Analysis of Adjective-Noun Word Pair Extraction Methods for Online Review Summarization

Koji Yatani, Michael Novati, Andrew Trusty, and Khai N. Truong
Department of Computer Science, University of Toronto
Toronto, Canada
{koji, atrusty, khai}@dgp.toronto.edu, michael@novati.ca

Abstract

Many people read online reviews written by other users to learn more about a product or venue. However, the overwhelming amount of user-generated reviews and variance in length, detail and quality across the reviews make it difficult to glean useful information. In this paper, we present a summarization system called Review Spotlight. It provides a brief overview of reviews by using adjective-noun word pairs extracted from the review text. The system also allows the user to click any word pair to read the original sentences from which the word pair was extracted. We present our system implementation as a Google Chrome browser extension, and an evaluation on how two word pair scoring methods (TF and TF-IDF) affect the identification of useful word pairs.

1 Introduction

Online review websites offer users a wealth of perspectives about products that can be purchased or locations that can be visited (which we will refer to as reviewed entities). However, the number of reviews for any given entity often extends well beyond what users can read quickly. User-generated reviews are also known to be unstructured documents [Jo and Oh, 2011]; they vary greatly in length, detail and focus in comparison to ones written by professional editors. These issues make it difficult for users to quickly and easily glean useful details about the entity.

There are several ways to provide a brief overview of user-generated reviews. For example, often reviewers are asked to rate their overall impression (usually from 1 to 5) of an entity. Although this rating gives the reader a quick understanding of how much the reviewer liked or disliked the entity, it does not offer information about why that rating was given. Several websites also allow readers to rate the usefulness of posted reviews, and order reviews by this usefulness rating. These features allow the user to browse reviews in a way that differs from needing to read them all. However, browsing in this manner may result in the user missing information that she could find to be important to her (*e.g.,* comments from the more recent reviews). Review websites may also provide separate ratings and reviews for different aspects or attributes of an entity (*e.g.*, "price," "food quality," and "atmosphere" for a restaurant review). However, most review websites let reviewers post their comments without any specific structure; thus automatic processing of reviews for such aspects is still challenging [Jo and Oh, 2011].

Instead of aggregating reviews for each aspect or topic for summarization, we focus on improving a tag cloud interface. A tag cloud is a visualization using a set of words (or "tags"), and is used in many online services including review websites. The size and color of each tag are associated with the importance or relevance in the original text. A tag cloud is considered useful for different user tasks, which include building an impression about the entity that the tag cloud visualizes [Rivadeneira *et al.*, 2007]. However, we argue that a standard tag cloud is not necessarily appropriate for summarization of online user-generated reviews. Figure 1a shows a tag cloud simply based on the frequency of words in reviews for a restaurant. Although it is possible to learn general information about the restaurant from this tag cloud (*e.g.*, it is a Japanese restaurant and the meals are probably good), the details that may be important for users are hard to identify for inclusion. For instance, "roll" is a term mentioned frequently in the reviews, but it is not clear what types of rolls were mentioned most often by reviewers.

Figure 1. A visual comparison of a standard tag cloud (a) and our Review Spotlight interface using adjective-noun word pairs (b). Both were created from the same review text. Our interface provides a more detailed overview than what a standard tag cloud is able to convey.

We developed a system which provides a quick overview of user-generated reviews called Review Spotlight. The system displays a list of adjective and noun word pairs that appear most frequently in the review text as shown in Figure 1b. With adjective-noun word pairs, the user can view more detailed information about the reviewed entity than with single words. Review Spotlight also performs a sentiment analysis on the extracted word pairs and uses different font colors to represent the level of positivity or negativity of each word pair. Review Spotlight also allows the user to click on a word pair to see additional contexts in which it is used in the original review text. The user interface does not arrange word pairs in a specific order so that the user can serendipitously acquire information even from word pairs with small fonts.

In this paper, we describe the implementation of our Review Spotlight system. We also present our evaluation on how different scoring methods affect the identification of useful word pairs. We then discuss possible improvements on summarization of online user-generated reviews using word pairs.

2 Related Work

Although user review summarization has been investigated heavily in computational linguistics [Hu and Liu, 2004], user interfaces employing it have not been studied as extensively. One approach for user review summarization is to collect opinions on different features in an entity (*e.g.,* food or service in a restaurant review). Liu *et al.* [2005] developed a system to visualize how many positive or negative reviews were posted about features of an entity using bar graphs, but the system was not evaluated from the user interface perspective. Carenini *et al.* [2006] used a Treemap visualization to show the information extracted from user-generated reviews organized in a tree structure based on the features. However, their user study showed that the participants were often confused by the Treemap visualization, and preferred text-based summarization. They also developed a system similar to Liu *et al.*'s work [2005], but they tailored it to allow the user to compare multiple entities [Carenini *et al.*, 2009]. Their user interface shows the distribution of positive and negative reviews along different features. However, they did not formally study the efficacy of their user interface.

A computational linguistics method often used for summarizing user-generated reviews is a sentiment analysis, which determines the semantic orientation of a given text. Turney [2002] and Pang *et al.* [2002] applied a sentiment analysis technique to analyze review text. Both systems used machine learning techniques to identify the semantic orientation of the phrases extracted using *n*-gram methods.

There are several systems that have applied sentiment analysis for tag cloud visualization. Dave *et al.* [2003] built a system to extract the tags from product reviews and display a sentiment score calculated based on those tags. They also used the *n*-gram methods (n=1~3) for extracting positive and negative features. Lee *et al.* [2007] developed a system in which the user can manually add tags to an entity, and can rate whether the added tag contains a positive or negative sentiment; the rated positivity/negativity of the tag is visualized using the font. Ganesan *et al.* [2008] incorporated emoticons into a tag cloud visualizing eBay seller feedback. For example, a smiley face is automatically added to a tag that their system recognized as a positive tag. Although the effect of word sentiment visualization has not been studied in detail previously, we believe that incorporating a sentiment orientation into a tag cloud visualization could be useful in allowing the user to quickly understand how positively or negatively the entity is reviewed.

3 Review Spotlight System

3.1 High-level Interface Design

Figure 2a shows the interface of the first Review Spotlight prototype. Based on findings from a formative study that we conducted [Yatani *et al.*, 2011], we focus on presenting the adjective-noun pairs frequently mentioned in the reviews. The font size of each word pair is set to be proportional to its number of occurrences. Review Spotlight also uses color to visualize the sentiment of each word, which will be discussed later.

Figure 2. Review Spotlight interfaces: a) First prototype. By clicking a word pair, the user can see the original sentences which the clicked word pair came from; b) Second prototype as a Google Chrome browser extension. In addition to the word pair and sentence display, it has a filter interface (checkboxes) for positive, negative, and objective word pairs and the histogram to control the time period of the reviews to be displayed in the interface.

When the user moves the cursor over an adjective-noun pair, Review Spotlight shows as many as four adjectives that are most frequently paired with that noun. In the example shown in Figure 2a, Review Spotlight shows "ridiculous," "worth," "bad," and "terrible" when the cursor is over the word pair "long wait." When the user clicks adjectives, the interface displays the number of occurrences of that adjective-noun word pair, and the sentences in which it appears in a separate textbox. Thus, Review Spotlight supports quick examination of review details to enable the user to test her impressions.

We then developed the second prototype as a Google Chrome browser extension (Figure 2b). We added several features in this prototype. The interface includes a histogram showing the number of reviews posted for a given entity along a timeline; this gives the user an overview of how the number of the reviews changes over time. The slider below the histogram allows the user to specify the time period for which Review Spotlight should summarize the reviews. The interface also provides checkboxes for selecting sentiment types ("positive," "negative," and "objective"). The user can specify which types of word pairs should be displayed by clicking the corresponding checkboxes.

3.2 Implementation

To generate a summarization from the review text, Review Spotlight performs the following five steps: 1) obtain the review text; 2) extracts adjective-noun word pairs; 3) counts each word pair's occurrences; 4) performs a sentiment analysis of each word pair; and 5) displays the word pairs. In this section, we explain how we implemented this procedure for the second prototype (Figure 2b). This prototype has two components: the client (a Google Chrome extension) and the parser server.

Review Spotlight first collects all the reviews for the given entity from review websites (*e.g.*, Yelp.com or Amazon.com). When Review Spotlight gets an HTML source of the first review page, it extracts the total number of the reviews. It also parses the source to find the reviews and stores them with the posted date. Review Spotlight also requests the next page to the Web server of review websites and parses its HTML source until it obtains all the reviews.

Review Spotlight then extracts adjective-noun word pairs from the review text. The extension sends each sentence to the server we built. We used WebSocket to support fast communication with the server. Our server passes the sentence to a backend computer cluster to parallelize part-of-speech (POS) tagging. Using a POS tagger developed by Tsuruoka and Tsujii [Tsuruoka and Tsujii, 2005], our parser located in the cluster nodes labels the part of speech for the words in the sentence, and returns it to the Review Spotlight client through the server. The client then pairs a noun with the closest adjective modifying it. Review Spotlight also extracts a word pair from a sentence with the "to be" verb. For instance, if Review Spotlight parses a sentence "The food is great," it extracts the word pair "great food." In addition, by focusing on adjective-noun pairs, Review Spotlight intrinsically removes noise introduced by common trivial words, such as articles and prepositions.

Review Spotlight then counts the number of occurrences of each word pair and groups word pairs by nouns. Next, the system eliminates the word pairs that only appear once in the original review text. It then calculates the font size for the extracted adjectives and nouns. The font size for a noun is determined linearly by its number of occurrences. The font size for an adjective is determined linearly by the number of occurrences of the word pair consisting of it and the associated noun. We set the minimum and maximum font sizes to 10 and 30 pixels, respectively.

Next, Review Spotlight performs a sentiment analysis on the word pairs using SentiWordNet 1.0 [Esuli and Sebastiani, 2006], a context-free, word-based sentiment analysis tool. A sentiment value provided by SentiWordNet 1.0 for each word consists of three scores (*i.e.,* positivity, negativity, and objectivity), and it is defined for each common use of the word. Review Spotlight first calculates the sentiment value for an adjective by taking the average of its sentiment values for all the use contexts. It then calculates the sentiment value for a noun by taking the average of the sentiment values of all paired adjectives weighted by the number of occurrences. It maps the sentiment value into the color scheme in which shades of green, red, and blue represent positive, negative, and neutral meaning, respectively; the darkness of the shade conveys the sentiment strength. Through a preliminary experiment with the prototype system, we determined that users preferred the noun coloring based on this weighted average over the coloring based on the average of the sentiment values defined in SentiWordNet.

After the sentiment analysis, Review Spotlight performs a spatial allocation function to place the extracted word pairs within a given space (600 x 250 pixels by default, but the dimensions can be adjusted). Review Spotlight places the word pairs randomly so that the user is not biased to any specific terms based on their placement position. Review Spotlight also adds padding around each word pair that is relative in size to the bounding box of the word pair to avoid visual complexity. Review Spotlight performs this spatial allocation, starting with the largest word pair to the smallest, until it cannot find a location for a new word pair which does not cause an overlap on any other word pairs that have been placed already. Review Spotlight then combines up to four adjectives that are most frequently paired with the noun. These adjectives become visible when the user moves the cursor over the word pair.

The Review Spotlight client re-renders the interface every time it has parsed five new reviews and updated the counts of the word pairs. In this way, the user can have a summarization without waiting for all the reviews being parsed.

4 User Study

We conducted two evaluations related to the usability of Review Spotlight. We briefly describe the interface evaluation and its findings, and focus on the evaluation of how well two different word pair scoring methods could emphasize useful word pairs.

4.1 Interface Evaluation

We conducted a laboratory study to evaluate how Review Spotlight addresses user requirements for an interface summarizing user-generated reviews compared to traditional review pages. Participants were asked to read reviews of two restaurants using either the Review Spotlight interface or original review webpages. Participants were then asked to decide which restaurant they would like to visit and provide the reasons for their decision verbally.

We found that participants could form detailed impressions about restaurants and decide between two options significantly faster with Review Spotlight than with traditional review webpages. Here, we summarize our findings closely related to word pair presentations from this laboratory study. The details of our experimental design and results are available in [Yatani *et al.,* 2011].

Word Pairs Participants Looked at

We did not provide the overall rating in Review Spotlight because we found that it does not always match word pairs that appeared in the interface. But participants successfully gained an idea of how positive or negative the reviews about the restaurant were. Our user study revealed that participants generally focused on word pairs containing particular subjective adjectives (*e.g.*, particular adjectives such as "good," "great," or "poor"). We also found that although the word pair coloring based on the sentiment analysis was intended for this purpose, most of the participants rather looked at the words, their font sizes, and their exact numbers of occurrences to obtain their general impression.

Handling Linguistic Analysis Problems

Participants noticed that some word pairs did not make much sense and that the color of the fonts did not often match to what they thought (*e.g.,* "impeccable" has a high negativity value in SentiWordNet 1.0). This type of problems could be addressed by incorporating a more sophisticated method of determining the relevant information and sentiment in the review [Jo and Oh, 2011]. Another linguistic analysis problem is that the meanings of some word pairs were context-dependent. For example, if one reviewer commented "Last time we went, we had and loved the grilled chicken," and another commented "I will avoid their grilled chicken next time," the current Review Spotlight implementation would detect that "grilled chicken" is a common pair despite the contrasting reactions. Similarly, the current Review Spotlight implementation does not accurately extract the word pairs that appeared in negative sentences (*e.g.,* "This is not a good restaurant").

The Review Spotlight interface allows the user to click on an adjective-noun pair to see the original review text containing that word pair so that she can learn additional information about the reviewed entity. But our user study showed that participants also used this feature to assure the true meanings of the word pairs.

Controlling Displayed Word Pairs

One participant suggested enabling control of the degree of subjectivity or objectivity in the word pairs. In some cases, users would look for objective (descriptive) words rather than subjective (evaluative) words to understand what the restaurant is like and what it offers. Yet, subjective word pairs were frequently used for impression formation in our user study, and thus allowing the user to adjust the amount of subjective/objective words appeared in the interface could improve the system. Based on this finding, we included a simple filter (checkboxes in Figure 2b) to include/exclude positive, negative, and objective word pairs from the interface.

4.2 Word Pair Evaluation

We also examined what word pairs people find useful in an overview of restaurant reviews and how different word pair scoring methods can impact the identification of word pairs which users would appreciate. We compare two methods for scoring the word pairs extracted from the review text: term frequency (TF) and term frequency-inverse document frequency (TF-IDF) [Salton and McGill, 1986]. TF is a commonly used method for a tag cloud visualization, and it is the scoring method which the current Review Spotlight uses. We hypothesized that the word pairs extracted by TF-IDF could convey more useful information to the user because TF-IDF can emphasize unique word pairs in the reviews of one entity. We also initially included entropy as a potential scoring method in our evaluation [Lu *et al.,* 2009]; however, we found that entropy over-weighted word pairs used very rarely, and as a result the highly-scored word pairs were not generally relevant (*e.g.*, "other place" or "23rd birthday"). Thus, we did not further examine the entropy method.

For three different restaurant categories (Italian, Japanese, and Mexican), we obtained the review text for ten restaurants located in the same city in United States with the most reviews on Yelp as of February 2011. The chosen restaurants had 673 reviews on average (SD=270), and were rated between 3 and 4.5 stars. After being processed, the restaurants had 8717 word pairs (SD=4248) on average. Figure 3 shows the distribution of word pairs across their number of occurrences; we note a similar pattern to what we have seen in other tag datasets [Venetis *et al.*, 2011]. We used the restaurant reviews from the same category to calculate the IDF. We then extracted 40

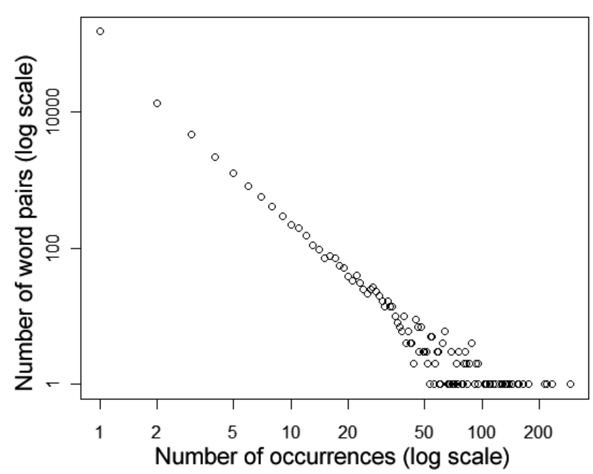

Figure 3.The distribution of word pairs across their number of occurrences.

Category	TF		TF-IDF	
	R	I	R	I
Italian	3.61 (0.59)	3.76 (0.59)	3.50 (0.54)	3.58 (0.54)
Japanese	3.83 (0.60)	3.69 (0.61)	3.74 (0.57)	3.69 (0.56)
Mexican	3.71 (0.72)	3.69 (0.71)	3.52 (0.65)	3.45 (0.68)

Table 1. The average and standard deviation of the relevance (R) and informative (I) ratings for 400 word pairs in each category.

words with the highest scores for each method.

We placed the extracted word pairs and the URL of the original review webpage on the Amazon Mechanical Turk (AMT) to obtain feedback about how relevant and informative each word pair is. We designed each task page on AMT (called HIT) to contain ten word pairs extracted from the review text of one restaurant. Five of the word pairs were selected from the top 40 TF word pairs, and the others from the TF-IDF word pairs. We asked the workers on AMT (referred as "raters" in this paper) first to read the original Yelp reviews and develop an impression for the restaurant. They were then asked to rate each word pair based on two criteria: how relevant and how informative the word pair was (relevance and informativeness). Each rating was gathered using a 5-Likert scale (5 for highly relevant or informative, and 1 for highly irrelevant or uninformative).

Each word pair was rated by three different raters; in total, there were 57 unique raters who rated 130 word pairs on average. We used the average of the ratings for the analysis described later. The raters were paid $0.05USD for each completed HIT.

4.3 Results of Word Pair Evaluation

We found that the relevance and informativeness ratings were highly correlated (Pearson's r=.75, p<.001). This corresponds with our intuition: an informative word pair is likely to be relevant to the reviewed entity. Table 1 shows the average and standard deviation of the relevance and informativesness ratings of 400 word pairs from the ten restaurants in each of the three categories. Raters generally rated the extracted word pairs as relevant and informative to some degree. In all categories, the word pairs generated by the TF method were rated as slightly more relevant and informative than the word pairs by the TF-IDF method.

We further analyzed what word pairs were highly rated and which method would be able to extract these word pairs. For example, Table 2 shows the fifteen word pairs with the highest relevance and informativeness ratings for one Japanese restaurant. We performed a similar inspection for all restaurants and discovered two trends on the difference of word pairs between the TF and TF-IDF methods.

We found that the TF-IDF method often did not highly score common positive word pairs which raters considered relevant and informative, such as "good food" and "good place" in Table 2. This result is in line with the findings from our interface evaluation that participants typically used par-

Word pair	TF Rank	TF-IDF Rank	R	I
fresh fish	1	2	5	5
great sushi	23	19	5	5
delicious sushi	(n/a)	40	5	5
good food	24	(n/a)	5	4.67
good place	11	(n/a)	5	4.33
delicious sashimi	(n/a)	21	4.67	5
good quality	38	(n/a)	4.67	5
good sushi	7	9	4.67	4.67
sushi place	6	14	4.67	4.67
sashimi platter	(n/a)	22	4.67	4.33
delicious fish	(n/a)	39	4.67	4.33
good service	29	(n/a)	4.67	4
fresh quality	(n/a)	33	4.33	5
grilled salmon	(n/a)	36	4.33	4.67
Japanese restaurant	18	(n/a)	4.33	4.67

Table 2. The fifteen word pairs with highest relevance (R) and informative ratings (I) for one Japanese restaurant.

ticular adjectives to build their impressions. We found that the TF method generally captured these word pairs, and contributed to slightly higher overall ratings of relevance and informativeness.

The word pairs scored by the TF-IDF method were often descriptive about specific food or dishes. For instance, "delicious sashimi" and "sashimi platter" in Table 2 were word pairs only extracted by the TF-IDF method.

The results imply that improvements could be achieved by combining multiple word pair scoring methods. However, it would not be as simple as just combining the results of multiple scoring methods (*e.g.,* 20 top words by the TF method and 20 top words by the TF-IDF method) as Table 2 illustrates. Future work is necessary to develop a more effective scoring method than what the TF or TF-IDF method individually offers in order to emphasize useful word pairs and to understand what word pairs users would appreciate in other types of reviews (*e.g.*, hotels and electronics).

5 Conclusions and Future Work

The overwhelming number of user-generated reviews and their inconsistent writing style often require a significant amount of time and effort to read, and can result in important information being obscured from the user. To counter such challenges, we developed Review Spotlight, a user interface that helps users with impression formation by summarizing user-generated reviews using adjective-noun word pairs. Our interface provides a quick overview of the reviews and allows the user to explore the details of reviews in which word pairs are mentioned.

A more effective scoring method for adjective-noun word pairs would improve the usability of Review Spotlight. As we presented, the TF and TF-IDF scoring methods both seem to perform moderately well. However, they still tend to contain word pairs which users would find rather irrelevant or uninformative. Venetis *et al.* [2011] examined four scoring methods to extract informative tags with CourseRank and del.icio.us datasets. However, their focus was on a tag cloud for displaying search results, and issues remain on how to apply their methods into summarization of online user-generated reviews.

One concern with using user-generated reviews is their reliability. The current Review Spotlight did not do any filtering of word pairs or review text. Thus, it could provide false impressions if the original reviews were posted under malicious motivations (*e.g.,* a reviewer posted false information or negative comments to decrease the venue's popularity). We can address this issue by adding more weight to reviews from highly-rated reviewers (*e.g.,* top 100 reviewers) or reviews other users found useful.

The current interface only supports reviews written in English. Adopting an international POS tagger like TreeTagger [Schmid, 1994] would enable the system to extract adjective-noun pairs from the text and display the review written in a different language in the same way it currently does with English text. However, adjective-noun pairs might not be the best way to summarize review pages in other languages. Thus, an investigation on what users would find to be more appropriate in other languages is necessary for internationalization.

References

[Carenini *et al.*, 2006] Giuseppe Carenini, Raymond T. Ng, and Adam Pauls. Interactive multimedia summaries of evaluative text. In *Proceedings of the International Conference on Intelligent User Interfaces*, pages 124–131, 2006.

[Carenini *et al.*, 2009] Giuseppe Carenini and Lucas Rizoli. A multimedia interface for facilitating comparisons of opinions. In *Proceedings of the International Conference on Intelligent User Interfaces*, pages 325–334, 2009.

[Dave *et al.*, 2003] Kushal Dave, Steve Lawrence, and David M. Pennock. Mining the peanut gallery: opinion extraction and semantic classification of product reviews. In *Proceedings of the International Conference on World Wide Web*, pages 519–528, 2003.

[Esuli and Fabrizio, 2006] Andrea Esuli and Fabrizio Sebastiani. SENTIWORDNET: a publicly available lexical resource for opinion mining. In *Proceeding of the International Conference on Language Resources and Evaluation*, pages 417–422, 2006.

[Ganesan *et al.*, 2008] Kavita A. Ganesan, Neelakantan Sundaresan, and Harshal Deo. Mining tag clouds and emoticons behind community feedback. In *Proceeding of the International Conference on World Wide Web*, pages 1181–1182.

[Hu and Liu, 2004] Minqing Hu and Bing Liu. Mining and summarizing customer reviews. In *Proceedings of the International Conference on Knowledge Discovery and Data Mining*, pages 168–177, 2004.

[Jo and Oh, 2011] Yohan Jo and Alice H. Oh. Aspect and sentiment unification model for online reviews analysis. In *Proceedings of the International Conference on Web Search and Data Mining*, page 815–824, 2011.

[Lee *et al.*, 2007] Sung Eob Lee, Dong Kwan Son, and Steve SangKi Han. Qtag: tagging as a means of rating, opinion-expressing, sharing and visualizing. In *Proceedings of the International Conference on Design of Communication*, pages 189–195, 2007.

[Liu *et al.*, 2005] Bing Liu, Minqing Hu, and Junsheng Cheng. Opinion observer: analyzing and comparing opinions on the Web. In *Proceedings of the International Conference on World Wide Web*, pages 342–351, 2005.

[Pang *et al.*, 2002] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 79–86, 2002.

[Lu *et al.*, 2009] Yu-Ta Lu, Shoou-I yu, Tsung-Chieh Chang, and Jane Yung-jen Hsu. A content-based method to enhance tag recommendation. In *Proceedings of International Joint Conference on Artificial Intelligence*, page 2064–2069, 2009.

[Rivadeneira *et al.*, 2007] A. W. Rivadeneira, Daniel M. Gruen, Michael J. Muller, and David R. Millen. Getting our head in the clouds: toward evaluation studies of tagclouds. In *Proceedings of the International Conference on Human factors in Computing Systems*, pages 995–998, 2007.

[Salton and McGill, 1986] Gerard Salton and Michael J. McGill. *Introduction to Modern Information Retrieval*. McGraw-Hill Inc., New York, NY, USA, 1986.

[Schmid, 1994] Probabilistic part-of-speech tagging using decision trees. In *Proceedings of the International Conference in New Methods for Language Processing*, pages 44–49, 1994.

[Tsuruoka and Tsujii, 2005] Yoshimasa Tsuruoka and Jun'ichi Tsujii. Bidirectional inference with the easiest-first strategy for tagging sequence data. In *Proceeding of the Conference on Human Language Technology and Empirical Methods in Natural Language*, pages 467–474, 2005.

[Turney, 2002] Peter D. Turney. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pages 417–424, 2002.

[Venetis *et al.*, 2011] On the selection of tags for tag clouds. In Proceedings of *Proceedings of the International Conference on Web Search and Data Mining*, page 835–844, 2011.

[Yatani *et al.*, 2011] Koji Yatani, Michael Novati, Andrew Trusty, and Khai N. Truong. Review Spotlight: A User Interface for Summarizing User-generated Reviews Using Adjective-Noun Word Pairs. *To appear in Proceedings of the International Conference on Human Factors in Computing Systems*, 2011.