"Beautiful picture of an ugly place". Exploring photo collections using opinion and sentiment analysis of user comments

Slava Kisilevich*, Christian Rohrdantz†, Daniel Keim‡
Department of Computer and Information Science
University of Konstanz, Germany
{slaks*,rohrdantz†, keim‡}@dbvis.inf.uni-konstanz.de

Abstract—**User generated content in the form of customer reviews, feedbacks and comments plays an important role in all types of Internet services and activities like news, shopping, forums and blogs. Therefore, the analysis of user opinions is potentially beneficial for the understanding of user attitudes or the improvement of various Internet services. In this paper, we propose a practical unsupervised approach to improve user experience when exploring photo collections by using opinions and sentiments expressed in user comments on the uploaded photos. While most existing techniques concentrate on binary (negative or positive) opinion orientation, we use a real-valued scale for modeling opinion and sentiment strengths. We extract two types of sentiments: opinions that relate to the photo quality and general sentiments targeted towards objects depicted on the photo. Our approach combines linguistic features for part of speech tagging, traditional statistical methods for modeling word importance in the photo comment corpora (in a real-valued scale), and a predefined sentiment lexicon for detecting negative and positive opinion orientation. In addition, a semi-automatic photo feature detection method is applied and a set of syntactic patterns is introduced to resolve opinion references. We implemented a prototype system that incorporates the proposed approach and evaluates it on several regions in the World using real data extracted from Flickr.**

I. INTRODUCTION

With the fast development of user-centered Internet technologies, we witness a rapid growth of Web resources, which not only allow users to consume textual information, but also to generate their own. This leads to dramatic improvements of products and services. For example, nowadays it is difficult to imagine that we would book a hotel room without checking the hotels overall ranking or without reading comments previously written by other users. We are also less inclined to buy a product without reading comments or ratings about its quality. In fact, written opinions have become essential components in decision-making processes. Furthermore, opinionated texts are now common in almost all parts of our life. They are essential parts of blogs, news, financial market reports, product reviews, etc. However, textual information generated on the Web almost grow at an uncontrollable pace, and manual skimming through user opinions has become a time consuming process.

There has been extensive research within the past ten years on automatic opinion and sentiment analysis. Different algorithms and approaches have been proposed for the analysis of customer feedback data from web surveys [1], movie reviews [2], [3], [4], news articles [5], product reviews [6], [7], financial blogs and news [8], [9], stock message boards [10], opinions in the domain of fast food restaurants [11], and blogs [12].

A typical task in opinion mining is to determine whether a document (review, comment) is bearing a positive or negative connotation [13], [2], [6], [3], [10], [11], [14]. If either connotation is present, the task can be formulated as a classification problem with two class labels (positive and negative) [15]. Three different kinds of approaches have been used: Unsupervised [2], semi-supervised [14] and supervised [1], [3], [12], [10], [16], [9] ones. Supervised machine learning approaches perform good if sufficient labeled training data exist (for example, in the movie reviews domain users assign ranks to movies along with their opinionated text). However, in domains where labels are not easy to acquire or where opinion orientation is measured on a real-valued scale [17], unsupervised approaches are more favorable.

In this paper, we consider the problem of opinion and sentiment analysis of users' comments written for photos, uploaded to photo sharing web sites. Photo sharing web sites, in general, allow users to maintain their own albums of photos. Users can view photos of other members and write comments for a particular photo. In this paper we work with photo comments from Flickr[1].

Before proceeding further with the analysis, we need to understand what are the similarities and what are the differences between the domain of user photo comments and other domains. Having manually examined hundreds of user comments, we found some similarity to blogs [12], where opinions are stated in the beginning of the paragraph. Similar to blogs, the same user can write several comments about the same photo, but usually the first comment contains the opinions and sentiments, while subsequent comments mostly include neutral information like responses to comments of others or the photo owner. The following example shows two comments from the same user. In the first comment there is an expression of sentiment like "Powerful place and story", while the second comment was made after the owner of the photo wrote his response.

---

[1]http://www.flickr.com/

(1) *this is great. I visited Dachau, but don't remember this part. but I hear they have added some things in the last 5 years. Powerful place and story, thanks for sharing*

(2) *I was there about 8 years ago and I don t recall this hall way. Was this one of the houses, or near the main complex where the museum and films were?*

As already mentioned, the owner of the photo can also participate in the discussion about his own photo. In this case, his opinions can introduce a certain bias, which suggests that comments of the photo owner should be excluded from the analysis. The following is a short example of two comments written by the owner of the photo to people as a response to their comments.

(1) *thanks for the comments. i also found the colors both beautiful and chilling...a very creepy place for sure*

(2) *Thanks! I was fortunate to actually capture the impression it made on me standing there in person*

Detailed inspection of user comments revealed that comments are noisy, relatively short, and with only few negations. They may be written in any languages, contain arbitrary syntactic structures and typos. Moreover, they may contain a mixture of opinions on the quality of the photo (usually positive) "Great shot", "Nice picture" and sentiments or moods expressed towards objects depicted on the photo ("Sad place"). As mentioned above, a widely used approach is to classify documents using a binary classification. This approach seems inappropriate in our case for two reasons: (1) Photo comments have two subjects of opinions (opinions on the photo and sentiment towards objects). Consequently, we will loose valuable information if the overall score will be a mixture of two opinion scores. (2) Since most of the opinions are positive, we will end up with most of the photo comments classified as positive. In order to draw a clear analysis, we propose two improvements over existing approaches. We extract two types of opinions: (1) opinions that relate to the photo quality, and (2) general sentiments targeted towards objects depicted on the photo. Supervised machine learning approaches are not feasible in our case since it is very hard to find agreements between human annotators on a real-valued scale, e.g. the difference in opinion strength between "Great shot" and "Amazing photo" cannot be clearly defined. For that reason, we propose an unsupervised approach for opinion scoring using concepts of word importance based on statistical properties derived from the field of information retrieval [18]. Further observations revealed that opinionated pieces of text are mostly accompanied by adjectives, which is in accordance with past findings [19], [20]. Based on these facts, we generated our own lexicon of adjectives extracted from the corpora of user comments, and analyzed its usage with respect to photo quality opinions and general sentiments, as well as their usage by commenters. We found that in the majority of cases adjectives are used directly with the subject of the opinion ("Great shot") and that the most frequently used adjectives are the same, even if different regions of the world are considered with photos of different topicality.

The latter suggested that a finite lexicon of adjectives can be used for opinion analysis of photo comments in many regions around the world. In addition, we also discovered an interesting property of the frequency of adjectives, which is perfectly described by Zipf's law [21]. Our approach is based on a sentence-level opinion analysis, which makes it scalable in dynamic environments like photo comments, where a comment can be added at any time by any of the members.

We developed a desktop Google Earth-based system [22] that combines map navigation and photo exploration using *opinion* and *sentiment* scores as well as a number of derived textual features. The system is capable of showing the photos filtered by one of the features, locating them on the map or seeing them sorted sequentially in an additional window. We believe that our approach can be a very useful extension of photo sharing web sites that will enrich and improve the currently available service capabilities.

The main contributions of the paper can be summarized as follows:

- Our model is based on the corpora extracted from users' photo comments.
- We construct and work with a finite lexicon of opinion words in contrast to the majority of approaches in which seed lists are used to infer scores of unknown opinion words.
- We develop a model that consists of two types of scores: *opinion* regarding the photo and *sentiment* towards the subject of the photo. For this purpose, we suggest a semi-automatic extraction of photo features and a set of syntactic opinion reference patterns.
- We model the orientation strength based on word distributions without using any external dictionaries, while the semantic orientation (positive or negative) of a word is determined by the predefined lexicon of positive and negative opinion-bearing words.
- We provide a continuous scale for opinion and sentiment orientation.
- Our approach allows dynamic updates of scores when new comments are added to the system, which makes the whole method readily applicable in real-world tasks.
- We demonstrate our approach using a Google Earth-based framework.

II. RELATED WORK

Existing approaches in the context of opinion analysis can be broadly divided into several categories. We will review the following categories as they are closely related to our work *opinion classification and orientation*, *lexicon generation* and *feature-based opinion analysis*. A more detailed overview can be found in [15].

A. Opinion classification and orientation

An unsupervised approach for review classification was applied in [2] using pointwise mutual information (PMI) between a phrase containing an adjective or adverb, and the positive word "excellent" and negative word "poor". The PMI

probabilities were calculated based on the number of pages, retrieved by the AltaVista search engine, that contain one of the phrases or two phrases together. The review was classified as positive or negative using the average opinion orientation of all phrases.

[23] proposed a similarity measures of adjectives for semantic orientation using the WordNet [24] synonymy relation. The idea is to count the geodesic distance (similarity relation) of an arbitrary word to a word *good* and *bad* and to determine its orientation based on its similarity relation to one of these words.

A Naive Bayes Classifier was used in [3] for classifying movie reviews, while [10] use Naive Bayes as one of five classifiers with majority voting. A Support Vector Machine (SVM) classifier was used by [1] for classifying customer feedback data. [9] applied SVM on financial blogs.

[17] proposed a real-valued scale opinion orientation based on a classification of adverbs (doubt, strong and week intensifiers, negation and minimizers), different verb categories (positive, negative, conjecture and declarative verbs) and complex relationships of adverbs, adjectives and verbs in the text.

B. Lexicon generation

[13] proposed an approach for the identification of semantic orientations of words using a seed list of positive and negative orientations and the conjunctions (and, or, but) between adjectives with known orientation. [5] used a seed lexicon containing eight positive and eight negative words from the news domain to classify a news article as positive or negative. A vector of words from an article was constructed and similarity between it and the vector of negative and positive seed words was measured using cosine measure. A novel approach was proposed in [7] to construct a domain sentiment lexicon using a seed list of sentiment words and relations of these words to specific topics (in the product domain). The key observation is that sentiment words are directly associated with product features. This observation was used to identify new sentiment words using features and new features using new sentiment words. [12] used a Wikipedia dictionary to determine the polarity of adjectives. [25] generated a dictionary called SentiWordNet using WordNet with three sentiment scores (positive, negative and objective) for each WordNet synset.

C. Feature-based opinion analysis

In addition to the approaches that try to detect the sentiment of sentences or even documents as a whole, the task of feature-based analysis is to investigate to which feature (e.g. entity, topic, attribute) a sentiment or opinions refers. This is also very important in our case since we want to separate opinions that refer to photo features from other opinions. Having identified a set of features and a set of opinion words with respective orientation values (+1/-1) in a sentence, the task is to assign the opinion words to features. Different approaches have been suggested in the past. Some of them use distance-based heuristics ([26], [27]). The closer an opinion word is to a feature word, the higher is its influence on the feature. Other approaches exploit advanced natural language processing methods, like dependency parsers, to resolve linguistic references from opinion words to features.

[28] extract pairs (opinion word, feature) based on 10 extraction rules that work on dependency relations involving subjects, predicates and objects gathered from the Minipar dependency parser[2]. [29] use lexico-syntactic patterns in a bootstrapping approach for subjectivity classification. They define a set of 13 syntactic templates (e.g. subject passiveverb) and concrete example patterns for such templates (e.g. subject was satisfied). However, the purpose is only to resolve relations between opinion holders and verbs for subjectivity classification.

III. Photo Comment Corpus

In this section we outline the photo comment collection, the creation of the corpus and the preprocessing techniques.

A. Data Collection

We collected photo comments from Flickr, the largest web community for photo and web sharing, using its publicly available API[3]. Since the API does not allow downloading metadata for a particular region in the World, the downloading was performed as follows: An initial user id was used to download his photo metadata (owner id, photo id, photo url, date a photo was taken, geotagged information, comments). Then, we downloaded all the users' contacts. To speed up the process of retrieving heterogeneous users, we retrieved all groups to which the individual user belongs, and using group information, we were able to retrieve all the people who belong to these groups. Beginning in June, 2009 (as part of another project) we collected metadata for about 90 million geotagged photos from about 7.6 million users by the end of April, 2010.

B. Development of Corpora

1) Region selection: Five regions (Dachau, Auschwitz, Wisla, Krakow and Warsaw) were defined for analysis. The rationale behind selecting these regions is that we want:

- To find differences in comment types between regions
- To find differences in the usage of parts of speech (adjectives and nouns)
- To build a model that better reflects different kinds of comments

We assumed that Dachau and Auschwitz concentration camps should contain special kinds of comments (negative emotions) that would differ from comments in general touristic places. Wisla, we assumed, is a neutral region visited rarely by tourists while Krakow and Warsaw were selected as large Polish cities that include many touristic attractions.

Table I summarizes the statistics related to the selected regions.

[2]http://webdocs.cs.ualberta.ca/~lindek/minipar.htm
[3]http://flickrnet.codeplex.com/

TABLE I
STATISTICAL INFORMATION RELATED TO FIVE REGIONS SELECTED FOR ANALYSIS

Region	Area	# commented photos	# owners	# commenters
Krakow	$120km^2$	8127	1257	23045
Warsaw	$60km^2$	8690	1140	22695
Wisla	$43km^2$	117	39	603
Auschwitz	$12km^2$	505	138	1687
Dachau	$14km^2$	329	121	1062

2) Preprocessing: For every region, we selected photos that contain at least one comment. Photos that do not contain comments were removed from further analyses.

Photo comments are very noisy and unstructured. They may contain HTML tags which should be filtered from the original text. In addition, they may contain different irrelevant sections that have to be removed such as URL links or invitations to join a group. Below are two examples of comments (punctuation is preserved) that require the removal of URL links and invitations to join groups

(1) *Greetings!Using the "blog this" function above your picture, we have linked your picture to our WordPress blog SIDDHARTHA*

(2) *Hi, I m a member of a group called Five Flickr Favs, and we'd love to have your photo added to the group*

Photo comments can be written in different languages or may contain mixtures of several languages. In order to analyze parts of speech usage, we had to apply a POS-tagger. Since there is no universal POS-tagger that can work on any language and we don't know exactly what languages are used in comments, we decided to remove all comments that are not written in English. For this, we used the TextCat language guesser[4]. The following languages were identified while scanning all the comments: *Polish, English, Swedish, Slovenian, Slovakian, Danish, Italian, Dutch, Spanish, French, German, Finnish, Albanian, Hungarian, Norwegian, Unknown.*

After removing all non-English comments, we were left with 4214 commented photos in Krakow, 4098 commented photos in Warsaw, 56 commented photos in Wisla, 311 commented photos in Auschwitz and 179 commented photos in Dachau.

In the next step, we applied the Stanford POS Tagger[5] on the English comments.

IV. METHOD

A. Definitions

Different terminology definitions are provided in the sentiment and opinion analysis literature. The terminology used in this paper mostly sticks to the definitions given in [15], but makes a clear distinction between opinions and sentiments. The important terms and their definition for this paper:

[4]http://odur.let.rug.nl/~vannoord/TextCat/
[5]http://nlp.stanford.edu/software/tagger.shtml

- **Photo Feature:** Nouns that describe the photo features - attributes, components or characteristics of the photo, e.g. "shot", "photo", "color", "composition", "light". Photo features in our case are usually directly related to the quality of the photo. It is common to distinguish between explicit and implicit features, i.e. features that are mentioned in a sentence and features that are not explicitly mentioned but implicitly referenced.
- **Orientation:** The semantic orientation of a word or a comment as a binary categorical variable with the parameter values "negative" and "positive" (the third possibility "neutral" is omitted in our scenario).
- **Orientation Strength:** The numerical strength of the orientation value ranging from 0 to ∞ in absolute numbers, whereas negative orientations are indicated by the algebraic sign "-".
- **Photo Opinion (PO):** Negative or positive user statements, that clearly refer to photo features of a certain photo, are summarized as the respective photo opinion. They express the users' opinions on the technical and artistic photo quality. For simplicity, we will only speak of opinions when we refer to photo opinions.
- **General Sentiment (GS):** Negatively or positively connoted user statements that cannot be attributed to a photo feature. As implied by the denomination, the general sentiment shall capture orientation statements that have a broader nature than opinions, i.e. sentiments and emotions that are evoked by the photo content. For simplicity, we will only speak of sentiments when we refer to general sentiments.
- **Ambiguity:** Not all users have concordant opinions or sentiments when commenting on a photo. However, at the end one single PO and GS value is computed for each photo that does not account for potential disagreements among users. Accordingly, for both PO and GS ambiguity values are provided that indicate perfect agreement (0.0) or complete disagreement (1.0) among users, as well as arbitrary real values in between.

B. Corpus-based lexicon generation

Opinion mining is heavily dependent on an opinion lexicon. The two common approaches for lexicon generation are dictionary-based and corpus-based ones. The former is based on bootstrapping a seed of opinion words from dictionaries like WordNet, SentiWordNet or Wikipedia, the latter is based on the corpus and, thus, inherently domain dependent.

We applied a corpus-based lexicon generation due to several reasons:

- We want to generate a new lexicon in the domain of photo comments since currently, at least to our knowledge, no such lexicon is publicly available
- Dictionaries like SentiWordNet may supply only a binary opinion orientation, while our task is to model opinion orientations on a real-value scale
- We want to investigate statistical properties of words used for commenting

It was shown in past research that there is a strong correlation between the presence of adjectives and opinions [19], [20]. Indeed, a careful analysis of photo comments showed that people often use short sentences like "Great photo", "Nice picture", "Sad place" to express their opinions or sentiments. The analysis also showed that the number of positive adjectives used in photo comments is higher than for negative ones and, overall, the number of positive comments is much higher than the number of negative comments. Any lexicon of positive and negative words will show that the words "great" and "nice" are positive. However, it is difficult to estimate which of these two words is "more positive than the other" using lexical features alone. For that reason, we decided to apply a measure, which is similar to the TF-IDF (Term Frequency-Inverse Document Frequency) measure used in information retrieval and text mining [18]. The idea is that standard opinion or sentiment words that are used frequently by majority of people receive lower scores than words that are used infrequently. In order to acquire word distributions, we extracted adjectives and nouns from the corpus, counted their occurrences in the five selected regions separately, and sorted them according to their occurrence from the highest to the lowest. Nouns were extracted in order to learn what words are commonly used as photo features. We used Yago-Naga stemmer[6] to convert all nouns into a singular form.

In order to minimize the bias of some active commenters, we counted word occurrence only once for each person. The reason why we selected five separate regions is that word occurrences may differ due to different topicalities. Moreover, the number of commented photos is different from region to region and the word distribution would inevitably be biased towards words used in regions with a lot of comments. Table II summarizes frequencies of adjectives in five areas.

An inspection of the results shows the following interesting patterns: The words *great, nice* and *beautiful* are the most frequent and equally ranked adjectives in all five regions, 33% of the adjectives are found within the 20 most frequent adjectives in every region, 58% of the adjectives are found in at least one region and 42% of frequent adjectives are found only in one region. This suggests that people use many common opinion words even if the context of photos is very different (Dachau concentration camp and Nature).

Another interesting finding is that the distribution of adjectives in all five regions can be described by Zipf's law, which stays that if f is the frequency of a word in the corpus and r is the rank, then

$$f = \frac{k}{r} \qquad (1)$$

where k is a constant for the corpus. When we take the logarithm of both sides, we obtain a linear function with the slope of -1. The slope coefficients we obtained for Krakow is -1.138, Warsaw: -1.136 , Auschwitz: -0.988 and Dachau: -0.95 (Wisla was excluded because it does not have enough

[6]http://www.mpi-inf.mpg.de/yago-naga/

TABLE II
20 MOST FREQUENT ADJECTIVES AND THEIR FREQUENCY IN FIVE SELECTED AREAS. WORDS THAT ARE COMMONLY USED IN FIVE REGIONS ARE COLORED IN YELLOW, IN FOUR REGIONS - GRAY, IN THREE - PINK, IN TWO - GREEN, IN ONE - WHITE

Krakow	Warsaw	Wisla	Auschwitz	Dachau
great,1469	great,1403	great,26	great,129	great,65
nice,864	nice,856	nice,14	nice,61	nice,29
beautiful,829	beautiful,756	beautiful,13	beautiful,57	beautiful,29
good,311	good,306	lovely,8	good,42	fantastic,17
wonderful,271	wonderful,257	awesome,7	powerful,31	powerful,14
lovely,238	amazing,215	amazing,6	amazing,30	excellent,14
amazing,202	cool,191	cute,6	impressive,27	awesome,11
interesting,200	lovely,184	good,5	sad,24	amazing,11
cool,196	fantastic,181	such,3	wonderful,24	sad,10
fantastic,168	excellent,174	excellent,3	excellent,22	impressive,10
excellent,153	interesting,173	wonderful,3	fantastic,18	very,8
awesome,137	awesome,166	right,2	awesome,17	interesting,8
very,129	very,133	pretty,2	interesting,16	such,7
perfect,116	perfect,104	cool,2	very,15	wonderful,7
gorgeous,74	gorgeous,79	new,2	strong,13	dark,7
such,71	cute,78	very,2	many,12	lovely,6
cute,68	little,61	fantastic,2	same,11	cool,6
much,62	such,55	terrific,1	white,11	scary,6
little,58	stunning,47	fierce,1	such,11	dramatic,6
black,55	impressive,45	perfect,1	cool,11	good,6

words for a reliable slope estimation). Apart the statistical properties of the word distribution, Zipf's law can be also explained in terms of "least effort" principle: [30] *the tension between the goal of the speaker to minimize production efforts by using only few words very frequently and the goal of the listener to minimize perceptual confusion by having a large vocabulary of distinct words.*

C. The Adjective Weighting Model

Having defined statistical and linguistic interpretations of the distributions of adjectives in the photo comments corpus, we are now ready to propose an adjective weighting model for opinion orientation.

We define the word opinion orientation w_{oo} using the principles of word relevance as defined in the TF-IDF measure and word distribution properties of Zipf's law as follows:

$$w_{oo} = orientation(w) * log(\frac{f_{w,r=1}}{f_w} + 1) \qquad (2)$$

where $orientation(w)$ is a function which assigns 1 if the word w is positive and -1 if it is negative, $f_{w,r=1}$ is the frequency of the word having the rank 1 (Equation 1) and f_w is the frequency of the word w in the whole corpus.

The difference between TF-IDF and our approach is that the importance of the word in TF-IDF is measured for every word independently, while opinion orientation score is calculated relative to the most frequent word in the corpus. Thus, if the most frequent word is "great" with frequency of 1469 (see Table II) and the word ranked second is "nice" with the frequency of 864, "great" will get a score of 0.30 (log (1469/1469 + 1)), while the score of "nice" will be 0.43 (log
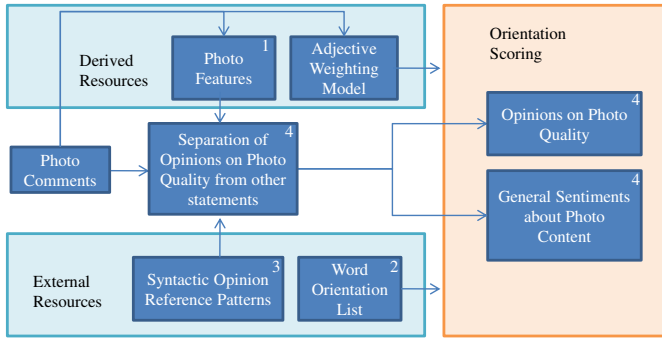
Fig. 1. Overview about the interdependence of the different core text analysis processes. The numbers correspond to the paragraphs in Section IV-D, where details are provided.

(1469/864) + 1). One is added to *log* to avoid zero score of the most frequent word.

We should note, that the word frequency in the Equation 2 is absolute and can be applied to five regions separately. In order to make a global model that takes into account different word distributions, we need to find the relative order of all words from five regions. We proceeded it as follows:

- We calculated a ratio $\frac{f_{w,r=1}}{f_w}$ for every word
- An average of ratios for every word was calculated taking these ratios for the same word $w_{i,n}$ from every region n
- If the word $w_{i,n}$ was not found among the lexicon of the region n, its ratio was assumed to have the ratio of the last word in the lexicon of the region n

After building a weighted ratio for every word, we applied Equation 2 to obtain the global adjective weighting model.

D. Automatic Opinion and Sentiment Analysis

The automatic opinion and sentiment analysis consists of several interdependent steps as outlined in Figure 1. The analysis relies on both resources derived from the photo comment corpus itself and external resources. The details are provided in the following subsections.

1) The Photo Features: In order to determine which opinions relate to the photo, first a list of photo features had to be compiled. For this purpose a term extraction method was created that exploits certain characteristics of photo features: (1) such features usually correspond to nouns, (2) such features should not depend significantly on the photo location, and (3) such features should be frequent in photo comments. Consequently, (1) all nouns where extracted, that (2) appeared in photo comments of at least 4 out of 5 locations, and finally (3) the 100 most frequent among these terms were extracted as candidate photo features. The list was then manually revised and finally, 60 out of these nouns where considered in the analysis as photo features. The top 10 frequent nouns present in at least 4 locations were, in decreasing frequency order, "shot", "photo", "color", "composition", "light", "picture", "capture", "love", "image", "work". Here, "love" is one example that was manually deleted. In this case we could observe that high

frequency of the noun "love" was due to a repeated error of the part-of-speech tagger, when occurrences of the verb "love" in very short sentences (e.g. "Love it!") were misclassified as nouns.

Implicit features: If sentences were shorter than 6 words and did not contain a noun, it was assumed that comments implicitly meant the photo (e.g. "I love it.", "Well done.", "Very nice.").

2) The Word Orientation List: As already mentioned, a manually enhanced version of the widely used Internet General Inquierer lexicon [31] was applied. It was used to determine the orientation of the word and incorporate it into Equation 2, i.e. +1 for positive, -1 for negative and 0 for neutral words (not contained in the orientation list). Before, words were reduced to their base form with Kuhlen's algorithm [32], in order to increase the number of matches.

All words not contained by the adjective weighting model of Section IV-C, because they either did not appear in the photo comments or belong to a different part-of-speech category, were allocated the weight 1.

3) Syntactic Opinion Reference Patterns: In order to detect references of opinion words to photo features, a set of syntactic opinion reference patterns was defined, based on linear word order part-of-speech sequences[7]. A very simple example is the pattern "JJ NN" which stands for an adjective (JJ) directly followed by a noun (NN). In this case we can be sure, that the adjective refers to the noun. Hence, if the noun is a photo feature then the adjective and its orientation can be assigned to this feature. While in theory recursive patterns of arbitrary length (e.g. JJ* NN) are possible in natural language, in practice such patterns do not appear to a noteworthy extent. We could observe that the limited pattern set we defined, covers the vast majority of cases. The whole pattern set is provided in Figure 2. One main advantage is that the patterns encode the available linguistic knowledge about opinion references without requiring the time-consuming parsing of a full syntax structure tree or a typed dependencies graph. Our syntactic reference patterns cover most of the cases that other approaches detect with dependency parses. This is due to the fact that in English adjectives are usually very close to the nouns they refer to. Only very exceptional and infrequent cases like relational phrases "'the photo, that shows a tree, is really nice'" cannot be resolved by our means. In case of verbs, our approach is not able to distinguish explicitly whether the feature is the subject or the object of the verb. In our tests, however, we could observe that this is not a problem. Verbs that express opinions ("'to hate/to like'") cannot have a photo feature as subject and in cases in which they are objects, they are covered by our analysis patterns. In addition, our method is less error-prone than dependency parsing, especially when applied to less formalized and clean writing, as in user-generated content.

---

[7] The used part-of-speech tags follow the Penn Treebank Tag-set definition: http://www.comp.leeds.ac.uk/ccalas/tagsets/upenn.html
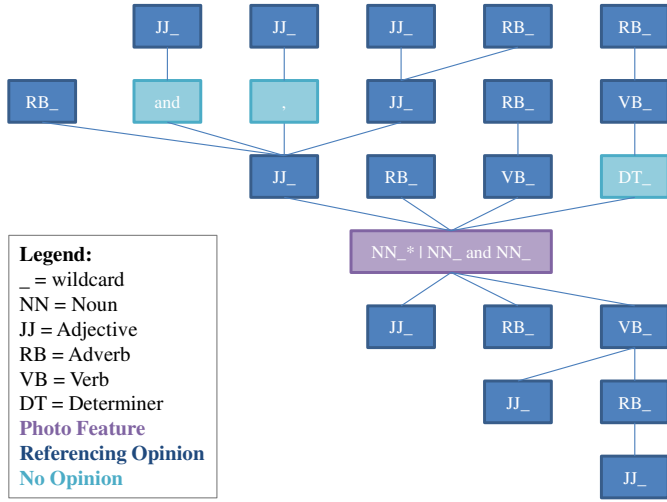
Fig. 2. Syntactic Opinion Reference Patterns. Word order patterns go from top (before photo features) to bottom (after photo features), the level indicates the exact position.

4) Identification and Separation of Photo Opinions and General Sentiments: One crucial point of the automatic text analysis is the detection and separation of (1) opinions about the photo quality (PO) and (2) general sentiments expressed about the photo content (GS).

The first part (1) is based on the extraction of photo features and the mapping of opinion statements to photo features. The described set of syntactic opinion reference patterns was applied for this mapping. For each photo feature in a sentence, all words were extracted that describe the feature according to one of the syntactic opinion reference patterns. The orientation scores of these words were then summed up to yield a photo opinion value. In this process, a simple heuristic is used to invert the orientation of negated words.

Accordingly, step (2) is based on all sentiment expressions that could not be attributed to photo features during step (1). This means that all words not referring to photo features were considered and their orientation scores summed up to yield a general sentiment value.

It should be noted that general sentiments only in very rare cases are falsely classified as photo opinions, whereas the contrary could be observed more frequently, due to different reasons (missing photo feature, implicitness).

Additionally, in both steps the ambiguity of comments is analyzed. This implies investigating whether different users express different opinions or sentiments on the same photo. The output of steps (1) and (2) are an opinion and sentiment value for each user comment. The ambiguity is then calculated separately for the opinions about a photo and the sentiments about a photo. Equation 3 shows how the opinion ambiguity value is calculated for a photo, based on the number of user comments with positive opinions ($\#pos$) and the number of user comments with negative opinions ($\#neg$). For the

sentiments this works analogously.

$$amb = \begin{cases} 0 & \text{if } (\#pos = 0) \vee (\#neg = 0), \\ \frac{Min(\#pos, \#neg)}{Max(\#pos, \#neg)} & \text{else.} \end{cases} \tag{3}$$

E. Statistical Proof of Concept

Because of the lack of an appropriate Gold Standard it is not easy to evaluate the sentiment and opinion analysis. Instead, we try to gain evidence for the suitability by performing statistical analysis. Table III shows mean and standard deviation values of opinions and sentiments for the different regions. As expected, the relative difference of mean photo opinion values between different locations is much smaller then that of mean general sentiments. This is in accord with our expectations, because the general sentiment is much more dependent on the location than the photo quality. There is a certain correlation between photo quality and general sentiments, which could be due to the fact that both cannot be separated unambiguously in all cases. However, the two concentration camp memorials Ausschwitz and Dachau, as expected, have very low general sentiment values and the two popular tourist places Warsaw and Krakow are allocated much higher values (even the same mean). Wisla, which we anticipated to be a rather neutral place, lies between the extrema with its general sentiment values. All in all, the statistics indicate that a reasonable separation of opinions and sentiments could be achieved.

Location	Op. Mean	Op. Stdv.	Sent. Mean	Sent. Stdv
Auschwitz	0.827	1.847	0.318	1.202
Dachau	0.776	2.16	0.268	1.026
Krakow	1.003	2.068	0.736	1.544
Warsaw	0.976	2.618	0.736	2.483
Wisla	0.945	1.209	0.516	1.314

TABLE III
AVERAGE AND STANDARD DEVIATION OF OPINION AND SENTIMENT SCORES

V. APPLICATION

In this section we demonstrate the desktop application that combines Google Earth[8], the custom engine built on top of Google Earth [22], and the navigation and filtering toolbox that implements the method for opinion and sentiment analysis of photo comments.

A. Usage Scenario

Our goal is to enrich the user experience by improving photo navigation in a selected region and adding more options for exploring the area. Google Earth has become a favorite platform among Internet users for map navigation and exploration. Google Earth contains different layers that include points of interest, photos, etc. At any time the user can navigate to a specific region in the World and explore the points of interest or photos that were taken there by tourists.

[8]http://earth.google.com/

The difficulty is that photos are displayed in Google Earth as small rectangular thumbnails. The actual image is displayed only when the user clicks on the thumbnail. To gain an actual view of the place, the user has to click on the thumbnails many (several) times and search through different photos. Similarly to Google Earth, the Flickr web site allows the navigation to a particular place using the provided search field. The web page will display large image thumbnails in a sequential order with an overall statistic of how many images were found. For example, for May 25, 2010, Flickr reports $256,827$ results when Warsaw is used as a keyword. Flickr allows sorting the results using three options *Relevant, Recent* and *Interesting*. In addition, Flickr allows locating photos on the map using its WorldMap[9] by providing the location and optionally the category (architecture, urban, forest). By issuing a search for Warsaw, Flickr found $73,174$ geotagged photos, displayed as image thumbnails on a horizontal strip and sortable according to two parameters: *Interesting* and *Recent*. The *relevant* option allows searching for images that contain the search keyword in their titles, while the *interesting* option is based on the non-disclosed algorithm that takes into account such features like *number of views, comments, etc*.

In our application, we implement two main features that are the core of the algorithmic part of the paper, namely *opinion* and *sentiment*, three derived features, namely *number of sentences in comments, opinion and sentiment ambiguities* and *the number of positive and negative opinion words in comments*. Additionally, we included an additional feature, which is part of the downloaded metadata *the number of times the photo was viewed*. The application has two main views and will be described in the following subsections.

B. Photo sorting

The control panel displayed in Figure 3(a) allows the user to receive information about the boundaries of the selected region (label 1). When the user changes the boundaries by manipulating the Google Earth map, the application connects to the server and updates statistical information related to the photos (label 2). In particular, the following statistics are sent by the server: total photos in the region, minimum and maximum opinion and sentiment scores, minimum and maximum number of sentences in comments, positive and negative opinion words and opinion ambiguity, minimum and maximum number of viewed photos.

The central part of the control panel contains a number of filtering options (label 3): filtering by *opinion*, *sentiment*, *sentences*, *ambiguity*, *number of positive and negative words*, etc. In addition, two *quantity filters* (label 4) allow limiting the number of displayed photos on the map view and in the control panel.

When one of the filtering options is invoked, the request is sent via REST protocol to the server along with all relevant information. The server generates two types of responses that are sent as one string to the client. The first response is

formatted as Keyhole Markup Language (KML), an XML-based language for the visualization of geographic entities and the one which is used by Google Earth. In our case it contains a photo URL and all the relevant information about the photos (opinion and sentiment scores as well as comments). The KML file is extracted from the response and delegated to the underlying Google Earth engine for visualization. The second response is used by the control panel to show N-top photos (label 5) filtered by one of the provided options.

If the user clicks on one of the photos, the information about the selected photo is displayed on the left side including the coordinate of the photo, comments and scores (label 6). A double-click on the photo positions the map around that photo on the map view.

C. Map navigation

Map navigation (Figure 3(b), label 1) allows exploring the photos using the map view after they were filtered by one of the available scores. The exploration is similar to the functionality provided in the stand-alone Google Earth version or Flickr MapView. However, the difference is that the thumbnail of the image is directly visible on the map. This can save time as it does not require clicking on every thumbnail to see the underlying images. When the image thumbnail is clicked, the large image is displayed along with its comments and scores (see Figure 3(b), label 2). Since the data interchange format is XML-based, information about the whole set of filtered photos or an individual photo can be saved into the file and later visualized in any application that supports KML (label 3).

VI. CONCLUSIONS

This paper introduces a practical unsupervised approach for improving the user experience during the exploration of (geotagged) photos on photo sharing web sites by filtering and sorting photos using opinions and sentiments expressed in user comments written for uploaded photos.

Our approach is able to identify two types of opinions from the comments: opinions that are related to the photo quality and general sentiments or moods expressed towards the objects shown on the photo. Unlike most of the existing approaches in which binary (negative or positive) opinion orientation is used, we model opinion orientation using a real-valued scale. Using linguistic features, we build a finite lexicon of adjectives and calculate their opinion strength using a word importance paradigm borrowed from the information retrieval field. The opinion orientation (negative or positive sign) is calculated using a predefined lexicon of positive and negative opinion-bearing words. The identification and separation of photo opinions is based on a semi-automatic method for photo feature extraction and a set of predefined syntactic opinion reference patterns. The overall opinion and sentiment scores for a photo is the cumulative sum of all scores in the comments. This allows a dynamic update of scores if new comments are written for the photo.
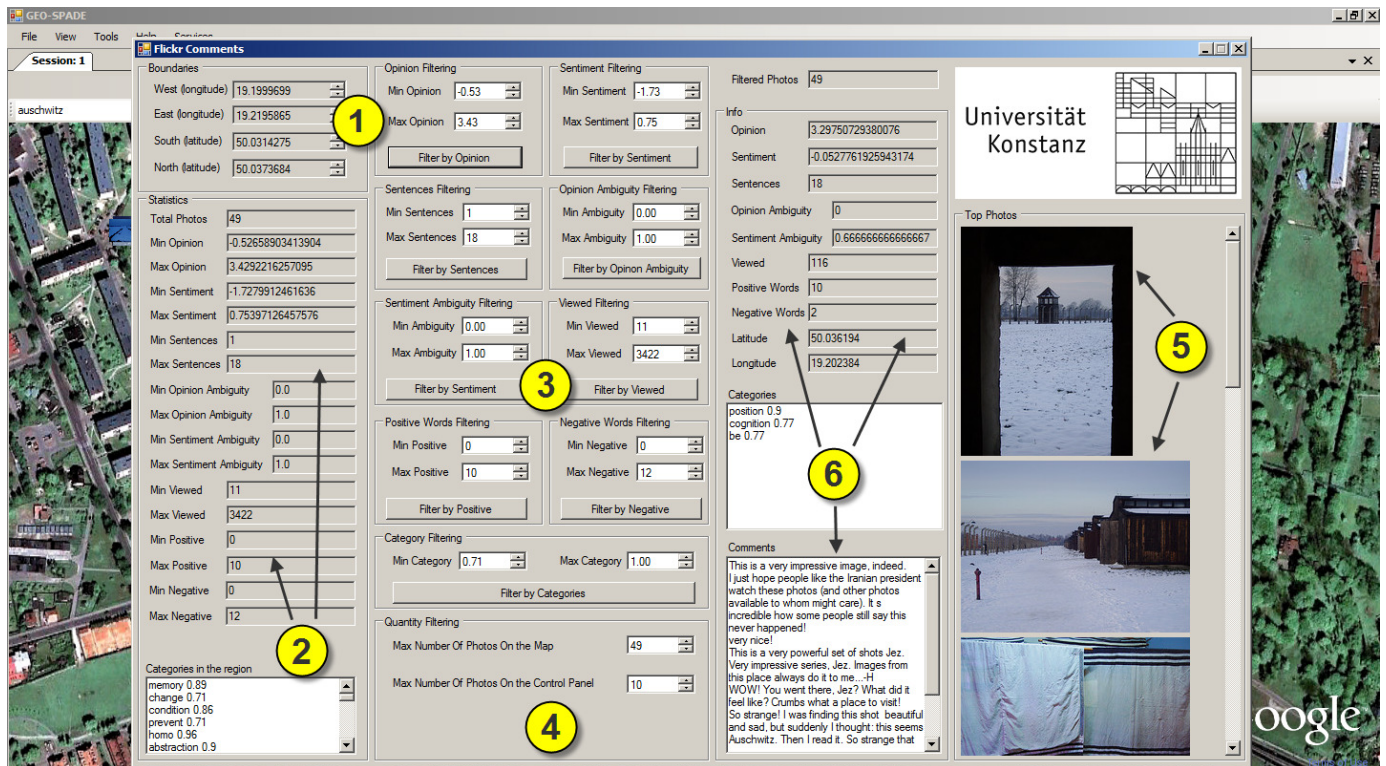
(a) Control view. Filtering according to opinions, sentiments and other derived textual features. N-top ranked photos according to the filtering parameter selected are displayed with the relevant information
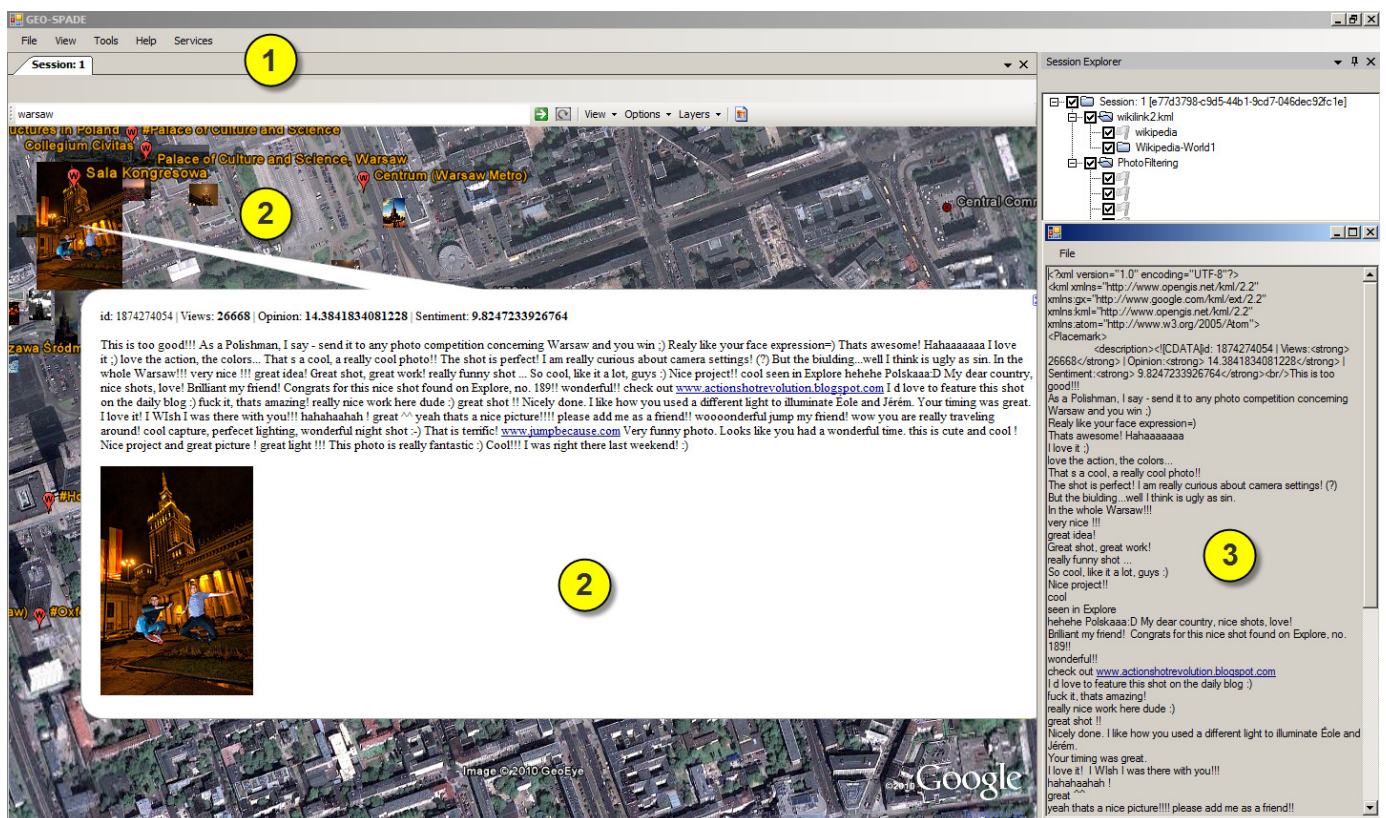
(b) Google Earth view (label 1). Clickable photo thumbnails are displayed on the map (label 2). The results of visualization can be seen and saved in KML format (label 3)

Fig. 3.   Google Earth-based application for photo search using opinion and sentiments scores

We implemented a prototype desktop Google Earth-based system that implements the method described in the paper. It allows the exploration of geotagged photos using opinion and sentiment scores combined with the visualization of photos on the map.

In our future work, we shall concentrate on the improvement of the score assignment algorithm and work on multi-lingual solutions.

ACKNOWLEDGEMENTS

REFERENCES

[1] M. Gamon, "Sentiment classification on customer feedback data: noisy data, large feature vectors, and the role of linguistic analysis," in *Proceedings of the International Conference on Computational Linguistics (COLING)*, 2004.

[2] P. Turney, "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*, 2002, pp. 417–424.

[3] F. Salvetti, S. Lewis, and C. Reichenbach, "Automatic opinion polarity classification of movie reviews," *Colorado research in linguistics*, vol. 17, no. 1, 2004.

[4] B. Ohana and B. Tierney, "Sentiment classification of reviews using SentiWordNet," in *9th. IT & T Conference*, 2009, p. 13.

[5] M. Sahlgren, J. Karlgren, and G. Eriksson, "SICS: Valence annotation based on seeds in word space," in *Proceedings of the 4th International Workshop on Semantic Evaluations*, 2007, pp. 296–299.

[6] K. Dave, S. Lawrence, and D. Pennock, "Mining the peanut gallery: Opinion extraction and semantic classification of product reviews," in *Proceedings of the 12th international conference on World Wide Web*, 2003, p. 528.

[7] G. Qiu, B. Liu, J. Bu, and C. Chen, "Expanding domain sentiment lexicon through double propagation," in *Proceedings of the 21st international jont conference on Artifical intelligence*, 2009, pp. 1199–1204.

[8] A. Devitt and K. Ahmad, "Sentiment polarity identification in financial news: A cohesion-based approach," in *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, vol. 45, no. 1, 2007, p. 984.

[9] N. O'Hare, M. Davy, A. Bermingham, P. Ferguson, P. Sheridan, C. Gurrin, and A. Smeaton, "Topic-dependent sentiment analysis of financial blogs," in *Proceeding of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion*, 2009, pp. 9–16.

[10] S. Das and M. Chen, "Yahoo! for Amazon: Sentiment extraction from small talk on the web," *Management Science*, vol. 53, no. 9, pp. 1375–1388, 2007.

[11] A. Fahrni and M. Klenner, "Old wine or warm beer: Target-specific sentiment analysis of adjectives," in *AISB 2008 Convention Communication, Interaction and Social Intelligence*, vol. 1, 2008, p. 60.

[12] P. Chesley, B. Vincent, L. Xu, and R. Srihari, "Using verbs and adjectives to automatically classify blog sentiment," *Training*, vol. 580, no. 263, p. 233, 2006.

[13] V. Hatzivassiloglou and K. McKeown, "Predicting the semantic orientation of adjectives," in *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics*, 1997, p. 181.

[14] S. Argamon, K. Bloom, A. Esuli, and F. Sebastiani, "Automatically determining attitude type and force for sentiment analysis," *Human Language Technology. Challenges of the Information Society*, pp. 218–231, 2009.

[15] B. Liu, "Sentiment Analysis and Subjectivity," *Handbook of Natural Language Processing, Second Edition,(editors: N. Indurkhya and FJ Damerau)*, 2009.

[16] A. Drake, E. Ringger, and D. Ventura, "Sentiment Regression: Using Real-Valued Scores to Summarize Overall Document Sentiment," in *2008 IEEE International Conference on Semantic Computing*, 2008, pp. 152–157.

[17] V. Subrahmanian and D. Reforgiato, "AVA: Adjective-Verb-Adverb Combinations for Sentiment Analysis," *IEEE Intelligent Systems*, vol. 23, no. 4, pp. 43–50, 2008.

[18] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," *Information processing & management*, vol. 24, no. 5, pp. 513–523, 1988.

[19] J. Wiebe, R. Brace, and T. O'Hara, "Development and use of a gold-standard data set for subjectivity classifications," in *Annual meeting-association for computational linguistics*, vol. 37, 1999, pp. 246–253.

[20] J. Wiebe, "Learning subjective adjectives from corpora," in *Proceedings of the National Conference on Artificial Intelligence*, 2000, pp. 735–741.

[21] R. Baayen, *Word Frequency Distributions*. Springer, 2002.

[22] S. Kisilevich, D. Keim, and L. Rokach, "A generic google earth-based framework for analyzing and exploring spatio-temporal data," in *12th International Conference on Enterprise Information Systems*, 2010.

[23] J. Kamps, M. Marx, R. Mokken, and M. De Rijke, "Using WordNet to measure semantic orientation of adjectives," in *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004*, vol. 4, 2004, pp. 1115–1118.

[24] C. Fellbaum, *WordNet: An electronic lexical database*. MIT press Cambridge, MA, 1998.

[25] A. Esuli and F. Sebastiani, "SentiWordNet: A publicly available lexical resource for opinion mining," in *Proceedings of LREC*, vol. 6, 2006.

[26] X. Ding, B. Liu, and P. Yu, "A holistic lexicon-based approach to opinion mining," in *Proceedings of the international conference on Web search and web data mining*, 2008, pp. 231–240.

[27] D. Oelke, M. Hao, C. Rohrdantz, D. A. Keim, U. Dayal, L.-E. Haug, and H. Janetzko, "Visual opinion analysis of customer feedback data," in *VAST '09: Proceedings of the 2009 IEEE Symposium on Visual Analytics Science and Technology*, 2009, pp. 187–194.

[28] A.-M. Popescu and O. Etzioni, "Extracting product features and opinions from reviews," in *HLT '05: Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, Morristown, NJ, USA, 2005, pp. 339–346.

[29] E. Riloff and J. Wiebe, "Learning extraction patterns for subjective expressions," in *Proceedings of the 2003 conference on Empirical methods in natural language processing*, 2003, pp. 105–112.

[30] M. Baroni, "Distributions in text," *Corpus linguistics: An international handbook*, vol. 2, 2009.

[31] V. Buvac, "Internet general inquirer," http://www.webuse.umd.edu:9090/.

[32] R. Kuhlen, *Experimentelle Morphologie in der Informationswissenschaft*. Verlag Dokumentation, Munchen, 1977.