ISBN 978-83-60810-47-7

Sentiment Analysis for Hotel Reviews

Walter Kasper, Mihaela Vela
DFKI GmbH
Stuhlsatzenhausweg 3
D-66123 Saarbrücken, Germany
Email: {Walter.Kasper,Mihaela.Vela}@dfki.de

***Abstract*—User reviews and comments on hotels on the web are an important information source in travel planning. Therefore, knowing about these comments is important for quality control to the hotel management, too. We present a system that collects such comments from the web and creates classified and structured overviews of such comments and facilitates access to that information.**

I. Introduction

TRAVEL planning and booking on the web has become one of its most important commercial uses. With the rise of the Web-2.0 user-generated reviews, comments and reports about their travel experiences play an increasing role as information source. Especially for hotel booking, such user reviews are relevant since they are more actual and detailed than reviews found in traditional printed hotel guides etc., they are not biased by marketing considerations as e.g. the hotels' home pages or catalog descriptions and reflects actual experiences of guests.

Though nearly every internet travel agency and hotel booking service nowadays offers also ratings and/or reviews of hotels, it is not that easy for hoteliers who want to know what is published about their hotels on the web to gather the user-generated information. A standard search engine like Google will give thousands of hits for a hotel. But, though there seems to be a huge number of sites providing user reviews, often these are just the same because many sites use the same source, such as *openholidayguide.com*. In other cases, the links lead only to some general page from which one can access reviews besides other information and lacking transparent navigation structure. Also, the links might point to some individual review but leaving it open whether there are other reviews on the site. An additional problem is that the Web-2.0 provides a large number of publication types: besides travel agencies and hotel booking services there are numerous blogs, fora, newsgroups, social networks etc. related to traveling.

Another problem concerns the kind of information: travel agencies and hotel booking services often only publish scalar ratings, e.g. scores between 1 and 5. Such scores are not very helpful for hotel managers as the numeric value does not provide information of what guests actually considered positive or objectionable. Also, the numeric scores are not comparable: when a 3-star hotel receives a higher score than a 4-star hotel, that does not imply that the one is better than the other. For hotel managers the textual user comments would be much more significant than the numeric scores since they would be interested to know *what* the users exactly commented on and *how* they thought of it.

Another problem for hotel managers is that of following updates and new reviews. Hotel booking services and travel agencies collect and publish user reviews systematically, e.g. by asking their customers for comments or ratings. So, new reviews appear quite frequently on their pages but it would be difficult to follow these by just using general search.

For the traveling user who is accessing reviews on the web for planning his travel, many of these considerations are not relevant, as he will be content with a momentary snapshot of reviews. But for hoteliers interested in user comments on the web a service that automatically and systematically collects and summarizes the relevant information from the web would be advantageous and perhaps even more useful than the paper forms many hotels use for gathering feedback from their guests.

The BESAHOT service presented in this papers aims at providing such a service for hotel managers that collects user reviews for hotels from various sites on the web, analyzes and classifies the textual content of the review and presents the result in a concise manner.

We will give an overview of the system in Section II and discuss the major components in more detail, the data acquisition from the web (Section II-B), the statistical polarity classification (Section II-C) and the linguistic information extraction (IE) components (SectionII-D). The user interface will be presented in Section III. In Section IV evaluation results for the analysis system will be presented. In Section V we will relate our work to other work in opinion mining.

II. Overview of the System

The system presented here is part of the BESAHOT project.[1] The target users are hoteliers who want to get actual overviews and summaries of textual comments about their hotel(s) on the web. At present, only German reviews from German sites are handled.

The BESAHOT system is an interactive web application based on the GWT framework. The core system on the server-side handles *data acquisition*, *analysis* and *storage* as shown in

This work was supported by a grant from the Saarland Ministry Of Economics and by the European Commission under contract number FP7-231527 (IKS).

[1]BEwertung SAarländischer HOTels im Web (Reviews of hotels in Saarland on the web). Saarland is one of the sixteen states of Germany.

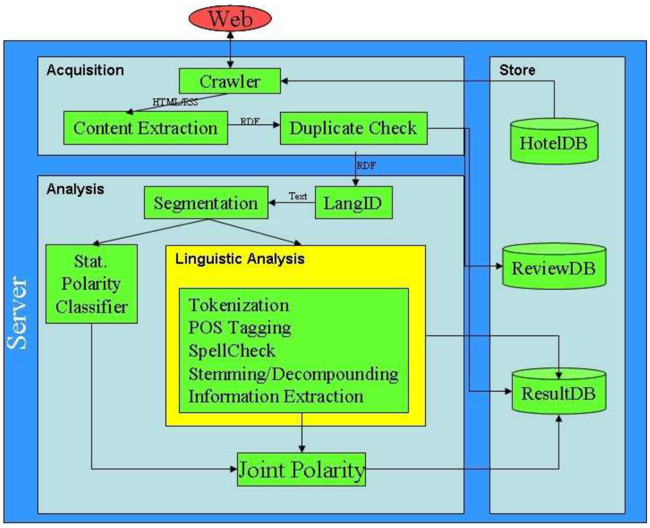

Fig. 1. BESAHOT system overview

Fig. 1. The user interface provides various types of summaries of the analyzed data, allows direct access to the information sources on the web as well as free text search.

New data retrieved from the web by the acquisition system (cf. Section II-B) are passed to the analysis system. The analysis systems first does a *language check* (*LangID*) to filter out reviews in other languages than German because even German hotel review sites occasionally contain reviews in other languages. The review texts then get segmented into segments ("sentences").[2] These segments are then subjected to further analysis by the *statistical polarity classifier* (cf. Section II-C) and *linguistic information extraction* components (cf. Section II-D) for finer grained analysis of the polarity and the topic of the review. Polarity values are always assigned to text segments, not to reviews as a whole.

The polarity values from the statistical and the linguistic classification are then combined into a joint global polarity value that is used for presenting the segments in the user interface.

Finally, the analysis results are stored together with the review segments in a special *ResultDB* optimized to the retrieval and interaction requirements of the user interface.

A. Polarities

In general, we distinguish three possible polarity values for text segments: the segment can express a *positive* opinion, a *negative* one or a *neutral* one. By neutral segments, we mean purely descriptive ones that do not carry an evaluation, such as *We spent three days at the hotel*. The delimitation of neutral/descriptive and evaluative text is not always easy, not even to humans. A remark like *no minibar* on the one hand just describes a fact but on the other hand the remark is probably meant as a negative comment describing a deficiency.

[2]We prefer the term "segments" to "sentence" because the segments are not always sentences in a linguistic sentence but just phrases.

Another problem for a polarity classification on text segments is that a segment might address more than one topic. For example, *clean rooms and friendly personnel* addresses the two topics *room* and *personnel* both rated as positive. But for a comment like *Room ok, but poor breakfast* it would be unclear what the overall polarity value of the comment should be, as there are actually two ratings on two different topics. Similar issues arise with respect to multiple ratings on the same topic as in *clean, but tiny room*.

The BESAHOT IE system is able to detect such multiple topics and ratings on a text segment. Nevertheless, as we have not yet found a good solution for handling these cases in the user interface, at present we prefer to disregard them in favor of a global polarity assignment, even if that sometimes might be a bit random. This will be further discussed in Section II-E and Section IV.

B. The Acquisition System

The acquisition of reviews from the web is handled by a web crawler. The *HotelDB* defines for each hotel a set of crawl configurations that define a start URL, URL patterns for links to follow, target URL patterns for pages containing reviews, the potential crawl depth and an indicator whether the content of a target page is mutable or not. The crawler handles HTML pages as well as RSS feeds. All the URLs usually point to dynamic web pages, that is, the content of the web pages can change between visits. Also, the web pages most times contain hundreds of links, most of them being irrelevant for retrieving reviews (e.g. advertisements, other hotels, etc). Therefore, filter patterns are used to restrict the crawler to follow only relevant links. The distinction between links to follow and target pages is required because the crawler often has to go through several intermediate pages to get at the review pages, e.g. from the hotel overview page to the review overview page to individual review pages and to more reviews.

At present, we ignore sites that present only numeric scores for hotel ratings and no textual reviews. Also, when we found that sites use the same data source for the reviews, we chose one of the sites as a representative and do not use the alternative sites for data retrieval.

When a target page is retrieved a *content extraction module* is applied that extracts the relevant textual content of the review but also other metadata such as scores and information about the reviewer/guests. The content extraction is based on XSLT scripts for known sites (*screen scraping*). If a page contains several reviews, for each of them a separate review instance is created. Extracted content is represented as RDF instance of a *Review* ontology defined in OWL (*Ontology Web Language*, http://www.w3.org/2004/OWL/). Fig. 2 shows an example of the structure.

Since the content of the web page is dynamic the system needs to determine whether it has seen a review before or whether it is a new review. The *duplicate check* uses review fingerprints created from the textual content without any formatting. This provides reliable and efficient tests independent of text size and formatting. Reviews that survive the duplicate

```
<bes:Review>
  <bes:about rdf:resource="urn:hotel:687_02"/>
  <bes:fullText>>
    Parkmöglichkeiten eingeschränkt ...
  </bes:fullText>
  <bes:reviewer>
    <bes:Guest>
      <bes:travelTime>Juli 2010</bes:travelTime>
      <bes:age>45-50</bes:age>
      <bes:guestType>
         geschäftlich allein reisend
      </bes:guestType>
    </bes:Guest>
  </bes:reviewer>
  <bes:source rdf:about="http://www.hotel..."/>
  <bes:rating>
    <bes:Rating>
      <bes:ratingCategory>
        Gesamtbewertung
      </bes:ratingCategory>
      <bes:ratingScore>
        8,1 von 10
      </bes:ratingScore>
    </bes:Rating>
  </bes:rating>
</bes:Review>
```

Fig. 2. Extracted content as RDF

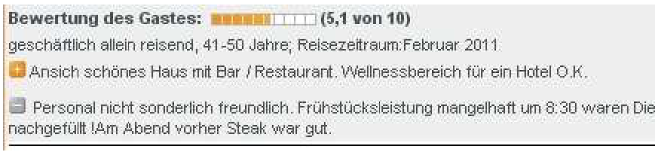

Fig. 3. Polarity classification by users.

check are stored in the *ReviewDB* and passed to the analysis system. The review texts there first are split into text segments that become the units of further analysis.

C. Statistical Polarity Classification

The statistical polarity classifier assigns to each text segment a polarity value. As a basis for statistical polarity classification we used the classification engine of [1]. This engine is based on *character n-grams* instead of terms. For our application this approach has several advantages.

- robustness against orthographic errors that are quite frequent in the reviews, especially transposed or omitted letters.
- robustness against unknown terms from word compounding that are very frequent in German
- it diminishes the sparse data problem as no huge training corpus is required
- applicability to short texts such as segments, not just longer documents

For getting training data for the statistical classifier we exploited the fact that on some hotel sites users themselves classify their contributions into positive and negative text items. An example is shown in Fig. 3.

So we collected a corpus of such hotel review texts from these sites and used them for training the classifier with 2 polarity classes (positive/negative). We use 4-grams with Goodman smoothing ([2]), trained on roughly 7200 text segments

TABLE I
CLASSIFIER BENCHMARK: 10-FOLD CROSSVALIDATION

Training	Precision	Recall	F-Measure
50%	0.90002	0.90017	0.90008
90%	0.92846	0.92855	0.92851

```
premod_nn :>
    (@seek(quantifiers_rule) &
          quantifier &
                [ NEGPOL #neg ])?
    (@seek(conj_adj_phrase) &
          %mods & property &
                [ NEGPOL #neg ])*
    @seek(noun_match) &
          gazetteer &
                [ SUPERCLASS #class,
                 SURFACE #surf,
                 POLAR #pol ]
    ->
    object & [OBJECT #surf,
                 CATEGORY #class,
                 NEGPOL #neg,
                 LEXPOL #pol,
                 RATING %<mods>].
```

Fig. 4. A SProUT rule for NPs

for each class. Crossvalidation benchmarks demonstrated a satisfactory performance as shown in Table I.

The benchmarks illustrate the robustness of the classifier: performance of the classifier does not increase very much when more data are used for training.

We use only two polarity values for the statistical classifier. An experiment to add a neutral category from manually classified data showed a clear performance degradation. Therefore, we preferred to leave the detection of neutral segments to the IE.

In Section IV we will further discuss the performance with respect to manually annotated data and the problem of multi-topic and neutral text segments.

D. Information Extraction

The main task of the linguistic analysis components in the BESAHOT system is to identify from a text segment its *topics* (what is talked about) and how these get rated within the segment. The core of that analysis is an information extraction (IE) component based on the SProUT platform (Shallow Processing with Unification and Typed Feature Structures; [3]). SProUT is a rule based IE system combining finite state technology with unification on typed feature structures for imposing type constraints on possible feature values and propagating constraints by coreferences. Fig. 4 shows an example for a rule in the SProUT system. The left-hand side of each rule consists of a regular pattern over the input sequence, while the right-hand side specifies the output structure. The *@seek* operator allows to call other rules and use their output.

The IE system is designed to supply answers to the following questions:

- *Topic* of the review segment: what is evaluated?
- *Dimension* of the evaluation: what properties are evaluated?

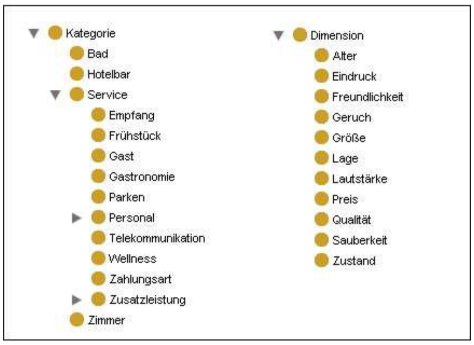

Fig. 5. Main topics and dimensions in the review ontology.

- *Dimension value*: what is the value on that dimension?
- *Polarity* of the evaluation: is it positive or negative or none at all (neutral)?

For the IE component we created a dictionary of *domain-specific terms* relevant for the hotel domain as well as a *sentiment dictionary* that associates basic polarity values with terms. Besides that, the dictionaries assign topic terms to a semantic category indicating what aspect of hotels this topic refers to, e.g. *Service*. Also, the *dimension* of evaluative terms are defined by the dictionary. Fig. 5 gives an impression of these categories and dimensions.

The IE system distinguishes several types of possible roles for a polarity value that influence in different ways what actual polarity is expressed in a segment.

- evaluative speech act indicators, such as *regrettably*. These can override any other polarity expressed.
- negation particles, e.g. *not* that will turn polarities in their scope to the opposite.[3]
- polarity modifiers, e.g. the *too* in a phrase like *too small* that can override the default polarity at phrasal level.
- “missing things” indicators, such as *without*
- negative and positive polarity items as well as idiomatic polarity expressions
- a default lexical polarity, e.g that *nice* expresses a positive rating

Fig. 6 gives an impression of the IE markup applied to stemmed text input, each line representing a text segment. Each colored sequence represents one or more semantic annotations on the text.

Fig. 7 depicts the semantic representation of *kein kostenloses schnelles WLAN (no free fast WLAN)* from IE as a feature structure. It can be read as follows: *WLAN* is the topic belonging to the *telecommunication* category. There are two properties attached that by default denote positive properties (*free, fast*), shown as values of the LEXPOL feature. But these occurrences are in the scope of a negation polarity, the NEGPOL value that is propagated down to the rating elements

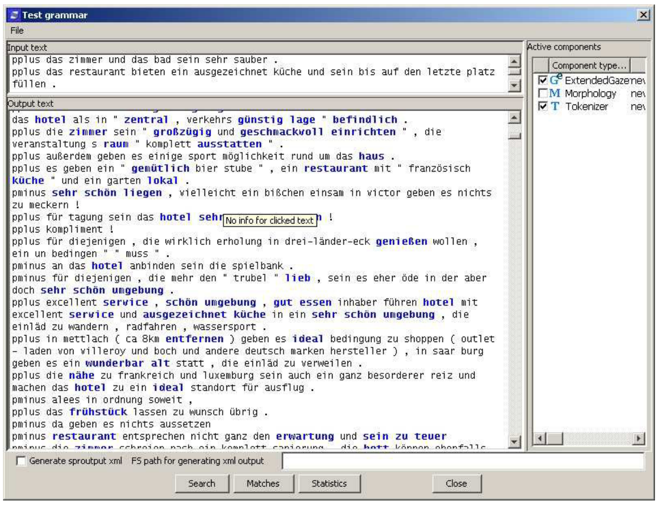

Fig. 6. Information extraction markup.

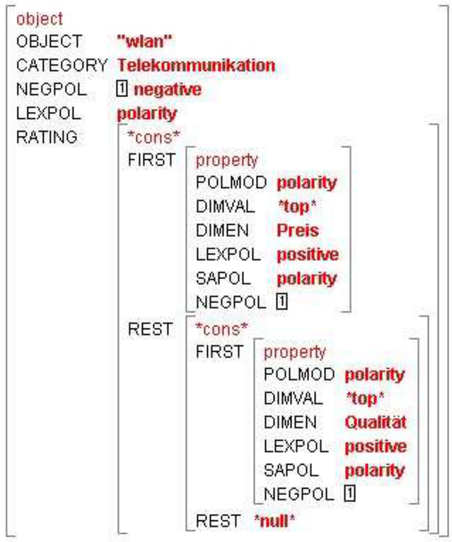

Fig. 7. Semantic representation for *kein kostenloses schnelles WLAN*

by a coreference and that will invert these default values.[4] This is handled by an IE postprocessor. So in the end we will have two negative ratings for the WLAN topic as being neither free nor fast.

Since the review texts often are not well-formed linguistically with respect to orthography and grammar, some preprocessing and normalization steps are applied before actually submitting the text to the SProUT IE system. Part of Speech (POS) tagging is used to reduce the search space for lexically ambiguous words and word forms. Also, to improve input quality a spellcheck is applied that automatically can correct frequent types of spelling errors like transposed and left-out

[3]Of course, this is a simplified assumption: *not bad* does not mean the same as *good*, but in this context we ignore such subtle distinctions.

[4]The value *polarity* on any of the ∗POL features that correspond to the different roles of the polarity values just designates a neutral value, that is, neither positive nor negative.

characters. To prevent over-correction the similarity measure between word and possible replacement must be set very high.

A frequent problem in processing German is word composition by which several terms are combined into a single word. This compounding generates new words missing from the usual dictionaries and so these are difficult to process. The SProUT morphology includes a decomposition component for German compounds that allows us to handle compounds as multi-word expressions and simplifies building the semantic dictionaries.[5] The morphological stem assignment is also used to correct strange POS tag assignments from the tagger for terms for which the morphology provides a more plausible POS.

After the SProUT IE has marked up the found structures, the resulting feature structures are passed to a postprocessor that evaluates the found structures and computes the final rating values for a segment taking into account the different types of polarities and their scopes. This postprocessor would recognize that the positive lexical default polarity values of the adjectives in Fig, 7 occur in the scope of a negative polarity marker and therefore would invert them such yielding finally 2 negative ratings instead of 2 positive and some negative polarity. Also, isolated annotations that cannot be related to ratings get eliminated here.

It is obvious that for the IE system the representation of multiple topics and multiple ratings in a text segment is not a problem. Also, we treat the absence of rating annotations in a segment as evidence that the segment belongs to the neutral polarity category.

E. Combining Statistical and IE Polarities

For each segment the statistical polarity classifier yields a positive or negative polarity value. More fine-grained polarity values are available even for parts of the segments (subsegments). We developed an experimental system that would use the IE to create finer phrases as subsegments of the text segments according to the recognized topic changes. Unfortunately, in many cases that resulted in text fragments that are incomprehensible without their syntactic context and so cannot be presented to users.[6] Therefore we kept to the approach to assign a global polarity value to the whole text segment, but the assignment of that global value would take into account both classification sources, the statistical value and the IE values. In that approach, the statistical value is regarded as baseline value and the ratings from IE are used to possibly correct that value. As an approach that would give the IE ratings preference to the statistical value proved unsatisfactory, we developed a method for using the IE ratings as length-normalized weights on the statistical values: for each polarity, the IE weight is defined as the number of ratings of that polarity divided by the token length of the segments. On short segments, the IE ratings thus will have larger weight than on longer segments. The global polarity values then are computed by combining the scores of the statistical classifier with these weights according to (1).

$$pol = \underset{p \in \{pos,neg\}}{\arg\max} \frac{sp(p)}{(1 + ie(p)/sl)} \tag{1}$$

where p is a polarity, $sp(p)$ its statistical score, sl the segment length and $ie(p)$ the number of the IE ratings with that polarity. This approach reconciles the confidence of the statistical classifier with the IE results better than a preference based approach. A side effect of the formula is that the statistical polarity value will be kept, if the IE does not yield ratings. The motivation for this is that the statistical classifier has larger coverage than the current IE. Therefore we keep the statistical polarity value and treat the absence of IE ratings as meaning "IE does not know" rather than "This is neutral polarity". This provides more flexibility for the user interface that can decide how to handle this case.

III. The User Interface

The BESAHOT system is a tool to support hotel managers in quality control. So it should provide them with fast and comprehensive overviews and summaries of how their hotel is rated on the web and how it is commented on by guests and visitors on the web.

Fig. 8 shows the main result overview that the user will see when accessing the BESAHOT service after selecting a hotel. The top panel displays some statistics about scores from source sites, normalized to a scale between 1 and 10, and about guest types, as far as this information could be extracted from the source web pages. Also, the time range can be restricted to show only recent reviews. The *Aktualisieren* button allows to start the crawler to search for new reviews on the web for the selected hotel.[7]

The main panel provides a summary of the reviews by displaying text snippets from the reviews according to their polarity and category. A click on a segment opens a popup panel that displays the full review text highlighting the displayed segment in context. This allows users to check the text in context and also makes it unnecessary to visit the source page, though this would be easy by just following the provided link to the source page. Additionally, the popup displays information about the guest that provided the rating.

For this display we exploit the IE's capability to identify neutral text segments: text segments that do not receive an IE rating here are omitted from the view. An open issue in designing the user interface is the handling of text segments belonging to more than on category. Adding these segments to each category tends to result in rather crowded and redundant category fields, impairing the usefulness. So, presently such

[5] Usually, the last component of a compound is regarded as the headword as that governs the morphological properties of the compound. Semantically, we found that often the other components are more significant.

[6] A possible solution would be the use of a text generator to generate some simplified text from the semantic structures of the IE instead of using only text pieces from the original review texts. At present, this is outside the scope of the project.

[7] This *Actualize* button exists only in the demonstration system. In the final system the server would automatically update the databases periodically.

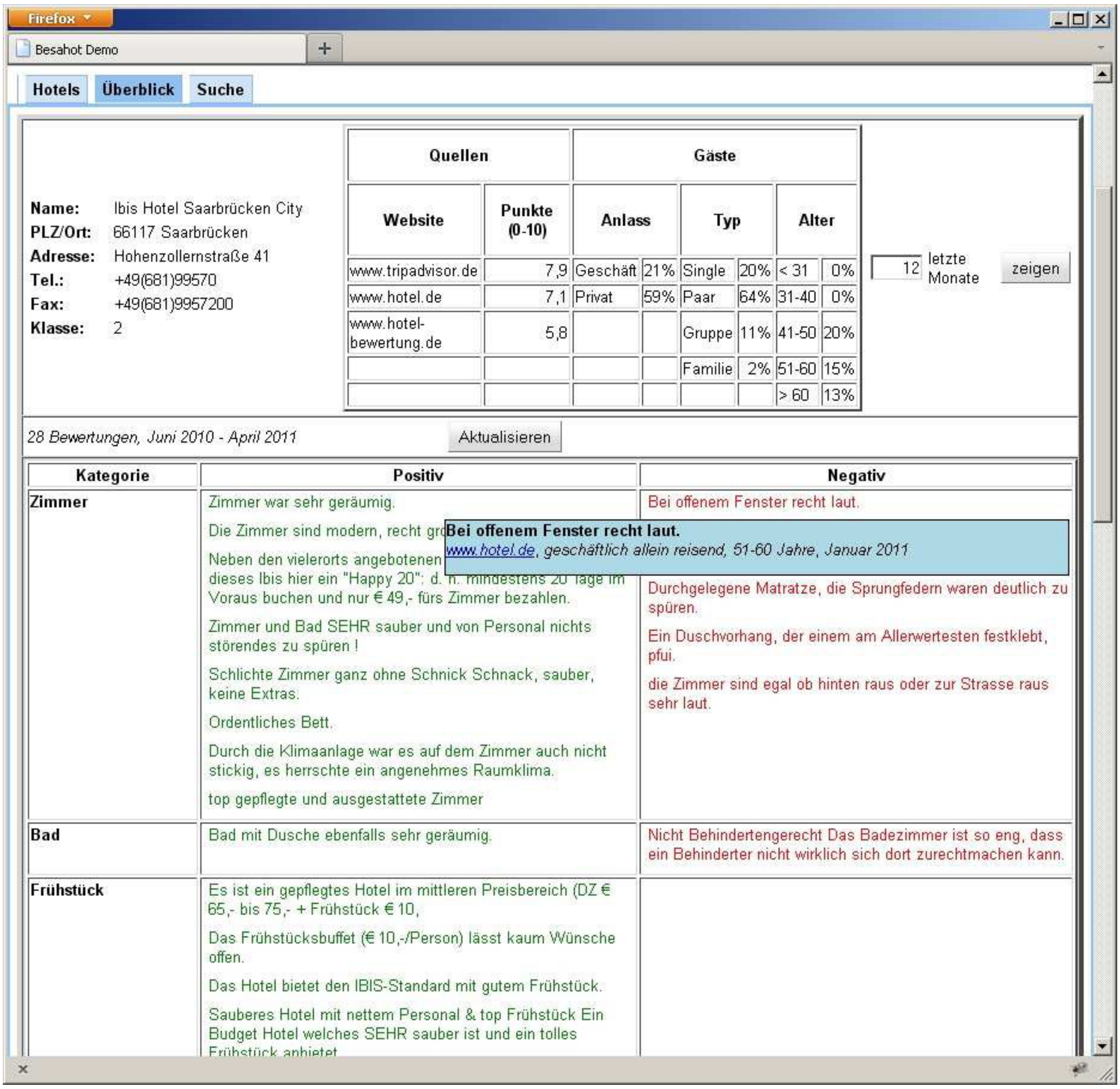

Fig. 8. Classified review summary for a given hotel.

ambiguous segments currently are displayed only in one category, preferably a dominant one.

In addition to the overview presentation, a free text search function allows users to search the review database by freely chosen keywords, independent of the predefined categories and polarity values.

A usability test for the user interface with members of the Saarland hotel association is in preparation.

IV. Evaluation

We evaluated the analysis system on a corpus of 1559 hotel reviews crawled from the web. These reviews contained 4792 text segments. For the evaluation, these segments were manually classified with respect to their polarity, including the neutral polarity besides positive and negative ones. Also, we annotated the segments whether they cover more than one topic. The distribution from this manual classification is shown in Table II.

TABLE II
Manual corpus classification

Segments	positive	negative	neutral	multi-topic
4792	2240	1183	938	431

We evaluated the performance of the statistical classifier alone, the IE system alone and the hybrid system combining the polarity classifications from the statistical classifier and the IE system as described in Section II-E. Evaluated on all segments, the results in Table III were achieved.

TABLE III
CLASSIFICATION ON ALL SEGMENTS

	Correct	False	Accuracy
Stat	3145	705	0,66
IE	2604	486	0,54
Stat+IE	3208	646	0,67

TABLE IV
CLASSIFICATION WITHOUT NEUTRALS

	Correct	False	Total	Accuracy	F-measure
Stat	3145	705	3854	0,82	0,80
IE	2604	486	3090	0,68	0,66
Stat+IE	3208	646	3854	0,83	0,81

It shows that the IE system currently covers less data than the statistical classification, but that it slightly improves the overall classification accuracy. These data relate the results to the complete corpus not taking into account the presence of neutral and multi-topic segments. As discussed in previous sections, the assignment of only positive/negative polarities in these cases can be a bit random, or, for the cases of neutral polarity that make up about 19% of the corpus, the positive/negative assignment is rather uninteresting.

Therefore, in a second experiment, we evaluated the classification performance on only the subset of manually verified positive/negative segments and achieved considerably better results, shown in Table IV.

These values demonstrate that it would be beneficial to be able to identify neutral and multi-topic/multi-polarity ratings. As mentioned in Section II-C the pure statistical classifier did not look promising in that respect. Therefore we evaluated how well the IE system would recognize the neutral and the multi-topic cases identified in our corpus. The results are shown in Table V.

These values look promising. We expect that improving the coverage of the IE system will also improve these figures. That will also provide a strong motivation for changing the interpretation of the absence of a polarity rating from IE as "don't know" to "classify that as *neutral*".

V. RELATED WORK

The development of the WWW and the possibility for customers/users to express their opinion online made the online available reviews interesting for both the vendor as well as for the potential customer. Therefore, the interest on opinions and sentiments of (former or future) customers has increased tremendously. In parallel, the development boosted research in opinion mining and sentiment analysis in recent years. Good overviews on existing opinion mining techniques and methods are given by [4] and [5].

TABLE V
RECOGNITION OF NEUTRAL AND MULTI-TOPIC POLARITY

	Correct	False	Total	Accuracy
Neutrals	682	256	938	0,72
Multi-topics	324	107	431	0,75

Most research in this area concentrates on opinions about products. Also, domains such as movie reviews or news found considerable interest especially in research, since large datasets and corpora are publicly available.

The goal of opinion mining can vary considerably. In many cases, one is only interested in a global overview: how many users/reviews rate a product positive or negative. For these, a global polarity classification is sufficient without having to go into details of a product. More fine-grained is an approach as that of [6] who present an opinion mining approach for news articles. They do not just global classification at document level but split up the review into phrases. Based on a predefined lexicon and contextual information they apply machine learning techniques for determining the polarity of the phrase. But different from our approach, they do not identify specific features that are evaluated.

Research in opinion mining often requires specific resources such as suitably classified corpora and sentiment dictionaries that associate terms with sentiments. For English, a large set of resources is publicly available for research. Therefore also most research is done on English data, such as ([7], [6], [8]). For opinion mining approaches that also do feature extraction for the rated product features, also domain-specific dictionaries can be needed that specify product-specific features.

For German (or other languages), there are less of such resources available, even though the situation starts to improve. A large sentiment dictionary for German has been built by [9] that we used to initialize our sentiment dictionary for the terms extracted from our hotel review corpora. The dictionary of domain-specific terms and concepts for the hotel and tourism domain we had to create ourselves.

While our IE system for feature extraction relies on manually created rules, there are a number of approaches to use machine learning techniques to achieve that, such as the work of [7] on mining opinions about products. They describe an unsupervised information extraction system which determines the relevant features and the corresponding opinion. The method uses relaxation labeling[10] for finding the semantic orientation of words in the context of given product features and sentences. A more linguistically inspired approach that resembles ours is described in [11].

The tourism domain in not one of the mainstream domains for opinion mining research. [8] uses a corpus of English reviews from *tripadvisor.com* in order to present a rule-based method for classifying opinions. Different from other approaches she takes also the context into account. This way she differentiates between the needs of a person on a business trip and the needs of the same person on a family trip. A larger English corpus also from *tripadvisor.com* is used in the study of [12] that uses linguistic preprocessing with the SENTIWordnet ([13]) but machine learning techniques for feature assignment. [14] describe in their work a framework for constructing Thai language resource for feature-based opinion mining for hotel reviews. Their approach for extracting features and polarity words from opinionated texts is based on syntactic pattern analysis. In general it is left unclear how the high number of misspelled and ungrammatical data, we found in our corpora, are handled in these approaches and how they affect the result.

In general, these approaches focus on research on specific technologies but there is little indication about what the results are used for in an application, who the users of the results are and how results can be used by them. In many cases the research is related to building recommendation systems so that the results are not directly used by humans but just by machines. The BESAHOT system, on the hand, targets explicitly human users, not machines.

Closely related to BESAHOT is the work on review summarization such as [15], [16]. Summarization there means extracting relevant sentences classified according to their polarity and some category, called *features* or *aspects* in these papers. They focus on adjectives as carriers of polarity and nouns/noun groups as designators for features, ignoring other word classes. Negation seems to be recognized only if adjacent to an opinion term. Irrelevance/neutrality is defined by thresholds on scores. The methods of feature extraction based on nouns in the context of opinion terms tend to yield high numbers of features. [16] therefore introduce a second level of manually created static high-level aspects that resemble more the high-level categories used in BESAHOT. It is unclear whether sentences belonging to more than one category are treated in the user interface in a special way. The BESAHOT-IE approach looks more flexible as it is not restricted to few word classes and it can handle larger contexts and relevant linguistic phenomena better than these approaches. Also, resources for the IE are easy to extend and to adapt for new data and phenomena.

VI. Conclusion

We presented a web based opinion mining system for hotel reviews and user comments that supports the hotel management in monitoring what is published on the web about their houses. The system is capable of detecting and retrieving reviews on the web, to classify and analyze them, as well as to generate comprehensive overviews of these comments. We showed that, despite some remaining issues, the system provides good performance for the analysis and the classification tasks. Further research will be necessary especially with respect to the demarcation of evaluative and neutral text as well as to the handling of multi-topic segments, especially for the user interface.

Besides that extension of coverage to more sites is under work. One further direction is to include web search into the data acquisition to find reviews on sites that only infrequently or just by chance publish guest comments on hotels registered on the BESAHOT service. Also, we are preparing a pilot test of the BESAHOT service with members of the Saarland hotel association to improve the information value and usability of the system.

References

[1] J. Steffen, “N-Gram Language Modeling for Robust Multi-Lingual Document Classification,” in *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC-2004).* Lisboa: ELRA, 2004, pp. 731–734.

[2] S. F. Chen and J. Goodman, “An empirical study of smoothing techniques for language modeling,” Computer Science Group, Harvard University, Cambridge (Mass.), Tech. Rep. TR-10-98, 1998. [Online]. Available: http://research.microsoft.com/~joshuago/tr-10-98.pdf

[3] W. Drozdzynski, H.-U. Krieger, J. Piskorski, U. Schäfer, and F. Xu, “Shallow processing with unification and typed feature structures — foundations and applications,” *Künstliche Intelligenz*, vol. 1, pp. 17–23, 2004.

[4] B. Pang and L. Lee, “Opinion mining and sentiment analysis,” *Foundations and Trends in Information Retrieval*, vol. 2, no. 1-2, pp. 1–135, 2008.

[5] B. Liu, “Opinion mining and sentiment analysis,” *Handbook of Natural Language Processing*, 2010.

[6] T. Wilson, J. Wiebe, and P. Hoffmann, “Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis,” *Computational Linguistics*, 2005.

[7] A.-M. Popescu and O. Etzioni, “Extracting product features and opinions from reviews,” in *Proceedings of the Human Language Technology Conference and the Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP)*, 2005.

[8] S. Aciar, “Mining context information from consumer’s reviews,” in *Proceedings of the Context-Aware Recommender Systems (CARS) Workshop*, 2009.

[9] U. Waltinger, “Germanpolarityclues: A lexical resource for german sentiment analysis,” in *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC)*, 2010.

[10] R. A. Hummel and S. W. Zucker, “On the foundations of relaxation labelling processes,” *PAMI*, 1983.

[11] L.-W. Ku, T.-H. Huang, and H.-H. Chen, “Using morphological and syntactic structures for chinese opinion analysis,” in *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3 - Volume 3*, 2009.

[12] S. Baccianella, A. Esuli, and F. Sebastiani, “Multi-facet rating of product reviews,” in *Proceedings of the 31th European Conference on IR Research on Advances in Information Retrieval*, ser. ECIR ’09. Berlin, Heidelberg: Springer-Verlag, 2009, pp. 461–472.

[13] A. Esuli and F. Sebastiani, “Sentiwordnet: A publicly available lexical resource for opinion mining,” in *In Proceedings of the 5th Conference on Language Resources and Evaluation (LREC-06*, 2006.

[14] C. Haruechaiyasak, A. Kongthon, P. Palingoon, and C. Sangkeettrakarn, “Constructing thai opinion mining resource: A case study on hotel reviews,” in *Proceedings of the Eighth Workshop on Asian Language Resouces*, 2010.

[15] M. Hu and B. Liu, “Mining and summarizing customer reviews,” in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2004.

[16] S. Blair-Goldensohn, K. Hannan, R. McDonald, T. Neylon, G. A. Reis, and J. Reynar, “Building a sentiment summarizer for local service reviews,” in *NLPIX2008*, Beijing, 2008.