

Inter-domain Opinion Phrase Extraction Based on Feature Augmentation

Gábor Berend¹, István Nagy T.¹, György Móra¹ and Veronika Vincze^{1,2}

¹Department of Informatics, University of Szeged

{berendg, nistvan, gymora}@inf.u-szeged.hu

²Hungarian Academy of Sciences, Research Group on Artificial Intelligence

vinczev@inf.u-szeged.hu

Abstract

In this paper, a system for the extraction of key argument phrases – which make the opinion holder feel negative or positive towards a particular product – from product reviews is introduced. Since the necessary amount of training examples from any arbitrary product type (target domain) is not always available, the possible usage of domain adaptation in the task of opinion phrase extraction is also examined. Experimental results show that models relying on training examples mainly from a different domain can still yield results that are comparable to that of the intra-domain settings.

1 Introduction

There has been a growing interest in the NLP treatment of subjectivity and sentiment analysis (see e.g. Balahur et al. (2011)) and that of keyphrase extraction, e.g. Kim et al. (2010). Product reviews serve as perfect objects for the combination of the above mentioned research areas as the opinion bearing phrases of a product review can be interpreted analogously to regular keyphrases of scientific documents, i.e. in both cases proper phrases have decisive role within the document where they were present. The fact that some review portals have the possibility to leave a set of pro and con phrases underlines this resemblance between opinion phrases and scientific keyphrases.

However, despite the somewhat common nature of opinion phrases and keyphrases, methods that work on the well studied field of scientific keyphrase extraction are not necessarily successful in the extraction of opinion phrases from product reviews. On the one hand, although proper phrases have their decisive role in both types of genres, opinion phrases are the ones that form the sentiments of the opinion holder, whereas in the case of scientific keyphrases they should be such phrases that summarize well the content of a document. Note the difference between opinion-forming phrases and those which summarize well the content of a document, i.e. one can frequently

use such phrases in a review that does not have much importance in the opinion-forming aspect, whereas in the case of scientific documents frequently used phrases tend to be proper keyphrases as well.

Most of the standard keyphrase extraction algorithms employ supervised learning, which makes the accessibility of training instances generated from reviews and the sets of opinion phrases assigned to them prerequisite. In the case of training an opinion phrase extractor on one domain, this criterion is not easily fulfilled in every case, due to the fact that it is not necessary that one can find abundant training examples for any kind of product types. For this reason, exploiting domain adaptation techniques during the task of opinion phrase mining among different domains might be useful. This paper examines the possible utility of domain adaptation in the inter-domain opinion phrase mining task.

2 Related Work

There have been many studies on opinion mining (Turney, 2002; Pang et al., 2002; Titov and McDonald, 2008; Liu and Seneff, 2009). Our approach relates to previous work on the extraction of reasons for opinions. Most of these papers treat the task of mining reasons from product reviews as one of identifying sentences that express the author's negative or positive feelings (Hu and Liu, 2004a; Popescu and Etzioni, 2005). This paper is clearly distinguishable from previous opinion mining systems as our goal is to find the reasons for opinions expressed and we aim the task of phrase extraction instead of sentence recognition.

This work differs in important aspects even from the frequent pattern mining-based approach of Hu and Liu (2004b), since they regarded the main task of mining opinion features with respect to a group of products, not individually at review-level as we did. Even if an opinion feature phrase

is feasible for a given product-type, it is not necessary that all of its occurrences are accompanied with sentiments expressed towards it (e.g. *The phone comes in red and black colors*, where *color* could be an appropriate product feature).

The approach presented here differs from these studies in the sense that it looks for the reason phrases themselves review by review, instead of multi-labeling some aspects. These approaches are intended for applications used by companies who would like to obtain a general overview about a product or would like to monitor the polarity relating to their products in a particular community. In contrast, we introduce here a keyphrase extraction-based approach which works at the document level as it extracts keyphrases from reviews which are handled independently of each other. This approach is more appropriate for the consumers, who would like to be informed before purchasing some product.

The work of Kim and Hovy (2006) lies probably the closest to ours. They addressed the task of extracting con and pro sentences, i.e. the sentences on why the reviewers liked or disliked the product. They also note that such pro and con expressions can differ from positive and negative opinion expressions as factual sentences can also be reason sentences (e.g. *Video drains battery.*). Here the difference is that they extracted sentences, but we targeted phrase extraction.

Most of the keyphrase extraction approaches (Witten et al., 1999; Turney, 2003; Medelyan et al., 2009; Kim et al., 2010) extract phrases from one document that are the most characteristic of its content. In these supervised approaches keyphrase extraction is regarded as a classification task, in which certain n-grams of a specific document function as keyphrase candidates, and the task is to classify them as proper or improper keyphrases. Here, our task formalization of keyphrase extraction is adapted from this line of research for opinion mining and we focus on the extraction of argument phrases from product reviews that induce sentiments in its author. As community generated pros and cons can provide training samples and our goal is to extract the users' own words, here we also follow this supervised keyphrase extraction procedure.

As stated earlier, abundant training examples are not necessarily available from a single domain (product type) in the case of opinion phrase ex-

traction, so domain adaptation techniques might be useful in the detection of opinion phrases. Formally, in the case of domain adaptation we are given two sets of instances, $S \subseteq D_S \in \mathbb{R}^n$ and $T \subseteq D_T \in \mathbb{R}^m$, D_S and D_T being the feature spaces of the source and target domain and S and T the set of source and target instances, respectively. Typically $|S| \gg |T|$ also holds for the sizes of the two distinct domains.

As a possible solution for domain adaptation Daumé and Marcu (2006) proposes an approach which learns three separate models, one for the source specific, target specific and general information as well. They also report that the usage of EM for the training of the models can be computationally costly.

Although the feature augmentation technique of Daumé (2007) uses a similar intuition (i.e. the existence of source-, target specific and general information), it is much simpler as it learns one model including both source and target domain instances in an extended feature space, instead of learning three models at a time. Here the original feature space is mapped to a higher-dimension space, so that source and target domain and general information are incorporated. To achieve this, the mapping Φ_S or Φ_T is employed to every instance x from the original feature space, depending on the fact whether the original vector x is representing a source or a target domain instance, respectively. The two mappings are of the forms $\Phi_S(x) = \langle x, x, \mathbf{0} \rangle$ and $\Phi_T(x) = \langle x, \mathbf{0}, x \rangle$, where $\mathbf{0}$ is the null vector.

3 Opinion Phrase Extraction

Experiments were inspired by the standard – mainly scientific – keyphrase extraction systems. In these systems, such in KEA (Witten et al., 1999) or Turney (2003), the extraction of such phrases (i.e. keyphrases) that circumscribe the main content of individual documents is regarded as a supervised learning task, where the author or reader-assigned keyphrases are used as positive training examples.

Here we adapted these standard scientific keyphrase extraction approaches to the task of opinion phrase extraction, however, in our case training examples were such phrases that make the author feel negative or positive towards a given object. Our setting was also similar to standard keyphrase extraction as the task of opinion phrase

extraction was regarded as a supervised learning task, where training instances are generated from consecutive n-grams of product reviews. Although the opinion phrase extraction setting shows resemblance to scientific keyphrase extraction, the different nature of scientific keyphrases compared to opinion phrases makes different approaches reasonable.

3.1 Feature Space

In our supervised learning approach, opinion phrase candidates were described by a set of features that were used in a MALLET (McCallum, 2002) implementation of the Maximum Entropy classifier. Opinionated phrases were finally determined by regarding those candidate phrases that were among the (top-5, 10 and 15) highest rated phrases based on the probability,

$$P(Class = +|X) = \frac{\exp(\sum_i^n \lambda_i f_i(+, X))}{\sum_{c \in C} \exp(\sum_i^n \lambda_i f_i(c, X))}$$

, where X is the feature vector describing a candidate phrase, n is the dimension of the feature space, the set $C = \{+, -\}$ refers to the set of possible outcome classes of an instance (i.e. proper and improper opinion phrases), λ_i is the weight determined by the model for the i^{th} feature and $f_i(c, X)$ is the feature function with respect to a class label c and the input vector X .

3.1.1 General Features

Since we assumed that the underlying principles of extracting opinionated phrases are similar to some extent to the extraction of standard (mostly scientific) keyphrases, features of the standard setting were applied in this task as well. The most common ones, introduced by KEA (Witten et al., 1999) are the **Tf-idf** value and the **relative position** of the first occurrence of a candidate phrase within a document. We should note that KEA is primarily designed for keyphrase extraction from scientific publications and whereas the position of the first occurrence might be indicative in research papers, product reviews usually do not contain a summarizing “abstract” at the beginning. For these reasons we chose these features as the ones which form our baseline system. **Phrase length** is also a common feature, which was defined here as the number of the non-stopword tokens of an opinion candidate phrase.

3.1.2 Task Specific Features

Due to the differences pointed out so far, different features can attribute to opinion phrase extraction compared to scientific keyphrase extraction. This subsection is dedicated to present some of the novel features that were introduced to favor the unique characteristics of opinion phrase extraction.

Opinionated phrases often bear special orthographic characteristics, e.g. in the case of *so sloooow* or *CHEAP*. Features that represent this phenomenon were also incorporated in the feature space: the first feature is responsible for **character runs** (i.e. more than 2 of the same consecutive characters), and another is responsible for **strange capitalization** (i.e. the presence of uppercase characters besides the initial one).

One feature used external information on the individual tokens of a candidate phrase. It relied on the **sentiment scores** of SentiWordNet (Baccianella et al., 2010), a publicly available database that contains a subset of the synsets of the Princeton Wordnet with positivity, negativity and neutrality scores assigned to each one, depending on the use of its sentiment orientation (which can be regarded as the probability of a phrase belonging to a synset being mentioned in a positive, negative or neutral context). These scores were utilized for the calculation of the sentiment orientations of each token of a candidate phrase. Surface-based SentiWordNet-calculated feature values for a candidate phrase included the *maximal positivity and negativity and subjectivity* scores of the individual tokens and the *total sum* over all the tokens of one phrase.

Sentence-based features were also defined based on SentiWordNet. Previous studies have shown that upon extracting keyphrases from scientific documents, the use of external knowledge such as checking Wikipedia to see whether there exists an article that has the same title as a candidate phrase can be beneficial. One possible use of SentiWordNet seems somewhat analogous to these findings since it was also used to gather **indicator terms** from sentences. Those elements of SentiWordNet synsets were gathered as potential indicator words for which the sum of the average positivity and negativity sentiments scores among all its synsets were above 0.5 (i.e. whose word forms are more likely to have some kind of polarity). Then for a given candidate phrase of a

	Mobiles	Movies
Number of reviews	2,009	1,962
Sentences/review	31.9	29.8
Tokens/sentence	16.1	17.0
Keyphrases/review	4.7	3.2
Candidate phrases/review	130.38	135.89

Table 1: Various statistics on the size of the corpora

given document, a true value was assigned to the SentiWordNet-derived indicator features that had at least one co-occurrence within the same sentence within the review of the candidate phrase.

SentiWordnet was also used to investigate the entire sentences that contained a phrase candidate. This kind of feature calculated the sum of every sentiment score in each sentence where a given candidate phrase was present. Then the **mean** and the deviation of the sum of the sentiment scores were calculated for each token of the phrase-containing sentences and assigned to the candidate phrase. The mean of the sentiment scores of the individual sentences yielded a general score on the **sentiment orientation** of the sentences containing a candidate phrase, while higher values for the **deviation** was intended to capture cases when a reviewer writes both factual (i.e. uses few opinionated words) and non-factual (i.e. uses more emotional phrases and opinions) sentences about a product.

A more detailed description on the framework and evaluation results dealing with the intra-domain setting (including human evaluation as well) can be found in Berend (2011). In addition to that system, here the feature augmentation technique of) was applied to improve inter-domain results.

4 Evaluation

Evaluation was carried out on two fairly different domains, i.e. on reviews dealing with mobile phones and movies from the site `epinions.com`. Section 4.1 presents the dataset of product reviews and the way the set of proper opinion phrases (which served as positive training examples) were determined for its elements, and Section 4.2 describes experimental results achieved on that dataset. The evaluation procedure was strict in the sense that only perfect matches (after some normalization step) were accepted, i.e. the normal-

ized version of an opinion phrase returned from a document must be identical to at least one of the normalized versions from the set of refined author keyphrases (pros and cons) of the very document.

4.1 Dataset

In our experiments, we crawled two quite different domains of product reviews, i.e. mobile phone and movie reviews from the review portal `epinions.com`. For both domains, 2000 reviews were crawled from `epinions.com` and an additional of 50 and 75 reviews, respectively.¹ This corpus is quite noisy (similarly to other user-generated contents); run-on sentences and improper punctuation were very common, as well as grammatically incorrect sentences since reviews were often written by non-native English speakers.

The list of pros and cons was inconsistent too in the sense that some reviewers used full sentences to express their opinions, while usually a few token-long phrases were given by others. The segmentation of their elements was marked in various ways among reviews (e.g. comma, semicolon, ampersand or the *and* token) and even differed sometimes within the very same review. There were many general or uninformative pros and cons (like *none* or *everything* as a pro phrase) as well.

In order to have a consistent gold-standard annotation for training and evaluation, we refined the pros and cons of the reviews in the corpora. In the first step, the **segmentation** of pros and cons was manually checked by human annotators. Our automatic segmentation method split the lines containing pros and cons along the most frequent separators. This segmentation was corrected by the annotators in 7.5% of the reviews. Then the human annotators also marked the general pros and cons (11.1% of the pro and con phrases) and the reviews without any identified keyphrases were discarded.

Linguistic analysis included the POS tagging (Toutanova and Manning, 2000) and syntactic parsing (Klein and Manning, 2003) of the reviews using Stanford CoreNLP.

4.2 Experimental Results

Several experiments were conducted in order to see the effect of domain adaptation in the opinion phrase extracting task. Where not stated differently experiments were carried out in 10-fold

¹All the data used in our experiments are available at <http://rgai.inf.u-szeged.hu/proCon>

	Mobiles			Movies		
	P	R	F	P	R	F
<i>Baseline</i>	1.72	1.84	1.77	1.21	1.93	1.49
<i>Target</i>	14.8	15.7	15.27	10.0	15.8	12.22
<i>Source</i>	3.5	3.7	3.58	3.2	5.0	3.92
<i>Mixed</i>	11.1	11.8	11.46	6.5	10.3	8.0
<i>Mixed_{DA}</i>	12.7	13.4	13.04	7.2	11.3	8.84

Table 2: Results obtained on the mobile and movie dataset relying on the top-5 ranked phrases.

cross validation, the results of which are present in Tables 2, 3 and 4. Results are reported in the form of Precision, Recall and F-score (indicated with P, R and F, respectively) at the levels of top-5, 10 and 15-ranked keyphrases for both mobile phones and movies. In the tables the row *Baseline* corresponds to that intra-domain setting when only the two standard features (i.e. tf-idf and position of first occurrence) were used. As for the rows *Target*, *Source* and *Mixed* the extended feature space was utilized and they indicate that the training and testing domains were the same, differed and originated from both source and target domains, respectively. The row *Mixed_{DA}* refers to the result when source and target domain documents were incorporated among the training instances (similarly to *Mixed*) and feature augmentation-based domain adaptation was applied as well.

First of all, before conducting experiments involving domain adaptation, the intra-domain performance of the opinion phrase extraction system was measured on the datasets. Intra-domain evaluation refers to the fact that during these runs all the instances of the training set were derived from the very same domain as the test instances. Obviously, these results can serve as an upper bound on the final results which used domain adaptation, i.e. on a more noisy training set. These results are found in the row *Target*.

Secondly, that case was investigated when the weights of the Maximum Entropy model that fitted the training instances the best were based on the different domain compared to the domain of the evaluation, meaning that no instances originating from the target domain were present during the creation of a particular model. Evaluating elements of the target domain based on the model that was learnt on a different source domain is present in the rows *Source* of Tables 2, 3 and 4. In this case it was not necessary to apply 10-fold cross validation, since the evaluation and the training of models took place on entirely different domains.

	Mobiles			Movies		
	P	R	F	P	R	F
<i>Baseline</i>	1.42	3.04	1.94	0.98	3.13	1.5
<i>Target</i>	10.4	22.0	14.11	7.0	21.9	10.63
<i>Source</i>	3.6	7.7	4.93	2.7	8.5	4.1
<i>Mixed</i>	8.0	16.9	10.82	4.6	14.6	7.05
<i>Mixed_{DA}</i>	8.6	18.3	11.72	5.0	15.8	7.65

Table 3: Results obtained on the mobile and movie dataset relying on the top-10 ranked phrases.

	Mobiles			Movies		
	P	R	F	P	R	F
<i>Baseline</i>	1.39	4.48	2.12	0.89	4.26	1.48
<i>Target</i>	8.0	25.4	12.17	5.3	24.6	8.67
<i>Source</i>	3.6	11.4	5.44	2.4	11.2	3.92
<i>Mixed</i>	11.1	11.8	11.46	3.7	17.4	6.13
<i>Mixed_{DA}</i>	6.7	21.2	10.17	4.0	18.6	6.53

Table 4: Results obtained on the mobile and movie dataset relying on the top-15 ranked phrases.

The row *Mixed* contains result achieved when models were created in such a manner that during 10 runs 10% of the target domain instances (choosing different elements every time) were added to the set of all the source domain instances. In these cases the evaluation took place on the remaining 90% of the target domain that were not selected to be added to the instances for the training originating from the source domain.

In the case of the results in the row *Mixed+DA* the selection of training and test instances was carried out exactly the same way as described in the case of the row *Mixed*, but this time the feature space was augmented as described in Daumé (2007) that is briefly outlined at the end of Section 2.

5 Discussion

Intra-domain results can be interpreted as an upper bound for a system that is based on domain adaptation, due to the fact that in the intra-domain setting data points that make up the set of training instances are drawn from the same distribution as the test instances. Similarly, when instances originating from a different source domain are added, it can easily bias the model on which predictions are based.

Best results in the intra-domain setting around an F-score of 15 might not seem so solid for the first time, but for the proper judgement of these results, it is worth to know that at the shared task of SemEval-2010 (Kim et al., 2010) that dealt with the extraction of keyphrases from scientific publi-

cations, the best performing system achieved an F-score of 19.3 when evaluating it against the top-15 author keywords. Naturally, product reviews are far more noisy and heterogeneous in language than scientific publications, and the determination of keyphrase-behaving opinion reasons is far more ambiguous and difficult. It is also true that the language of product reviews is more ‘creative’, i.e. there are more possibilities to express proper and similarly functioning keyphrases compared to the scientific genre, which makes exact match-based evaluation more prone to underestimate the results in the case of opinionated texts.

The fact that the highest F-scores for keyphrases are achieved when the number of extracted phrases is around the average number of pro and con phrases per reviews (i.e. between 4.7 and 3.2 for mobiles and movies, respectively) also suggests that our ordering of keyphrase candidates is quite effective (since once we find the number of keyphrases a document has, performance cannot really grow anymore).

It is also unequivocal from the results of the rows *Source* of Tables 2, 3 and 4 that training a model solely on one source domain (without any target domain instances) and evaluating it on a different target domain causes severe drop in performance. Despite the serious decline in the result in the latter settings, giving a small set of target domain documents (having a size equalling only to 10% of the size of the source domain documents) yields much better results.

However, since $|S| \gg |T|$, it is still true that the effect of adding elements from T to the training set is easily oppressed by the much higher mass of the element of S . It is shown that the simple, yet efficient method of feature augmentation can still help, yielding final domain-adaptation results that are comparable to those results when the training and the testing took place within the same domain.

Besides all, it can also be seen that the domain of mobile phones seems to be an easier task (which was confirmed by human annotator agreement rates as well).

6 Conclusions and Future Work

In this paper an extension of the standard scientific keyphrase extraction was introduced, and a possible way to overcome the absence of abundant tagged training examples was shown, using the simple method of feature augmentation. Using

the simple feature augmentation domain adaptation technique, results achieved on the target domain were comparable to those settings when the parameters of our model were estimated on a large set of instances from the very same domain as the test instances. However, this highly idealistic assumption that one has access to a fair amount of training material from the domain of the target documents is not always met. In these cases domain adaptation approaches seem to be useful.

The basic idea of treating opinion phrases similarly to scientific keyphrases raises the question whether domain adaptation methods would work in the aspect of scientific articles and product reviews as well. Although these two genres definitely seem to be more distant from each other than two sets of reviews dealing with different product families, we find it as one possible way to extend this work to thoroughly examine this particular question.

Acknowledgments

This work was supported by the Project “TÁMOP-4.2.1/B-09/1/KONV-2010-0005 – Creating the Center of Excellence at the University of Szeged”, supported by the European Union and co-financed by the European Regional Development Fund and by the project BELAMI financed by the National Innovation Office of the Hungarian government.

References

- Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC’10)*, Valletta, Malta, May. European Language Resources Association (ELRA).
- Alexandra Balahur, Ester Boldrini, Andres Montoyo, and Patricio Martinez-Barco, editors. 2011. *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA 2.011)*. Association for Computational Linguistics, Portland, Oregon, June.
- Gábor Berend. 2011. Opinion expression mining by exploiting keyphrase extraction. In *Proceedings of the 5th International Joint Conference on Natural Language Processing, IJCNLP’11*, Chiang Mai, Thailand, November.
- Hal Daumé, III and Daniel Marcu. 2006. Domain adaptation for statistical classifiers. *J. Artif. Int. Res.*, 26:101–126, May.

- Hal Daumé, III. 2007. Frustratingly Easy Domain Adaptation. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 256–263, Prague, Czech Republic, June. Association for Computational Linguistics.
- Minqing Hu and Bing Liu. 2004a. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '04, pages 168–177, New York, NY, USA. ACM.
- Minqing Hu and Bing Liu. 2004b. Mining opinion features in customer reviews. In *Proceedings of the 19th national conference on Artificial intelligence*, AAAI'04, pages 755–760. AAAI Press.
- Soo-Min Kim and Eduard Hovy. 2006. Automatic identification of pro and con reasons in online reviews. In *Proceedings of the COLING/ACL 2006 Main Conference Poster Sessions*, pages 483–490, Sydney, Australia, July. Association for Computational Linguistics.
- Su Nam Kim, Olena Medelyan, Min-Yen Kan, and Timothy Baldwin. 2010. Semeval-2010 task 5: Automatic keyphrase extraction from scientific articles. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, SemEval '10, pages 21–26, Morristown, NJ, USA. Association for Computational Linguistics.
- Dan Klein and Christopher D. Manning. 2003. Accurate unlexicalized parsing. In *Proceedings of the 41st ACL*, pages 423–430.
- Jingjing Liu and Stephanie Seneff. 2009. Review sentiment scoring via a parse-and-paraphrase paradigm. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 161–169, Singapore, August. Association for Computational Linguistics.
- Andrew Kachites McCallum. 2002. Mallet: A machine learning for language toolkit. <http://mallet.cs.umass.edu>.
- Olena Medelyan, Eibe Frank, and Ian H. Witten. 2009. Human-competitive tagging using automatic keyphrase extraction. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 1318–1327, Singapore, August. Association for Computational Linguistics.
- Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up?: sentiment classification using machine learning techniques. In *EMNLP '02: Proceedings of the ACL-02 conference on Empirical methods in natural language processing*, pages 79–86, Morristown, NJ, USA. Association for Computational Linguistics.
- Ana-Maria Popescu and Oren Etzioni. 2005. Extracting product features and opinions from reviews. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 339–346, Vancouver, British Columbia, Canada, October. Association for Computational Linguistics.
- Ivan Titov and Ryan McDonald. 2008. A joint model of text and aspect ratings for sentiment summarization. In *Proceedings of ACL-08: HLT*, pages 308–316, Columbus, Ohio, June. Association for Computational Linguistics.
- Kristina Toutanova and Christopher D. Manning. 2000. Enriching the knowledge sources used in a maximum entropy part-of-speech tagger. In *Proceedings of the 2000 Joint SIGDAT conference on Empirical methods in natural language processing and very large corpora*, EMNLP '00, pages 63–70, Stroudsburg, PA, USA. ACL.
- Peter Turney. 2002. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 417–424.
- Peter Turney. 2003. Coherent keyphrase extraction via web mining. In *Proceedings of IJCAI*, pages 434–439.
- Ian H. Witten, Gordon W. Paynter, Eibe Frank, Carl Gutwin, and Craig. 1999. Kea: Practical automatic keyphrase extraction. In *ACM DL*, pages 254–255.