number of domain independent features was used to generate a much larger number of domain dependent features for classification. We then applied naïve Bayes classification in a two-step classification process: first distinguishing attributes, components and tasks; and then classifying the extracted attribute terms by their polarity.

The paper is structured as follows. In section 2, we describe the system architecture. Section 3 shows results for two domains (computer science and health). In section 4, we present an evaluation of results and discuss issues and shortcomings of the current implementation. In section 5, we present related research and in section 6, our conclusions and directions for future work.

2 System architecture

2.1 Corpus processing

Our patent collection is a set of 7,101,711 US patents in XML-markup form from Lexis-Nexis. We divided the collection into subcorpora by application year and high-level domain using the patents' classification within the USPTO hierarchy. The XML markup was then used to extract the relevant portions of patents for further analysis. These sections included title, abstract, background, summary, description and claims. References, other than those embedded in the sections above, were omitted, as they contain many entity types (people, publications, and organizations) that are not particularly useful for our current task. The text of each section was extracted and broken into sentences by the Stanford tagger (Toutanova, 2003) which also tokenized and tagged each token with a part of speech tag.

We then chunked adjacent tokens into simple noun phrase chunks of the form (ADJECTIVE)? (NOUN)* NOUN.[2] We will hereafter refer to these chunks as *terms*. The majority of these patent terms fall into one of three major categories:

Components: the physical constituents or processes that make up an invention, as well as the objects impacted, produced by or used in the invention.

Tasks: the activities which inventions, their components or beneficiaries perform or undergo.

Attributes: the measureable dimensions of tasks and components mentioned in the patent.

To generate features suitable for machine learning of these semantic categories, we used a small set of lexico-syntactic relationships, each defined with respect to the location of the term in a sentence:

prev_V: the closest token tagged as a verb appearing to the left of the term, along with any prepositions or particles in between. (*cached_in, prioritizing, deal_with*)

prev_VNpr: a construction of the form <verb><NP><prep> appearing to the left of the term. Only the head noun in the NP is retained (*inform|user|of, provides|list|of, causes|increase|in*)

prev_Npr: a construction of the form <noun><prep> appearing to the left of the term. (*restriction_on, applicability_of, time_with*)

[1] http://www.uspto.gov/web/offices/ac/ido/oeip/taf/us_stat.htm

[2] We blocked a set of 246 general adjectival modifiers (e.g., *other, suitable, preferred, entire, initial,…*) from participating in terms.

prev_Jpr: a construction of form <adjective> <prep> appearing to the left of the term. (*free_from, desirable_in, unfamiliar_with*)
prev_J: a construction of form <adjective> <prep> appearing to the left of the term. (*excessive, considerable, easy*)

These features were designed to capture specific dependency relations between the term and its premodifiers and dominant verbs, nouns, and adjective phrases. We extracted the features using localized rules rather than create a full dependency parse.[3] One additional feature internal to the term itself was also included: last_word. This simply captured the head term of the noun phrase, which often carries generalizable semantic information about the phrase. Each feature instance was represented as a string comprising a prefix (the feature type) and its value (a token or concatenation of tokens).

2.2 Classification

For each term appearing in a subcorpus, the collection of co-occurring features across all documents was assembled into a single weighted feature vector in which the weight captured the number of documents for which the feature occurred in conjunction with the given term. We also calculated the document frequency for each term, as well as its "domain specificity score", a metric reflecting the relative frequency of the term in specialized vs. randomized corpora (see section 3).

In order to avoid the need to create manually labeled training data for each patent domain, we employed bootstrapping, a form of semi-supervised learning in which a small number of labeled features or seed terms are used in an iterative fashion to automaticaly identify other likely diagnostic features or category exemplars. Bootstrapping approaches have previously shown considerable promise in the construction of semantic lexicons (Riloff, 1999; Thelen, 2002, Ziering, 2013). By surveying common prev_V features in a domain-independent patent subcorpus, we selected a small set of domain-independent diagnostic lexico-syntactic features ("seed features") that we felt were strong indicators for each of the three semantic categories. The set of seed features for each category is shown below. Semantically equivalent inflectional variants were also included as features.

Attribute: *improve, optimize, increase, decrease, reduce*
Component: *comprise, contain, encompass, incorporate, use, utilize, consist_of, assembled_of, composed_of*
Task: *accomplish, achieve, enhance, facilitate, assisting_in, employed_in, encounter_in, perform, used_for, utilized_for*

We then utilized these manually labeled generic features to bootstrap larger feature sets F for domain-specific subcorpora. For each term t in a domain-specific subcorpus, we extracted all the manually labeled features that the term co-ocurred with. Any term which co-occurred with at least two labeled feature instances and for which all of its labeled features were of the same class was itself labeled with that class for subsequent use as a seed term s for estimating the parameters of a multinomial naïve Bayes classifier (Manning et al, 2008). Each seed term so selected was represented as a bag of its co-occurring features.

The prior probability of each class and conditional probabilities of each feature given the class were estimated as follows, using Laplace "add one" smoothing to eliminate 0 probabilities:

$$\hat{P}(c_j) = \frac{|\,S_j\,|}{|\,S\,|}$$

$$\hat{P}(f|c) = \frac{count(f,c) + 1}{count(c) + |\,F\,|}$$

[3] The compute time required to produce dependency parses for the quantity of data to be analyzed led to the choice of a "leaner" feature extraction method.

where S_j is the set of seed terms with class label *j*, *S* is the set of all seed terms, *count(f,c)* is the count of co-occurrences of feature *f* with seed terms in class *c*, *count(c)* is the total number of feature co-occurrences with seed terms in class c, and *F* is the set of all features (used for Laplace smoothing).

Using the naïve Bayes conditional independence assumption, the class of each term in a subcorpus was then computed by maximizing the product of the prior probability for a class and the product of the conditional probabilities of the term's features:

$$C = \underset{c \in C}{\operatorname{argmax}} P(c) \prod_{f \in F} P(f|c)$$

Terms for which no diagnostic features existed were labeled as "unknown".

Once the terms in a subcorpus were categorized as *attribute*, *component*, or *task*, the terms identified as attributes were selected as input to a second round of classification.[4] We used the same bootstrapping process as described for the first round, choosing a small set of features highly diagnostic of the polarity of attributes. For positive polarity, the seed features were: *increase, raise, maximize*. For negative polarity: *avoid, lower, decrease, deal_with, eliminate, minimize, reduce, resulting_from, caused_by*. Based on co-occurrence with these features, a set of terms was produced from which parameters for a larger set of features could be estimated, as described above. We then used naïve Bayes classification to label the full set of attribute terms.

3 Results

We present results from two domains, health and computer science, using a corpus consisting of all US patent applications submitted in the year 2002. The health subcorpus consisted of 19,800 documents, while the computer science subcorpus contained 51,058 documents. A "generic" corpus composed of 38,482 patents randomly selected from all domains was also constructed for the year for use in computing a "domain specificity score". This score was designed to measure the degree to which a term could be considered part of a specific domain's vocabulary and was computed as the log(probability of term in domain corpus / probability of term in generic corpus). For example, in computer science, the term *encryption technology* earned a domain specificity score of 4.132, while *speed* earned .783 and *color* garnered .022. Using a combination of term frequency (# of documents a term occurs in within a domain) thresholds and domain specificity, one can extract subsets of terms with varying degrees of relevance within a collection.[5]

3.1 Attribute/Component/Task (ACT) Classification

The bootstrapping process generated 1,644 features for use in the health domain and 3,200 in computer science. Kullback-Leibler divergence is a commonly used metric for comparing the difference between two probability distributions (Kullback and Leibler, 1951). By computing Kullback-Leibler divergence $D_{KL}(P||Q)$ between the distribution P of classes predicted by each feature (i.e., the probability of the class given the feature alone based on the term seed set labels) and the prior class distribution Q, we could estimate the impact of individual features in the model. Table 1 shows some of the domain-specific features in the health and computer science domains, along with the category each tended to select for.[6]

Using the features generated by bootstrapping, the classifier was able to label 61% of the 1,335,240 terms in health and 81% of the 1,391,402 terms in computer science. The majority of unlabeled terms were extremely low frequency (typically 1). Higher frequency unlabeled terms were typically from categories other than those under consideration here (e.g., *john wiley, j. biochem, 2nd edition*). The distribution of category labels for the health and computer domains is shown in Table 2.

---

[4] We found relatively little evidence of explicit sentiment targeted at component and task aspects in patents and therefore focused our polarity analysis on attributes.

[5] Similar to Velardi's use of "domain relevance" and "consensus" (Velardi, 2001).

[6] Although it is possible to use KL-Divergence for feature selection, it is applied here solely for diagnostic purposes to verify that feature distributions match our intuitions with respect to the classification scheme.

Table 1. Features highly associated with classes (a[ttribute], c[omponent], t[ask]) in the health and computer science domains, along with an example of a term co-occurring with each feature in some patent.

Health			Computer Science		
Feature	**Class**	**Term**	**Feature**	**Class**	**Term**
prev_V=performed_during	t	biopsy	prev_V=automates	t	retrieval
prev_V=undergone	t	angioplasty	last_word=translation	t	axis translation
prev_V=suffer	a	hypertension	prev_Npr=reduction_in	a	power usage
prev_Npr=monitoring_of	a	alertness	Prev_Npr=degradation_in	a	audio quality
prev_V=binds_to	c	cytokines	prev_V=displayed_on	c	oscillograph
prev_Npr=salts_of	c	saccharin	last_word=information	c	customer information

Table 2. Number and percentage of category labels for health and computer domains (2002)

Category	Health	Computer Science
attribute	88,860 (10.8 %)	56,389 (6.5%)
component	680,034 (83.2%)	716,688 (83.2%)
task	48,002 (5.8 %)	88,786 (10.3%)

Tables 3a and 3b show examples of machine-labeled terms for the health and computer science domains. When terms were ranked by frequency, given a relatively relaxed domain specificity threshold (e.g., .05 for health), the top terms tended to capture broad semantic types relevant to the domain. As this threshold was increased (e.g., to 1.0 for health), the terms increased in specialization within each class.[7] As the table entries show, while the classification is not perfect, most terms fit the definitions of their respective classes. Note that in the health domain in particular, many of the "components" reflect objects acted upon by the invention, not just constituents of inventions themselves. Symptoms and diseases are interpreted as attributes because they are often measured according to severity and are targets for reduction.

Table 3a. Examples of ACT category results for health domain at two levels of domain specificity (ds).

Component (ds .05)	(ds 1.0)	Attribute (ds .05)	(ds 1.0)	Task (ds .05)	(ds 1.0)
patients, tissue, blood, diseases, drugs, skin, catheter, brain, tablets, organs	mitral valve, arterial blood, small incisions, pulmonary veins, anterior chamber, intraocular lens, ultrasound system, ultrasound energy, adenosine triphosphate, bone fragments	disease, infection, symptoms, pain, efficacy, side effects, inflammation, severity, death, blood flow	cosmetic properties, cardiac activity, urination, tissue temperature, gastric emptying, arousal neurotransmitter release, atrial arrhythmias, thrombogenicity ventricular pacing	treatment, administration, therapy, surgery, diagnosis, oral administration, implantation, stimulation, parenteral administration, surgical procedures	invasive procedure, ultrasound imaging, systole, anastomosis, spinal fusion, tissue ablation, image, reconstruction, cardiac pacing, mass analysis, spinal surgery

[7] The domain specificity thresholds chosen here differ between domains in order to compensate for the influence of the size of each domain's subcorpus on the terminology mix in the "generic" domain corpus against which domain specificity is measured. In the future, we plan to compensate directly for these size disparities in the score computation.

Table 3b. Examples of ACT category results for computer domain at two levels of domain specificity.

Component (ds 1.5)	(ds 3.0)	Attribute (ds 1.5)	(ds 3.0)	Task (ds 1.5)	(ds 3.0)
data, information, network, computer, users, memory, internet, software, program, processor	web applications, object access protocol, loans, memory subsystem, function call, obligations, source file, file formats, lender centralized database	errors, security, real time, traffic, overhead, delays, latency, burden, sales, copyright, protection	interest rate, resource utilization, resource consumption, temporal locality, system errors, transport layer security, performance bottleneck, processor capacity, cpu utilization, shannon limit	access, communication, execution, implementation, communications, management, task, tasks, stores, collection	network environments, business activities, database access, server process, search operation, client 's request, backup operation, project management, program development, document management

3.2 Polarity Classification

For the polarity classification task, the system assigned positive or negative polarity to 80,870 health and 73,289 computer science attributes. While not all the system labeled attributes merited their designation as attributes, the large quantity so labeled in each domain illustrates the vast number of conditions and dimensions for which inventions are striving to “move the needle” one way or the other, relative to attributes in the domain. Examples of the system’s polarity decisions are shown in Table 4. The system’s labels suggest that the default polarity of attributes in both domains is nearly evenly split.

Table 4. Examples of (pos)itive and (neg)ative polarity terms in health and computer science domains

Domain	# attributes	% of total	Examples
health ***pos***	43807	54%	ambulation, hemodynamic performance, atrial rate, anticoagulant activity, coaptation, blood oxygen saturation
neg	37063	46%	bronchospasm, thrombogenicity, ventricular pacing, withdrawal symptoms, fibrin formation, cardiac dysfunction
computer science ***pos***	32291	44%	transport layer security, processor capacity, cpu utilization, routability, network speeds, microprocessor performance
neg	40998	56%	identity theft, deadlocks, system overhead, memory fragmentation, risk exposure, bus contention, software development costs, network latencies, data entry errors

4 Evaluation and discussion

In order to evaluate the classification output, we first selected a subset of terms within each domain as candidates for evaluation based on the twin criteria of document frequency and domain specificity. That is, we wished to concentrate on terms with sufficient presence in the corpus as well as terms that were likely to express concepts of particular relevance to the domain. Using a frequency threshold of 10 this yielded 19,088 terms for the health corpus and 35,220 for computer science with domain specificity scores above .05 and 1.5 respectively. For each domain, two judges annotated approximately 150 random term instances with ACT judgments and approximately 100 machine-labeled attributes for polarity. The annotation tool displayed each term along with five random sentences from the corpus that contained the term, and asked the judge to choose the best label, given the contexts provided. An “other” option was available if the term fit none of the target categories. For the polarity task, the “other” label included cases where the attribute was neutral, could not be assigned a polarity, or was improperly assigned the category “attribute”. An adjudicated gold standard was compared to system labels to measure precision and recall, as shown in table 5.

Table 5a. Health domain: precision, recall and F-score for ACT and polarity classification tasks

Task	**Category**	**Precision**	**Recall**	**F-score**
ACT	attribute	.70	.44	.54
	component	.76	1.0	.86
	task	.86	.29	.43
Polarity	positive	.53	.85	.65
	negative	.77	.93	.84

Table 5b. Computer domain: precision, recall and F-score for ACT and polarity classification tasks

Task	**Category**	**Precision**	**Recall**	**F-score**
ACT	attribute	.80	.62	.70
	component	.86	.96	.90
	task	.43	.33	.38
Polarity	positive	.67	.88	.76
	negative	.75	.86	.80

Although the size of the evaluation set is small, we can make some observations from this sample. Precision in most cases is strong, which is important for the intended use of this data to characterize trends along each dimension using terminology statistics over time. The lower scores for tasks within the ACT classification may reflect the fact that the distinction between component and task is not always clear cut. The term “antivirus protection”, for example, describes a task but it is classified by the system as a component because it occurs with features like “prev_V=distribute” and “prev_V=provided_with”, which outweigh the contribution of the feature “last_word=protection” to select for the type task. To capture such cases of role ambiguity, it may be reasonable to assign some terms to multiple classes when the conditional probabilities for the two most probable classes are very close (as they are in this case). It may also be possible to integrate other forms of evidence, such as syntactic coordination patterns (Ziering, 2013) to refine system decisions.

One shortcoming of the current polarity classifier is that it does not attempt to identify attributes for which the polarity is neutral or dependent upon further context within the domain. For example, the attribute “body weight gain” is labeled as a negative. However, in the context of premature birth or cancer recovery, it may be actually be a positive attribute. Testing whether an attribute co-occurs with conflicting features (e.g., prev_V=increase and prev_V=decrease) could help spot such cases.

5 Related work

Text mining from patents has focused on identifying domain keywords and terminology for analytics (Tseng, 2007). Velardi’s (2001) approach, using statistics to determine domain relevance and consensus is very similar to that adopted here. We have also drawn inspiration from sentiment analysis, proposing an ontology for patents that reflects their review-like qualities (Liu, 2012). Most relevant is the work on discovering aspects and opinions relating to a particular subject such as a camera or restaurant (Kobayashi, 2007). There are many subtleties that have been studied in opinion mining research that we have finessed in our research here, such as detecting implicit sentiment and attributes not expressed as noun phrases. Wilson et al (2005, 2009) addressed the larger problem of determining contextual polarity for subjective expressions in general, putting considerable effort into the compilation of subjectivity clues and annotations. In contrast, our aim was to test whether we could substantially reduce the annotation effort when the task is focused on polarity labeling of attributes within patents. We hypothesized that the specialized role of patents might permit a more lightweight approach amenable to bootstrapping from a very small set of annotations and feature types.

Bootstrapping has been successfully applied to developing semantic lexicons containing a variety of concept types (Riloff, 1999; Thelen, 2002). It is often applied iteratively to learn new discriminative features after a set of high probability categorized terms are identified during an earlier round. While this increases recall, it also runs the risk of semantic drift if some terms are erroneously labeled. Given that the majority of unlabeled terms after a single round in our system are either extremely low frequency or not relevant to our ontology, we have not felt a need to run multiple iterations. Ziering (2013) used bootstrapping to identify instances of the classes *substance* and *disease* in patents, exploiting the tendency of syntactic coordination to relate noun phrases of the same semantic type. Given the general nature of coordination, a similar approach could be used to find corroborating evidence for the classifications that our system produces.

6 Conclusion

We have described an approach to text data mining from patents that strikes a middle ground between undifferentiated keywords and rich, domain specific ontologies. Motivated by the interpretation of patents as “positive reviews”, we have made use of generic lexico-syntactic features common across patent domains to bootstrap domain-specific classifiers capable of organizing terms according to their roles as components, tasks and attributes with polarity. Although the majority of keywords in a domain are categorized as components, the ontology puts tasks and attributes on an equal footing with components, thereby shifting the emphasis from devices and processes to the goals, obstacles and targets of inventions, information which could be valuable for analysts attempting to detect trends and make forecasts. In addition to more rigorous evaluation and tuning, future research directions include testing the approach across a wider range of technology domains, incorporation into time series analysis for forecasting, and mining relationships between terms from different categories to provide an even richer terminological landscape for analysts to work with.

Acknowledgements

This research is supported by the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior National Business Center (DoI/NBC) contract number D11PC20154. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoI/NBC, or the U.S. Government.

References

de Miranda, G. M. Coelho, Dos, and L. F. Filho. (2006) Text mining as a valuable tool in foresight exercises: A study on nanotechnology. Technological Forecasting and Social Change, 73(8):1013–1027.

Kobayashi, N., Inui, K. and Matsumoto, Y. (2007) Extracting aspect-evaluation and aspect-of relations in opinion mining, Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, Prague, Czech Republic, pp. 1065–1074.

Kullback, S. and Leibler, R. (1951). "On Information and Sufficiency". Annals of Mathematical Statistics 22 (1): 79–86.

Liu, B. (2012): Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies, Morgan & Claypool Publishers.

Manning, C., Raghavan, P. and Schütze, H. (2008) Introduction to Information Retrieval. Cambridge University Press.

Riloff, E. and Jones, R. (1999) Learning dictionaries for information extraction by multi-level bootstrapping. In Proceedings of the 16th National Conference on Artificial Intelligence and the 11th Innovative Applications of Artificial Intelligence Conference, pp. 474–479.

Riloff, E. and Shepherd, J. (1997) A corpus-based approach for building semantic lexicons. In Proceedings of the Second Conference on Empirical Methods in Natural Language Processing, pp. 117–124.

Shih, M.J., Liu, D.R., and Hsu, M.L. (2008) Mining Changes in Patent Trends for Competitive Intelligence. PAKDD 2008: 999-1005.

Sheremetyeva S. 2009. An Efficient Patent Keyword Extractor As Translation Resource Proceedings of the 3rd Workshop on Patent Translation in conjunction with MT-Summit XII Ottawa, Canada.

Thelen, M. and Riloff, E. (2002) A bootstrapping method for learning semantic lexicons using extraction pattern contexts. In Proceedings of the Conference on Empirical Methods in Natural Language.

Toutanova, K., Klein, D., Manning, C. and Singer, Y. (2003) Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network. In Proceedings of HLT-NAACL 2003, pp. 252-259.

Tseng, Y.-H., Lin, C.-J., and Lin, Y.-I. (2007). Text mining techniques for patent analysis. Information Processing & Management, 43(5):1216 – 1247.

Velardi, P., Fabriani, P. and Missikoff, M. (2001) FOIS '01 Proceedings of the international conference on Formal Ontology in Information Systems - Volume 2001, pp. 270-284.

Wilson, T., Wiebe, J and Hoffmann, P. (2005). Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. Joint Human Language Technology Conference and the Conference on Empirical Methods in Natural Language Processing (HLT-EMNLP-2005).

Wilson, T., Wiebe, J and Hoffmann, P. (2009). Recognizing Contextual Polarity: an exploration of features for phrase-level sentiment analysis. Computational Linguistics 35(3).

Woon, W. L., Henschel, A., and Madnick, S. (2009) A Framework for Technology Forecasting and Visualization. Working Paper CISL# 2009-11 , Massachusetts Institute of Technology.

Yang, S.Y., Lin, S.Y., Lin, S. N., Lee, C. F., Cheng, S. L., and Soo, V. W. (2008) Automatic extraction of semantic relations from patent claims. International Journal of Electronic Business Management, Vol. 6, No. 1, pp. 45-54 (2008) 45.

Ziering, P., van der Plas, L. and Schütze, H. (2013) Bootstrapping Semantic Lexicons for Technical Domains. In Proceedings of the Sixth International Joint Conference on Natural Language Processing, pp. 844–848, Nagoya, Japan, October 2013. Asian Federation of Natural Language Processing.