Chapter 1
Intelligent News Aggregator for German with Sentiment Analysis

Danuta Ploch

Abstract The comprehensive supply of information from different points of view, e.g., from the thousands of news articles published online every day, is a tremendous advantage of the digital era. However, the immense amount of news material poses a significant challenge to interested readers: It is hardly possible to fully digest this wealth of information, so that the need for systems supporting intelligent news consumption arises. This chapter describes an approach to automatically mining opinions from topically related news article clusters. We focus our work on the extraction of quotations from German news articles and on analyzing the quotations according to the sentiments they express. Our approach is realized as a news aggregation system capable of handling real-world news streams. We describe the architecture and interface of our news aggregator, and present a rule-based method for quotation extraction as well as our supervised approach to sentiment analysis. We evaluate the implemented models on two human-annotated datasets, which can be made available upon request.

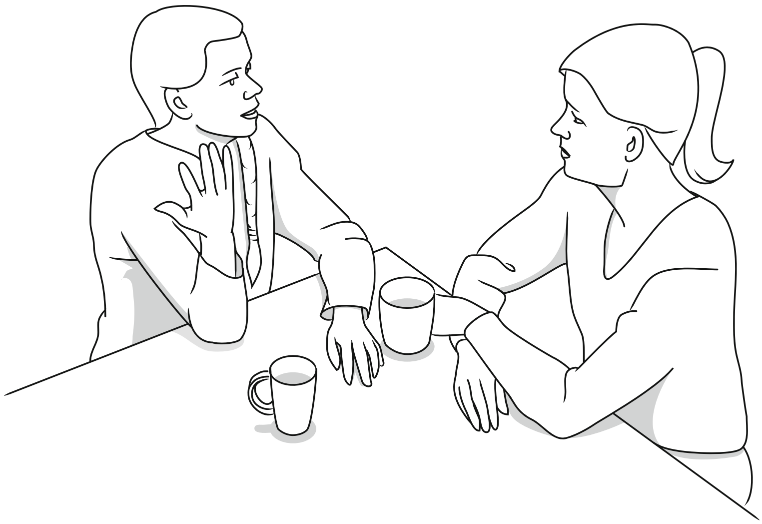
"**As Many Heads, So Many Opinions**" (**Horace**)

Since her 15th birthday Clara dreamed of becoming a journalist. She worked for the school newspapers and was a member of the debating society. Now, after passing the exams, Clara finally started living her dream. She remembered the day when she received the good news that they had accepted her application for an internship at the local newspaper. "That's a great chance for you, don't ruin it", her father told her. What a typical statement from her dad, she thought. Why would she ruin it? And what is there to ruin in an internship anyway? In fact, her tasks so far seemed quite forward. Even Dad would be able to do that, she thought with a grim on her face. At the beginning her tasks were restricted to copying articles and typing e-mails, but her boss quickly recognized her talent and assigned her an important research task for

D. Ploch (✉)
Technische Universität Berlin, Berlin, Germany
e-mail: danuta.ploch@dai-labor.de

F. Hopfgartner (ed.), *Smart Information Systems*, Advances in Computer Vision and Pattern Recognition, DOI 10.1007/978-3-319-14178-7_1

a special issue on “espionage amongst friends”, a topic that gained attention after the revelations of the US-whistleblower Edward Snowden in 2013. In particular, Clara and her internship friend Martha were supposed to analyze newspaper reports about the behavior of US-president Barack Obama and the German Chancellor Angela Merkel before and after the revelations. The interns should identify the politicians’ attitudes toward national security-related topics and if they possibly changed after the publication of the secret documents by Snowden. How the politicians have commented on the revelations and spying in general? Did they disagree on all points or was there also agreement? Which statements about spying among allies were the most controversial? The interns were also advised to work out whether Snowden’s actions influenced Merkel’s positions to crucial election topics during her campaign for the German federal election in 2013.



Clara and Martha split up their task into two parts. While Martha took over the investigation concerning Barack Obama, Clara started to search for news articles about Angela Merkel. She entered ‘Merkel Snowden 2013’ in the search field of her search engine and immediately thousands of news articles were presented to her as a never ending result list. “It will take days to go through all that news articles”, Clara whined. But since Clara has always been a fighter, she sorted the news articles by date and began with her research—news by news. Clara realized soon that sorting news articles by date was not really helpful for detecting topics and positions. Although she limited the date range and disabled displaying duplicates, the mass of news articles was overwhelming and she had to face problems like off-topic news articles concerning nonpolitical issues. During the coffee-break she met her friend Martha and complained: “I’m feeling like Cinderella: ‘The good ones go into the pot, the bad ones go into your crop’. “I thought journalism would be fun”, she added and couldn’t hide her sense of frustration. “But it is!”, answered Martha, Haven’t you tried out our in-house news archiving system? It does all the stressful preprocessing-work for you. Depending on your search query the system clusters news articles to topics and identifies thematically connected topics. It even extracts quotations and their polarity toward the contained opinion target". Clara had already worked for 2 months for the newspaper but had never heard of such a system. Fortunately, Martha told her now because 10 min ago she was on the verge of discarding her evening plans and working overtime instead of going to the cinema. She grabbed her coffee cup and returned to the office with a smile on her face—her evening was saved.

1.1 Introduction

In times of Twitter, Facebook, and other social media services news is broadcast around the world in no time at all. If something more or less newsworthy happens, users immediately take their smartphones and post it. Today's social media services bring a lot of benefits. For example, it is not possible to imagine reporting without Twitter from crisis regions. Still, since potentially everyone may distribute information, the question of reliability arises. In 2013, the ambiguous Twitter hashtag `#nowthatchersdead` confused a wide range of Twitter followers.[1] Many users interpreted the hashtag as "Now that Cher's dead" and retweeted that the pop queen Cher has died. In reality, the death of the former Prime Minister Margaret Thatcher ("Now Thatcher's dead") was announced. This example shows how fast rumors may come up and be spread around if no professional journalists are involved. Thus, editorially written news articles still remain one of the most important information sources.

In comparison to user-generated reports on the Web, journalistic work of reputable news agencies is considered reliable and credible. News articles are thoroughly researched and well formulated and they often report not only the piece of news itself but also provide additional context and background information. Besides, local newspapers or special issues on specific topics may cover topics not mentioned in social media which still are important for a large number of readers. As with user-generated content, the amount of editorially prepared news material is remarkable. In general, print media publish their data also on the Web and in addition there are numerous online news papers available.

The range of available news material allows users to stay informed about what happened and to look at news from different points of view. At the same time the considerable number of news articles complicates their handling and requires tools for helping the users to search and browse them. News aggregation systems support users in exploring news articles by analyzing and organizing news articles. To avoid information overload, they first detect (nearly) duplicate news articles and hide them from the reader. This is necessary because news articles published by well-known press agencies like *dpa*[2] (Deutsche Presse-Agentur GmbH) are redistributed by various news providers. Then, their main task is to identify news articles dealing with

[1] http://news.msn.com/pop-culture/confused-by-thatcher-tweets-cher-fans-upset-by-numbernowthatchersdead/.

[2] http://www.dpa.de/.

the same news story or topic from a large stream of text messages. As news stories evolve over time, they continuously group and rank news articles in order to constantly reveal relevant and hot topics and to enable their monitoring.

News texts do not only report facts about what has happened but also reflect opinions of involved entities such as persons or organizations. They serve therefore as a valuable source of opinions and help users making their decisions based on them. Depending on the news type the opinions are directed toward a wide variety of topics or other entities. For example, before political elections news articles echo the politicians' attitudes toward current election issues and influence the vote behavior. The perception of products or services is a precious piece of information for companies and often a key factor in a company's decision-making process.

As with events and topics, news aggregation services facilitate the search for and exploitation of opinionated text. Manually finding and evaluating opinion-relevant parts may be infeasible for users. Therefore, the detection of subjective text parts and the classification of text into different types of opinions are crucial tasks in news processing systems.

In this chapter we focus on news aggregation services to organize and analyze news articles. Section 1.2 describes approaches for grouping news articles depending on events and topics. We start by describing methods to detect and track short-term events in news streams. Then, we discuss the clustering of events into more abstract meta-topics. News articles often contain citations that underline reported issues. Therefore, Sect. 1.3 concentrates on the extraction and evaluation of citations. Section 1.4 covers services to analyze news material with regard to expressed opinions. We present a news aggregation system that incorporates all introduced steps in Sect. 1.5 and conclude the chapter in Sect. 1.6.

1.2 News Aggregation Model

News aggregation systems like Google,[3] Bing[4] and Yahoo![5] organize and present news articles from a large number of sources in order to offer users a comprehensive supply of information. The enormous amount of news material published every day requires a continuous and suitable preparation. In addition to the standard categorization of news article into the main columns such as "Politics", "Economy", "Sports", etc., and sorting the news items by date and/or language news aggregators apply Topic Detection and Tracking (TDT) techniques to group news articles related to the same events. Enhanced news aggregators offer additional services based on a deep analysis of the news sources and material. For example, Google assesses the sources and categorizes the content as *opinionated*, *detailed* or *preferred by the user*. We introduce a system that not only focuses on a high-level classification of news

[3] https://news.google.com/.

[4] http://www.bing.com/news/.

[5] http://news.yahoo.com/.

articles but also examines the content of the articles in order to grasp their meaning. Besides TDT at topic-level, the proposed system recognizes more abstract topics and performs quotation extraction and sentiment analysis based on the identified quotations. The system is capable of offering deep insight into single events and topics by highlighting named entities along with direct and indirect quotations. The users may inform themselves about involved entities, compare their comments, and learn about the perception of the entities and topics in the media landscape.

The proposed system was developed in close collaboration with Neofonie GmbH.[6] It is modeled with a processing pipeline as the central component. The system's structure is schematically outlined in Fig. 1.1. When going through the processing pipeline, the documents are enriched with more and more information. For each crawled news article a linguistic preprocessing is performed. The news articles are split into tokens and sentences and annotated with part-of-speech tags, named entities and lemmas (Sect. 1.2.1). Subsequently, the news articles are mined. On the

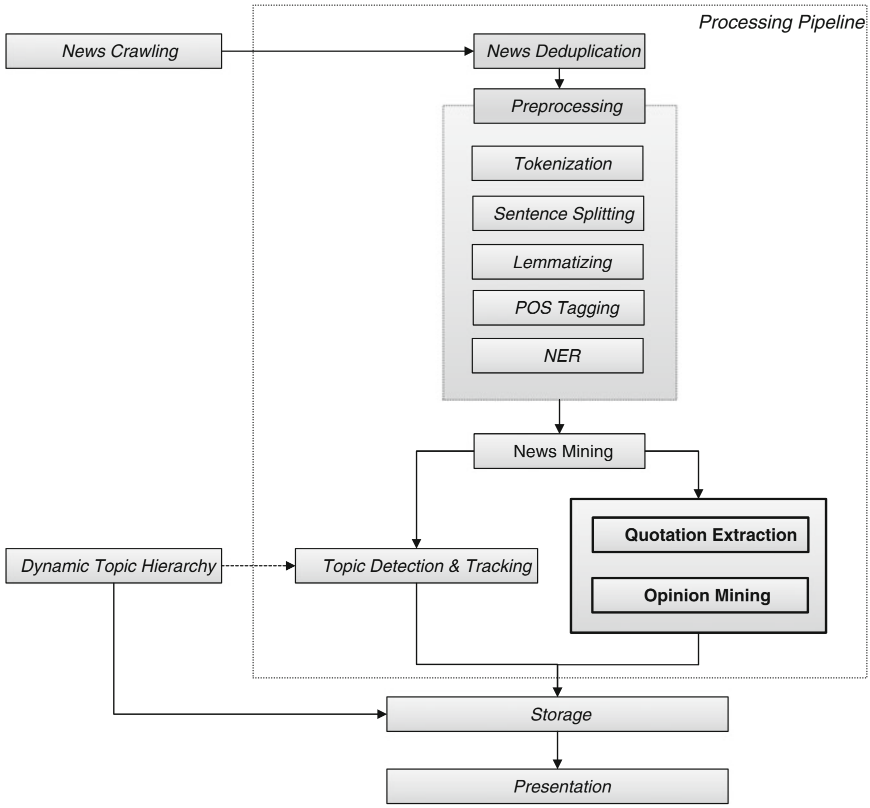


Fig. 1.1 System overview

[6] http://www.neofonie.de/.

one hand the news texts are regarded as objective information sources reporting facts. They are clustered according to events by the Neofonie GmbH (Sect. 1.2.2). On the other hand the system aims at identifying subjective parts in the form of quotations (Sect. 1.2.4) and at determining the sentiment polarity of the expressed statements (Sect. 1.2.5). In parallel to the pipeline, the Neofonie GmbH assigns the news articles to automatically identified abstract meta-topics, which connect thematically related topics (Sect. 1.2.3). The graphical user interface is described in Sect. 1.5 where also screenshots of the system are presented.

1.2.1 Preprocessing

The initial step for the analysis of news articles is linguistic preprocessing. After having crawled and deduplicated the news articles, we first split the text of each article into tokens and sentences. This is an important prerequisite for various linguistic tasks such as part-of-speech (POS) tagging, chunking, and also our quotation extraction approach. The next step is lemmatizing all words of the text. Mapping words to their canonical form allows looking them up in dictionaries in steps that follow. We find verbs starting quotations in this way. The task of POS taggers is to assign each word of a text its part of speech. We use the output of the POS tagger at several points in our system. For instance, we make use of POS information to compile feature vectors for our supervised sentiment analysis approach. The system exploits a lexicon-based named entity recognition approach. We use the German version of Wikipedia[7] for identifying and linking named entities. The named entities contribute to the concept vectors required for our topic detection and tracking approach (Sect. 1.2.2).

1.2.2 Topic Detection and Tracking

The Topic Detection and Tracking program defines an event as "something that happens at some specific time and place along with all necessary preconditions and unavoidable consequences". Topics (or stories) comprise a triggering event and all directly related events and activities [17]. As news stories evolve over time, the task of TDT approaches is to either identify news articles starting a topic or to assign news articles to existing topics. In our system we employ an incremental agglomerative clustering approach for TDT. We represent each news document as a vector of concepts including named entities. Each cluster represents a topic and is specified by a centroid vector with averaged concept weights of the covered news articles. Incoming news documents are compared to the centroid vectors of all existing topics. If

[7] http://de.wikipedia.org/.

the similarity to all existing centroids is lower than a predefined threshold then the incoming news article starts a new topic. The similarity measure is a combination of the cosine similarity between two vectors and a time-dependent penalty. Including a time-dependent penalty favors the assignment of news articles to new topics and at the same time avoids an infinite growth of existing topics [20].

1.2.3 Dynamic Topic Hierarchy

Classical news aggregators organize news articles, and the topic they belong to, in top-level categories like "Politics", "Economy", or "Sports". In order to better navigate and track the development of related news stories, news stories may be organized in a hierarchy. The arrangement of topics in a hierarchy links not directly related topics and supports readers to recognize relationships between them. This is especially helpful in the context of searching news archives and recommending related news articles for further reading. Since news stories evolve over time and new events happen continuously we do not sort the news articles in predefined hierarchies such as the Metadata Taxonomies for News by the International Press Telecommunications Council[8] (IPTC) but propose the creation of a dynamic topic hierarchy arising from the current news situation. Based on previously detected topics (Sect. 1.2.2) we build thematically connected meta-topics and assign labels to them. We select the most probable headline from the set of news articles belonging to the meta-topic.

1.2.4 Quotation Extraction

Quotations are a common stylistic device to clarify and strengthen a statement. Basically, a distinction is made between direct and reported speech. Considering quotations in news articles, they underline reported facts and may express positions or views of the cited persons or organizations. By employing quotations at specific points in an article the author highlights statements that are especially significant and worth to be cited. In addition, quotations may be a suitable source for identifying subjective passages of a news article. In our system we apply a rule-based approach to quotation extraction. We address the extraction of direct and reported speech and assign a speaker to each identified quotation. Our solution normalizes quotation marks, makes use of linguistic annotations to detect reporting verbs or phrases that introduce quotations, the boundaries of direct and reported quotation parts, and finally the speaker, which we also call quotation holder in the following. We describe our approach to quotation extraction in detail in Sect. 1.3.

[8] http://www.iptc.org/site/NewsCodes/.

1.2.5 Sentiment Analysis

Sentiment analysis aims to detect and assess opinionated text. Often the synonym "opinion mining" is used to refer to the same task. Journalistic content such as news articles is considered to be a reliable information source and may serve as a basis for various natural processing (NLP) tasks. News articles are objective reports and ideally, the authors do not express their attitudes. But this does not mean that news articles cannot contain any opinionated text. Usually, news texts reflect opinions and sentiments of newsworthy people, which are directed toward topics or toward other entities. The identification of passages that echo subjective statements, attitudes, or views and distinguishing them from the objective parts reporting facts is the first step in many sentiment analysis approaches before analyzing the identified passages. Exploiting quotations is one way to tackle the problem of mining opinions from news articles. Quotations are a trustworthy form of mirroring what people or organizations have said in an original and genuine manner. In addition, given a quotation, its speaker is usually the opinion holder, the entity that expresses the opinion. We propose a supervised two-stage approach where we first identify subjective quotations and, second, classify the subjective quotations in either POSITIVE or NEGATIVE. All remaining quotations are regarded as NEUTRAL (Sect. 1.4.3). We explore a range of features established in sentiment analysis of other text genres (e.g., product reviews, tweets) and examine to what extent they are suitable to separate subjective from neutral and positive from negative quotations (Sect. 1.4.6).

1.3 Quotation Extraction

Quotations report what persons or organization have said. In news articles they are often used to confirm claims made by the author and indicate the importance of the transported information. Thus, they may be an important piece of news article for various text processing tasks such as sentiment analysis or news summarization. In this section we present our approach to quotation extraction, which covers both the extraction of direct and indirect quotations as well as the assignment of a quotation speaker.

1.3.1 Introduction

Quotations repeat a speech, text, or statement expressed by a speaker and can be distinguished into direct and reported speech. We refer to reported speech also as indirect speech and to a speaker also as quotation holder in the following. Quotations are composed of a reporting and a reported clause. Following Krestel et al. [24] Fig. 1.2 shows an example of the structure of both, a direct and an indirect quotation. The reporting clause introduces the quotation. Besides the quotation holder, it may

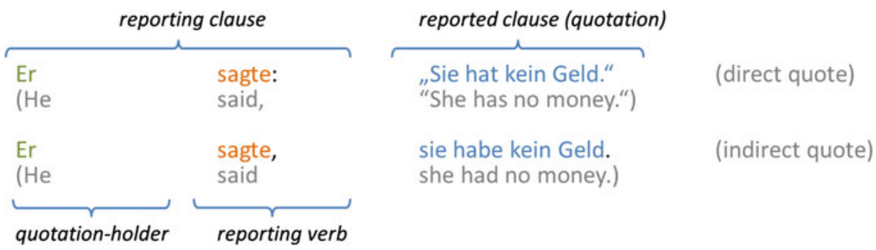


Fig. 1.2 The structure of quotations. A quotation is composed of a reporting and a reported clause. The reporting clause introduces the quotation. It includes the quotation speaker and an optional reporting verb. The reported clause encompasses the quoted content/text

also contain a reporting verb such as "sagte" (said) or "berichtete" (reported) and other circumstantial information such as the addressee or diverse other descriptive text. The reported clause encompasses the actual content that has been said.

Direct speech repeats things that have been said by a speaker as they are without any modifications. The repeated text is enclosed by quotation marks. In contrast to direct speech, *indirect speech* reports statements by modifying them grammatically or even rephrasing them. The grammatical change indicates that the expression was not uttered by the author, but by the original speaker. Indirect speech is composed of a main (reporting) and a subordinate (reported) clause. In German, the reported clause of an indirect speech often is introduced by the conjunction "dass" and uses the subjunctive mood for verbs. Quotations may consist of several reported clauses and we define quotations as *mixed* if both, quoted and unquoted reported clauses build the quotation.

Our Contribution. In our work we process German news articles and extract direct and indirect quotations along with a quotation speaker. We propose a rule-based approach that exploits linguistic information. Modeled as a processing pipeline our quotation extraction component first enriches news articles with linguistic annotations, which then are used to mine the complete quotation. We detect units of direct quotes by applying a pattern that takes into consideration different types of quotation marks. We exploit the presence of reporting verbs and other common phrases indicating quotations to underpin direct quotation candidates found by our pattern and to locate potential indirect quotations. In order to assign each quotation a speaker we make use of the output generated by a named entity recognizer and a part-of-speech tagger. We compile a list of candidate speakers and then apply rules that consider the type of the candidates and the proximity to the reporting verb to determine the quotation speaker. For evaluating our approach we manually created a quotation corpus from a set of German news articles. The corpus provides for each quotation its boundaries, the quotation speaker, a reporting verb or phrase, and the type of the quotation (direct, indirect or mixed). The corpus is available upon request and signing of an agreement.

1.3.2 Related Work

In the past, numerous solutions for the task of quotation extraction from newspaper material were proposed. The approaches differ in which technique they use, which language they support, in whether they extract direct, mixed, or indirect quotations (or all), and in how detailed they determine specific quotation units such as the quotation holder and other circumstantial information.

The majority of previously published work detects reporting verbs in news articles from a predefined or precompiled list and then extracts quotations based on rules that are derived by experts. An exact analysis of the news material in advance and the knowledge about the structure of quotations allow the identification of more or less fine-grained patterns that may vary from language to language. Usually, the patterns differ in the presence and position of lexical terms or syntactic information. There is consensus that for each quote a speaker needs to be extracted, because the information without the assignment of a speaker is of little use in most cases. Thus, many researchers represent quotations as a triple consisting of the quoted text, the quotation holder, and an optional reporting verb or a quotation introducing phrase.

The rule-based system presented by Pouliquen et al. [41] extracts around 2,600 direct quotations per day from a multilingual news stream. In order to keep the system extensible to other languages, the approach does not rely on linguistic information but on lexical patterns. The system recognizes quotation marks, reporting verbs, and person names (along with further information such as temporal or spacial modifiers, titles, and determiners) and applies three general and a couple of language-specific rules to find quotations. A simple named entity disambiguation solution serves as the accurate assignment of quotation holders. Still, the system misses quotations with speakers referenced by pronouns, since it does not perform anaphora resolution.

Krestel et al. [24] assemble a set of six basic patterns to extract quotations from news articles in English. They detect the most frequent reporting verbs using a finite state transducer and implement the identified patterns as a regular grammar. Existing GATE[9] components provide additional circumstantial information required during the quotation extraction process. In contrast to [41], that limit their approach to direct quotations, the authors treat indirect quotations as well.

The great part of the effort on quotation extraction and attribution has been made for English texts [24, 26, 35, 38]. Still, several publications focus their work on other languages than English. In particular, quotation extraction for Portuguese [10, 39] and French has been studied [11, 52].

Sarmento and Nunes [10] present a system that handles Portuguese news articles. It finds direct and indirect quotations by applying 19 patterns and by exploiting a list of 35 reporting verbs. The system does not implement anaphora resolution for pronouns or noun phrases and therefore detects only speakers referenced by their proper name. The authors evaluated their approach manually on 570 quotations extracted by system.

[9] https://gate.ac.uk/.

De La Clergerie et al. [11] present an approach to quotation extraction from French news articles. Their rule-based approach includes a comprehensive linguistic processing chain with a deep parser. A postprocessing component constructs direct and mixed quotations based on parsing results, 230 quotation verbs, and direct speech parts signaled by quotation marks that were retrieved in previous processing steps. As with [11], the authors in [52] focus their quotation extraction approach on French news articles. Again, a rule-based approach is driven that exploits an automatically created lexicon of reporting verbs. The authors recognize 16 patterns matching indirect quotations and implement them as an unlexicalized grammar using an finite state machine.

Besides the rule-based systems [1, 10, 11, 24, 26, 41, 45, 52] a range of supervised approaches has been presented for the task of quotation extraction [35, 38, 39]. Fernandes et al. [39] propose a supervised solution using an Entropy Guided Transformation Learning (ETL) algorithm. They automatically generate rules instead of manually designing them. The work regards quotation extraction as a two-task problem. First, their system identifies quotations and, second, the quotations are associated with a speaker. Recognized named entities and the output of a co-reference component serve as a basis for the speaker assignment. To solve the subtasks different sets of features (named entities, terms, co-references, part-of-speech tags, etc.) are applied to the ETL algorithm. The developed system is capable of extracting direct and mixed quotations from Portuguese news articles. In order to train their system, the authors create the GLOBOQUOTES corpus.

The approach to quotation extraction from English texts proposed by O'Keefe et al. [35] makes use of supervised techniques as well. The authors solve the quotation extraction part by using a regular expression looking for text between quotation marks. Regarding quote attribution, which means finding the speaker of a quote, they cast the problem to a sequence labeling task. Inspired by Elson and McKeown [13], the authors encode news articles by replacing specific terms with symbols and by removing unnecessary information. Then, a set of features is calculated which includes distance, paragraph, nearby, quote, and sequence features, again following Elson and McKeown [13]. In order to efficiently predict the target speaker from a list of candidate speakers, the authors compare different types of class models and sequence decoding. They examine the effects of creating feature sets with and without gold standard labels. They conduct their experiments on three different datasets and find that when leaving out gold labels for feature calculation the performance drops significantly for classic literature but remains comparable regarding news articles.

Pareti et al. [38] focus their work on the extraction of indirect and mixed quotations from English-language news articles. The authors explore two supervised algorithms, namely a Conditional Random Fields (CRF) and a Maximum Entropy (ME) classifier. The token-based CRF classifier predicts IOB labels (I-inside, O-outside or B-beginning), marking the beginning and the end of a quotation, whereas the ME classifier decides whether a phrase-structure parse node is or is not a quotation. The classifiers largely rely on the same features but also incorporate classifier-dependent features. Instead of using a predefined list of reporting verbs, the authors train a k-nearest neighbor (k-NN) classifier working with 20 feature types, that predicts whether an identified verb group introduces a quotation or not. The authors conclude that the token-based approach using a CRF classifier outperforms the rule-based baseline as well as the constituent-based approach using the EM classifier for all quotation types. Regarding quotation attribution, the authors transfer four methods described in O'Keefe et al. [35] for direct quotation and find that they all are suitable for indirect and mixed quotations.

There is few scientific work that aims the extraction of quotations exclusively from German news articles. To the best of our knowledge only Pouliquen et al. and Akbik and Schenck [1] deal with German language news articles. While Pouliquen et al. include German into their multilingual system as one of many languages, Akbik and Schenck present a system that automatically collects news from the main German news sites and then extracts direct quotations from these news articles. Their approach detects text between quotation marks as quote candidates and uses a named entity recognizer to identify potential speakers. Then a set of heuristics is used to determine the resulting quote-speaker tuples.

1.3.3 Approach

The proposed approach for extracting direct and reported speech from German news articles is rule-based. For each quotation the system identifies a speaker, a reporting verb, or a preparative phrase (like "…, *so* Angela Merkel".), and the quotation text with all its parts. We divide the task into five subtasks and model our quotation extraction approach as a processing pipeline where the news articles are annotated in each step of the pipeline with further information. Figure 1.3 demonstrates the included components and the working flow. Starting with a document preprocessing component we perform linguistic analysis like part-of-speech tagging and lemmatizing that serve as a basis for further processing steps. The normalization of quotation marks is important at this point as well. Detecting a reporting verb helps to identify the reporting clauses and is also a strong indicator of indirect speech. We therefore search for them in the next step of our pipeline. Subsequently, we identify the reporting clauses of direct and indirect quotations and determine the quotation parts and exact boundaries of the entire quotation. Note that the boundaries of indirect or mixed quotations may be ambiguous and in many cases difficult to recognize even by humans. In the last step of our pipeline we attribute one or more quotation holders to the previously identified quotations.

1.3.3.1 Document Preprocessing

Quotation marks normalization. News articles may contain malformed markup. Especially in systems with automatic news harvesting from heterogeneous sources the collected texts may be erroneous, e.g., in terms of incomplete articles, misplaced

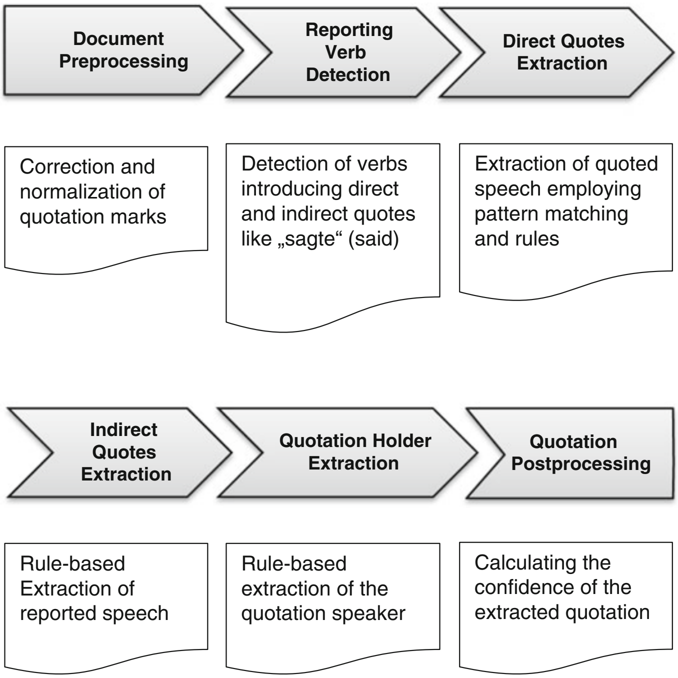


Fig. 1.3 The components of the quotation extraction pipeline

meta information or missing quotation marks. Regarding quotation extraction, texts released by different publishers may also contain varying styles of quotation marks like ≪ ≫, “ ” or ‘ ’. Since quotation marks are crucial indicators for both direct speech and quotation boundaries, the correction and normalization of quotation marks is an important subtask in quotation extraction systems. In our system a document preprocessing component corrects errors arising from inconsistent quotation marks. It first replaces all quotation marks with uniform quotation marks and then counts the number of quotation marks. The component does not patch texts with an odd number of quotation marks, but adjusts quotations with different start and ending quotation marks like quotations starting e.g., with “and ending with ‘.

Sentence Detection. Quotations may consist of several sentences or sentence parts. For example, the quotation *“Wir sind noch immer hier. Wir kämpfen noch immer”, sagte Santorum. (“We are still here. We are still fighting”, Santorum said.)* is composed of two sentences. In such case the quotation extraction must recognize both parts and determine correct quotation boundaries. A sentence detection is therefore an essential preprocessing step. Furthermore, other linguistic algorithms used in news text analysis require sentences as a basis for their calculations. Our quotation extraction pipeline uses the Apache OpenNLP[10] Maximum Entropy sentence detec-

[10] https://opennlp.apache.org/.

tor. The sentence detector uses a predefined model trained on Tiger corpus data [6] for the German language together with a self-developed heuristic, i.e., we created a set of likely ('.', '?' and '!') and unlikely (Fr., Hr., Prof. …) ends of a sentence and check the output of the OpenNLP sentence detector. If a sentence chunk has an ending that should never be an ending we merge again sentence parts that were incorrectly split.

Lemmatization. Determining the lemma of a word is a necessary step for our lexicon-based reporting verb finder described in (Sect. 1.3.3.2). In German the lemma of a noun normally is the "nominative singular" and of a verb it is the "infinitive present active" form. In our quotation extraction pipeline we make use of a morphological lemmatizer that looks up the words of the news article in a lexicon. The lexicon[11] was generated with "Morphy",[12] a software tool for the morphological analysis of German text [25].

Part-of-Speech Tagging. Part-of-speech tagging is the task of predicting the grammatical category (noun, verb, adjective, …) of a word based on the word's definition and its surrounding context. In computer linguistics each word of a sentence is assigned a label from a predefined set of part-of-speech labels. For German often the "Stuttgart-Tübingen-Tagset" (STTS)[13] is used for labeling [44]. The proposed quotation extraction approach works with the Apache OpenNLP maximum entropy part-of-speech tagger. Together with the predefined model trained on the Tiger corpus the tagger predicts STTS labels for words of a given text.

Noun and Verb Chunking. The chunking component analyzes sentences and determines verb and noun groups. The groups then serve as input for the recognition of potential reporting verbs or quotation holders. For example, a speaker may be referenced as "die deutsche Bundeskanzlerin" (the German chancellor) or a reporting verb may be a compound of two words like "teilte mit" (informed). The phrases output by the chunker help to determine the correct boundaries. Our processing pipeline uses the Apache Open NLP maximum-entropy-based chunker. To recognize noun chunks we use an out-of-the box model distributed by Gunnar Aastrand Grimnes.[14] For the recognition of verb chunks we trained a model on the Tiger Corpus [6]. The Tiger Corpus contains 50,000 sentences in German taken from the "Frankfurter Rundschau" which are POS-tagged and annotated with syntactic structure.

Named Entity Recognition. When citing persons or organizations a pronoun, a noun phrase or the proper name of an entity can be used to reference the quotation speaker. State-of-the-art named entity recognizers mainly detect top-level entities like persons, organizations, and locations which may be a starting point for the detection of quotation holders. We integrated the Stanford named entity recognizer into our quotation extraction pipeline. The Stanford recognizer is implemented as a Conditional Random Field classifier [16]. We use a pre-trained model for German provided by [15] that labels tokens as person, organization, location, and miscel-

[11] http://www.danielnaber.de/morphologie/.

[12] http://www.wolfganglezius.de/doku.php?id=cl:morphy.

[13] http://www.ims.uni-stuttgart.de/forschung/ressourcen/lexika/TagSets/stts-table.html.

[14] http://gromgull.net/blog/category/machine-learning/nlp/.

Table 1.1 The list of reporting verbs used for the quotation extraction approach

German reporting verbs				
Sagen	Behaupten	Aussprechen	Abraten	Teilen
Meinen	Warnen	Erwähnen	Raten	Klären
Fragen	Betonen	Bejahen	Ausfragen	Aufklären
Denken	Loben	Ermahnen	Ausplaudern	Mitteilen
äußern	Zugeben	Beichten	Erklären	Begründen

laneous. Since the Stanford classifier sometimes misses some named entities we decided to augment the list of named entities returned by the Stanford classifier by the named entities identified by the part-of-speech tagger described above. The type of the entities recognized in this way is tagged as UNKNOWN since the tagger marks the entities without providing a type.

1.3.3.2 Reporting Verb Detection

The detection of reporting verbs, that is verbs introducing quotations, is especially important for the recognition of reported speech and a quotation holder. Our reporting verbs detection approach is lexicon-based. We manually assembled a list of 25 common reporting verbs. We started with a set of six seed reporting verbs and extended the set by adding synonyms from Wortschatz Leipzig.[15] The Wortschatz Leipzig also outputs a frequency class that reports the relation of a word's frequency to the most frequent word in the corpus. We pruned the list by removing rare words (high frequency class) and very ambiguous words. Table 1.1 gives an overview of the common German reporting verbs that the reporting verb detector uses in our quotation extraction approach. Analyzing a text the reporting verb detector checks for each word's lemma if it occurs in the list. The corresponding words are then treated as reporting verb candidates for the quotations to extract.

1.3.3.3 Direct Quotation Extraction

All quotations within quotation marks are regarded as direct quotes (quoted speech). The direct quote collector detects quotations employing pattern recognition and handcrafted rules. It first compiles a set of quotation candidates (text parts enclosed by quotation marks) and then applies the set of handcrafted rules to them to construct the final direct quote.

The applied pattern is composed of different combinations of left and right quotation marks which must enclose at least one character. In order to avoid the detection of single words or phrases that are emphasized with quotation marks, quotation candidates are discarded if they consist of less than four words. Furthermore, we check

[15] http://wortschatz.uni-leipzig.de/.

whether the quotation candidates contain a verb. Our investigations have shown that quoted phrases with less than four words and without a verb are in most cases simply highlighted text parts such as proper names. A direct quote may be composed of several quotation candidates. That is why the component examines each quotation candidate and decides whether it is the beginning of a new quotation or the part of a compound quotation. It searches the environment of the quotation candidate for incomplete sentences (sentences not ending with a *period*, *exclamation* or *question mark*) and reporting verbs. Incomplete preceding sentences are concatenated to the quotation candidate. If the preceding sentence has been completed the component checks whether it contains a reporting verb. Sentences with a reporting verb are concatenated to the direct quotation candidate, because our experiments have shown that these sentences often are reporting clauses that provide a quotation speaker. Sentences following a quotation candidate are processed in a similar way. If a sentence contains a reporting verb or is incomplete, it is concatenated to the quotation candidate. Subsequent sentence parts containing the word "so" are also attached. We cover in this way cases like " '…', so Angela Merkel". Quotation candidates are connected to each other if a quotation candidate directly succeeds a reporting clause or a quotation candidate.

1.3.3.4 Indirect Quotation Extraction

Reported speech is not put in quotation marks. It is composed of a main (reporting) and a subordinate (reported) clause. In German, the reported clause often is introduced by the conjunction "dass" and uses the subjunctive mood for verbs. In order to extract indirect quotations from a news article we apply a rule-based approach. The indirect quote extraction depends on the output of the direct quote collector. Therefore, the indirect quote collection must succeed the direct quote collection in our processing pipeline. Our approach is to first identify a reporting and a reported clause and then construct the final indirect quotation. The indirect quote collector exploits the occurrence of reporting verbs. To avoid duplicate quotation extraction (identifying quotations as direct and indirect) the collector exclusively regards reporting verbs that have not been already assigned to a direct quotation. If a detected reporting verb is not already part of a direct quotation, we assume that the verb indicates the reporting clause of an indirect quotation. We build up an indirect quotation by analyzing the surrounding sentences or sentence parts. A strong indicator for the reported clause is the presence of the conjunction "dass" (that) together with the finite verbs "sei, seien, habe, werde, würde, würden" that are usually used in reported clauses to repeat what someone has said. The occurrence of "dass" and one of the verbs implies a reported clause and we infer a quotation. The quotation encompasses the reporting and the reported clause. Sentences containing a reporting verb in the reporting clause but missing "dass" in the reported clause are treated in the same way, if they contain the finite verbs mentioned above. We also detect indirect quote indicated by ', so' (as) and ', hieß es' (it was said).

1.3.3.5 Quotation Holder Extraction

The aim of the quotation holder extraction is to attribute speakers to the identified direct and indirect quote. Our approach is based on the observation that quotation holders in most cases are named entities or references to named entities that are mentioned nearest the reporting verb. For example, we choose the pronoun 'er' (he) regarding the fragment ', sagte er dem Spiegel' (, he said to 'Der Spiegel'). To determine a quotation's holder we first create a set of candidates. As candidates we consider named entities, pronouns (only "er" (he) and "sie" (she)) and noun chunks from the reporting clause. We exclude candidates originating from the reported clause. Then, we sort the list by proximity to the reporting verb but prioritize named entities and pronouns over noun chunks. Pronouns are still left in order with named entities, so that passages like ", sagte *er* zu *Angela Merkel*" (*he* said to *Angela Merkel*) do not get assigned to the wrong holder. If no reporting verb has been assigned to the quotation we search for the word "so" in the reporting clause and sort the candidates according to how near they are placed to the word "so". Concerning direct quotations there also may be quotations without a reporting verb and the word "so", since they are detected with the aid of quotation marks. In this case we simply select the candidate nearest to the reported clause. Our approach to quotation holder extraction also includes a simple form of co-reference resolution. If we determine a person as quotation holder we attempt to resolve its name to the longest form of it in the text. If the assigned speaker is a pronoun then we choose the first named entity before the quotation.

1.3.4 Corpus

We manually annotated a corpus of 714 news articles containing direct and reported speech. The news articles are all in German and were published over a time period of three months from February 23, 2012 to May 21, 2012. The corpus allows the evaluation of determining quotation text boundaries and of recognizing reporting verbs and quotation holders.

For the annotation process we had to assure a sufficient coverage of direct and reported speech. That is why we preprocessed the news stream provided by Neofonie GmbH and preselected some news documents before we started with the annotation procedure. We automatically detected different types of quotation marks and a set of predefined reporting verbs within the news articles. Thereafter we randomly sampled 1,000 news articles. We chose:

- ***250 news articles*** containing at least one direct quotation (text passages identified by the occurrence of quotation marks and that are longer than 24 characters)
- ***250 news articles*** containing at least one of the following reporting verbs: "sagte", "berichtete", "berichteten", "gestand", "erklärte", "erklärten"
- ***500 news articles*** without any restrictions.

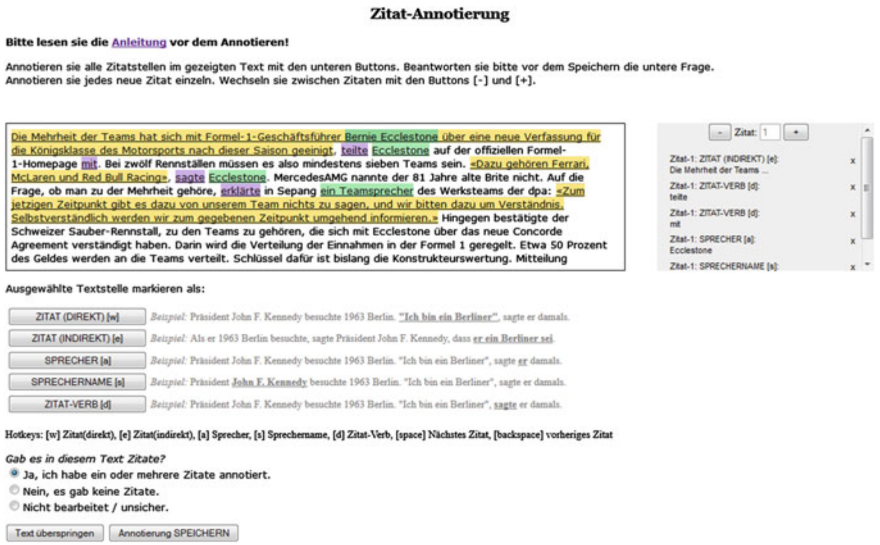


Fig. 1.4 The annotation tool used for the creation of the quotation extraction corpus

We asked the annotators to identify all quotations in a news article and advised them to mark for each quotation the quoted text, the quotation holder, and a reporting verb if available. A screenshot of the annotation tool is shown in Fig. 1.4. For quotation holders not referenced by their proper name but by, e.g., a personal pronoun or only by the last name, the annotators should assign the full proper name if possible. If a quotation or a reporting verb was composed of several parts, the annotators were asked to mark all parts (**teilte** der Sprecher **mit**, *the spokesman said*). They were also advised to mark if a news article does not contain any quotes at all.

We succeeded in annotating 714 news articles. 339 of the news articles were annotated twice, 27 three times, and 2 even four times. The remaining 347 news articles were annotated by only one annotator. The annotators exactly agreed upon the quotations in 287 news articles. At that point we speak of exact agreement if the boundaries of the quotation holder, the reporting verb, and the quote text match accurately comparing the annotated tokens. Finally, the resulting corpus of 287 news articles contains 383 quotations, whereof 256 quotations are direct, 98 indirect, and 29 mixed (including at least a direct and indirect part) quotations. A news article contains 1.3 quotations in average. 87 % of the quotations are attributed with a reporting verb. We succeeded in annotating a quotation holder for each quotation. For 202 quotation holders we could resolve the reference and assign proper names.

1.3.5 Evaluation

We evaluate our quotation extraction approach using a human-annotated corpus of 287 news articles where at least two annotators exactly agreed upon the contained

Table 1.2 Results for the quotation extraction

	Reporting clause						Reported clause		
	Holder			Verb					
	P	R	F1	P	R	F1	P	R	F1
All quotations	0.801	0.649	0.717	0.932	0.728	0.817	0.862	0.821	*0.841*
Direct quotations	0.791	0.672	**0.727**	0.914	0.679	0.779	0.89	0.895	**0.892**
Indirect quotations	0.852	0.596	0.701	0.989	0.815	**0.893**	0.747	0.782	0.764
Mixed quotations	0.727	0.653	0.688	0.852	0.767	0.807	0.913	0.505	0.65

The direct quotation extraction performs best. Our component achieves an F1-score of 0.89 for the extraction of the reported clause and an F1-score of 0.73 for assigning a speaker. The extraction of indirect quotations is more difficult. Still, achieving an F1-score of 0.76 for the extraction of the reported clause, it produces reasonable results

quotations. In order to measure the performance of our approach we make use of standard information retrieval measures and compute a token-based recall, precision and F1-score. We regard all overlapping tokens as true positives. All missing tokens are regarded as false negatives and all unnecessarily annotated tokens as false positives. We then calculate the overall performance by summing up all intermediate results and by calculating final micro-averaged results for the subtasks of holder, verb, and reported clause extraction.

Table 1.2 summarizes the obtained results. The results meet our expectations. We achieve the best micro-averaged F1-score of 0.89 for the extraction of direct quotation parts. With an F1-score of 0.76 our approach for extracting reported speech performs less effective but still reasonable. Considering quotation holders, the proposed algorithm behaves comparably for all quotation types. It achieves an F1-score of 0.72. The detection of reporting verbs or clauses introducing a quotation performs quite well with an F1-score of 0.82. It is striking that the extraction of reported clauses of mixed quotations is most challenging. Here, our algorithm does not exceed an F1-score of 0.65.

In order to facilitate a manual evaluation of our extraction approach and also for its further refinement and improvement, we developed a web demonstrator that visualizes the results calculated by our component (Fig. 1.5). The upper field allows to insert an arbitrary text[16] which our quotation component analyzes subsequently. The demonstrator shows the identified quotations in the preview window below. The detected text spans are highlighted in specific colors. At a glance the user sees whether the extraction was successful or whether the algorithm provides erroneous annotations.

[16] Note that our approach is calibrated on news articles and could produce insufficient markup for text types other than news articles.

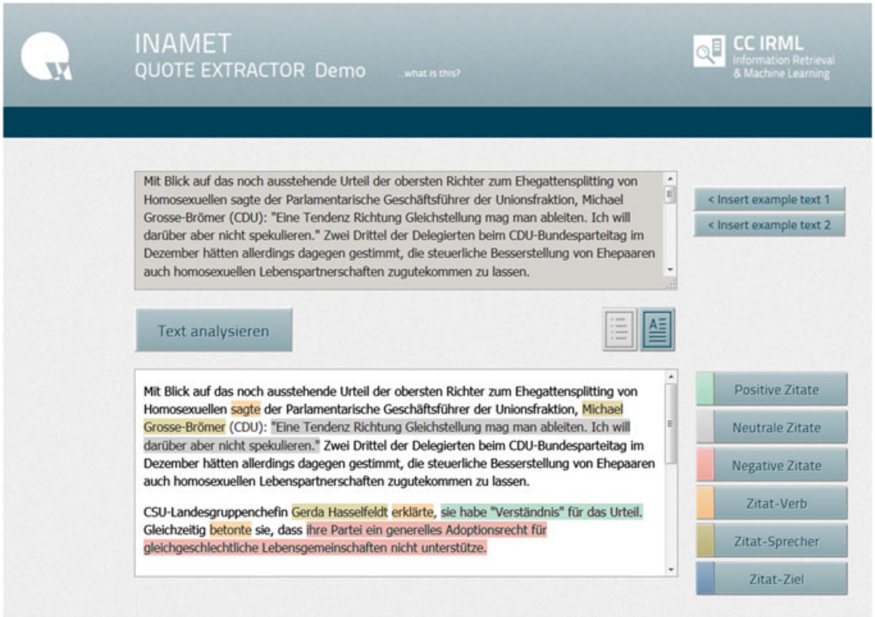


Fig. 1.5 The web demonstrator for visualizing automatically extracted quotations. It allows to insert arbitrary text into the *upper* field which is then analyzed by our quotation extraction component. The results are highlighted in the area *below*. The users can choose which information the demonstrator should display and which should be hidden

1.3.6 Conclusion

We presented an approach to quotation extraction that includes the extraction of direct and indirect quotations and the assignment of a speaker to each quotation. Our approach is rule-based and relies on a handcrafted list of reporting verbs. The implemented rules are manually created as well. As valid speakers we allow text spans covering pronouns (she and he), noun phrases, and named entities. We resolve pronouns to appropriate named entities mentioned earlier in the text. The results achieved with our unsupervised approach compare favorably with other approaches, and do not rely on the availability of training data. Especially the extraction of direct quotations and attributing them to a speaker already works very satisfactorily. Regarding indirect quotations, finding the boundaries of the reported clause and the correct quotation holder is more challenging, regarding mixed quotations even more. Since our approach to indirect quotation extraction relies on a list of reporting verbs and clues, potential results are limited to those parts near such predefined reporting verbs or clues. Thus, our future work includes among others, the extension of our approach by an automatic reporting verb recognizer [38]. We plan to detect previously unseen reporting verbs as well as the disambiguation of verbs. For example, the ambiguous verb "to add" may lead to a mistake by regarding it as a reporting speech indicator in a wrong context. By automatically determining reporting verbs and phrases we aim to improve the recall of indirect quotations. We also intend to complement our co-reference resolution approach with a state-of-the-art component. We want to determine co-reference chains and then choose the correct one as quotation speaker [39]. In order to consolidate quotations uttered by one speaker across different news documents, we plan to link speakers to Wikipedia entries by applying a named entity disambiguation approach. Our future work will also include the incorporation of supervised approaches. On the one hand we plan to treat quotation extraction as a sequence labeling task and on the other hand we plan to train a binary classifier that predicts quotations at sentence level. Our goal is to identify appropriate features for German-language texts and to let an ensemble combine the output of the rule-based approach with the output of the new classifiers.

1.4 Sentiment Analysis

Publicly available texts such as product reviews, social media contributions, or news articles discuss almost every thinkable entity and topic. Besides transporting facts, the texts often cover opinions as well, and even more than facts, the expressed opinions may influence readers. Regardless of whether someone wants to buy a new camera or wants to find out which party to vote in the next election, people in general read first what other people think and what experiences they have had before making their own decisions. That is the reason why companies are interested in a positive perception of their products and services in the media. Here, well-analyzed opinionated texts may serve as a basis for a multitude of sentiment-related applications like reputation monitoring or opinion summarization systems. Sentiment analysis may also be performed to improve other natural language processing tasks that rely on factual data. Separating opinionated text from objective text turned out to be beneficial for information extraction [43]. In this section we show how opinions can be extracted from news articles. We focus our work on news articles because they are a reliable information source and mirror the opinions of newsworthy people, which often serve as role models. We first introduce the term "sentiment analysis" and "opinion mining" and then present our comprehension of "opinions". The main part describes our supervised approach to sentiment analysis. We limit our approach to quotations because we assume that quotations are the most subjective parts of news articles.

1.4.1 Introduction

Sentiment analysis aims at identifying subjective language in texts and determining the orientation and strength of expressed opinions or sentiments toward the corresponding targets. Often the term "opinion mining" denotes the same task and is used synonymously [36]. The task of sentiment analysis may be decomposed in subjectivity detection and polarity (or orientation) classification. The goal of subjectivity detection is to distinguish objective from opinion-oriented text. While objective parts solely report facts without any personal assessment or evaluation, opinionated text parts may contain any type of subjective expressions that reflect the private state of a holder. This includes personal attitudes, views, statements, feelings, etc., expressed by the text's author or by other people mentioned or cited in the text. Having detected a subjective text part, the type and orientation of the expressed opinion must be determined. The most common orientations are *positive* and *negative*. For example, a literature critic may conclude that a reviewed book is excellent, which can be regarded as a positive opinion toward the book. Other classification schemes distinguish between supporting and opposing expressions. This variant of sentiment classification allows to contrast different points of view toward topics or political issues.

Our definition of opinion is driven by Liu and Zhang [28]. Following the authors, we define opinion as a quintuple consisting of a

- **target entity** (e.g., a product, an individual, a topic)
- **target aspect of the entity** (e.g., product features, subtopic)
- ***orientation*** (e.g., positive/negative/neutral)
- **holder** (the entity holding the opinion)
- **time** (when the opinion was expressed)

In order to gain a valuable opinion information, not all parts of the quintuple must necessarily be extracted. The perception of a movie in the media, e.g., can be inferred without knowing when the review was submitted or by whom. However, the extraction of a target entity and the valuation of the entity is essential. The opinion target may be a mentioned (named) entity or concept with a concrete text-anchor, but also an abstract topic that makes the identification of the opinion target especially difficult. In contrast to concrete entities, abstract topics are nonlocal information that have to be mined from the context. In our scenario, which is sentiment analysis based on newspaper quotations, the structure of quotations already provides an opinion holder. Knowing who uttered a quotation, we consider the quotation holder being also the opinion holder.

Our Contribution. We present work that aims at determining opinion orientation in German news articles. In contrast to many other sentiment analysis systems we focus our approach on direct and reported speech identified previously in the news articles. We assume that quotations are the most subjective parts of news articles and that they transport the opinions of the cited speakers as they are. We cast the task of sentiment classification to a three-class problem and label each quotation as either negative, positive, or neutral. Objective quotations reflecting facts are marked as neutral as well. We propose a supervised approach where we first search for subjective quotations and, second, decide the polarity of the subjective quotations. Both steps are separately solved by a Support Vector Machine classifier. As part of our work we examine the effectiveness of diverse sentiment classification features to find the most suitable feature set for both the subjectivity detection and polarity classification task. We train and evaluate our approach on a human-annotated corpus of German quotations. The corpus consists of 742 neutral, 71 positive, and 38 negative quotations. It can be made available for research purposes after signing an agreement.

1.4.2 Related Work

Related research on sentiment analysis varies from simple lexicon-based approaches looking up words in opinion lexicons to supervised approaches exploiting linguistic features and enhanced machine learning algorithms. The main part concentrates on the classification of text as either POSITIVE, NEGATIVE or NEUTRAL toward a specific entity or topic explicitly mentioned in the text.

Sentiment analysis treats texts at different levels. There is work examining entire document texts like entire reviews or news articles, attempting to predict the overall sentiment of a document text [37, 49]. But there is also work that performs sentiment analysis at statement level [5, 18, 46], sentence level [23, 33, 43, 54] or even phrase level [49]. Often, sentiment analysis work on reviews also aims at extracting product properties and the opinions toward these properties, which is called aspect-oriented sentiment analysis [21].

Sentiment analysis has also been applied to different text types. A great part of the work examines customer reviews, like product [21, 49] or movie [37] reviews. Since reviews are meant to share experiences and report opinions, they contain many subjective text parts and are therefore predestined for sentiment analysis. Yet, reviews can also contain objective parts summarizing the properties of the reviewed entities. Regarding movie reviews, one challenge is to separate plot information, which itself may be characterized as positive or negative, from opinions toward the movie. All work treating customer reviews must handle challenges arising from user-generated content such as potential spelling mistakes and grammatical errors.

Early work in classifying product reviews used lexicon-based techniques together with natural language processing algorithms in order to create opinion summarization. Hu and Liu [21] propose a three-stage approach to aspect-based opinion summarization. They first search for product features in customer reviews by applying association mining with some pruning. Then, the authors determine the polarity of sentences mentioning the features. Whether a sentence has to be classified as positive or negative results from the orientation of the individual opinions words (adjectives) in the sentence that is summed up to an overall orientation. The orientation of opinion words is pre-calculated based on a list of seed adjectives and the application of WordNet's information on synonyms and antonyms. Similar to Hu and Liu, Turney [49] categorizes product reviews in either 'recommended' or not 'recommended' by calculating the average sentiment orientation of the review's phrases. Turney calculates the orientation of phrases containing adjectives and adverbs by determining the mutual information between a phrase and the words "excellent" and "poor" and subtracting both values to obtain a final sentiment orientation score.

The so far discussed customer reviews are predominantly medium or long texts. With the mass distribution and utilization of social media services like Twitter or Facebook in the recent years, a part of the sentiment analysis work shifts toward the analysis of short texts generated by users. Because of the language used in such texts, new challenges arise for the task. Often, users write their texts colloquially and they do not care about spelling and punctuation. In addition, the texts mostly are very short and comprise phrases rather than complete sentences. Considering Twitter, a short message must not exceed 140 characters.

As one of the first, Go et al. [18] classify English-language tweets according to a query as either positive or negative. They adopt a supervised approach using diverse classifiers including a Naive Bayes, a Maximum Entropy, and a Support Vector Machine (SVM) classifier. In order to train the classifiers, the authors propose using tweets containing positive or negative emoticons (mapped to ':(' and ':)') as noisy labeled training data. The authors explore a range of standard text classification features like word uni- and bigrams and part-of-speech tags for representing the tweets. After having evaluated their approach on manually tagged tweets from different categories (177 negative and 182 positive tweets independent of emoticons), Go et al. conclude that the automatically created training dataset is suitable for training the examined algorithms, which solve the task reasonably. Using a combination of word uni- and bigrams the Maximum Entropy classifier achieves an accuracy of 83 %. Yet, there are no large differences between the classifiers and feature sets.

In comparison to customer reviews, news articles may express opinions less explicitly. Since journalists (ought to) write objectively and avoid emotional language, the identification of the implied opinions is especially challenging. In addition, the opinion holder must be extracted. Different from customer reviews, it is not the author's opinion expressed in the news article but the opinion of other people and organizations the article deals with. In 2006 Kim and Hovy [23] approached the task of opinion mining in English news articles by proposing a four-stage system. The authors extract opinions, determine the opinion topic, and assign an opinion holder by applying semantic role labeling. The authors separate subjective from objective sentences, perform semantic role labeling utilizing opinion-related frames and frame elements from FrameNet,[17] and choose the opinion target and holder out of the semantic roles. Finally, the extracted opinion triples consisting of the holder, topic, and opinion are stored in a database.

The work proposed by Nakagawa et. al [33] addresses sentiment classification at sentence level. The authors use conditional random fields with hidden variables, representing polarity of dependency sub-trees, to infer the polarity of the entire subjective sentences. The approach was evaluated on English and Japanese opinion texts and is promising. Among others, it was evaluated on Japanese news articles with an accuracy of 83 %, which shows its effectiveness on this text type. However, the work bases on subjective sentences and skips the task of subjectivity detection.

Strongly related to our work is the work of Balahur et al. [2–4]. The authors apply sentiment analysis to news articles. Although the team mainly explores approaches

[17] https://framenet.icsi.berkeley.edu/fndrupal/.

to classify English news quotations, it is an important requirement to develop approaches that are applicable with as little effort to other languages, since the authors incorporate their results into the Europe Media Monitor (EMM) news engine.[18] EMM collects and processes news articles from multilingual news sources. In [2] Balahur and Steinberger present reflections on how sentiment analysis applied to news articles differs from sentiment analysis on highly subjective texts like online reviews. The authors find that the task of sentiment analysis on news articles can be decomposed into three subtasks: determining the sentiment target, distinguishing between "good and bad news content" and "good and bad sentiment expressed on the target", and classifying explicitly expressed sentiments on the sentiment target that does not require any world knowledge or the interpretation by the reader. The inter-annotator agreement increases if the task is clearly defined in advance. Finally, the authors work out three possible perspectives on news articles: the author's (news bias research), the reader's (interpretation by readers influenced by their backgrounds), and the text's view. Each view requires a different approach to sentiment analysis and the authors limit their work to identifying sentiments concretely expressed in the text. In [3] Balahur et al. provide a comparison of different sentiment resources and classification strategies to categorize news quotations as positive, negative, and neutral. In their studies, the authors also examine how a preceding subjectivity detection step affects the classification results. They conclude that using large sentiment lexicon and a previous subjectivity filtering improves the results considering vocabulary-based methods. Straightforward bag-of-words approaches are limited and not effective enough for sentiment analysis on news quotations. The authors also conclude that exploiting sentiment annotations based on single topics are not suitable for the open-domain sentiment analysis on news. Thus, they propose a topic-dependent sentiment analysis with specialized models.

The work of Balahur et. al in [4] analyzes two aspects of sentiment analysis for English-language quotations in news articles. First, the authors examine how different word windows around an opinion target influence sentiment classification accuracy. Second, they exclude sentiment-bearing words that are category-specific words at the same time, in order to separate good or bad news content from positive and negative sentiments toward the opinion targets. The sentiment score is calculated by summing up the sentiment scores of all quotation words. As a result, the authors argue that taking into account only a word window around the target entity instead of including the entire quotation text yields better results. Considering the lexicons, the authors find that there are large differences between their performances and that combining them helps. However, the accuracy of the approach does not only depend on a large lexicon.

Sentiment analysis on German-language texts has been applied in [27, 31, 46]. Momtazi presents a rule-based approach to classify sentiments toward celebrities mentioned in short German social media texts. In order to label the short texts as positive or negative and assign the strength of the sentiment, the author creates and applies a sentiment dictionary and a list of booster and negation words. Mom-

[18] http://emm.newsbrief.eu/overview.html.

tazi evaluates her approach on a hand-annotated dataset of 500 short texts about celebrities. Her unsupervised polarity classification method outperforms standard supervised classifiers on the given dataset. The main contribution of Li et al. [27] is an annotation scheme for labeling sentiments in German political news articles and a dataset manually annotated according to the presented scheme. Each annotation frame consists of the text anchor (labeled as idiom, phrase, word or compound noun), the target, the source, and auxiliary words that may be intensifiers, diminishers, or negations. In addition, the opinion frames may be marked with the attitude's polarity, type (context-dependent or -independent), and intensity. With the aid of the relation extraction tool DARE[19] and the annotated relation examples, rules are automatically learned to extract the opinion source, target, and polarity. In their first experiments the authors achieved promising results. Another corpus for German sentiment analysis on news articles is contributed by Scholz et al. [46]. The corpus consists of around 1,500 statements labeled with the viewpoint (corresponding to our "opinion target"), either CDU[20] or SPD,[21] and the tonality of the statement (positive, neutral, negative). The authors use parts of the dataset to generate sentiment dictionaries containing entries scored with different measures. For the sake of media response analysis, the authors evaluate news material and propose a supervised machine learning approach which is similar to ours. Scholz et al. analyze the news in two stages. First, they detect subjective statements and, second, they classify the subjective statements as either positive or negative. In contrast to our findings, the subjectivity detection seems to perform better on the dataset of Scholz et. al. than the polarity classification part.

Detailed surveys on opinion mining and sentiment analysis can be found in [28, 36, 48].

1.4.3 Sentiment Classification

We solve the problem of quotation sentiment classification by employing a supervised two-stage approach. We first apply a subjectivity detection step where we mark quotations as either neutral or subjective. We then classify all subjective quotations according to their polarity in either positive or negative quotations. As a result of our sentiment classification approach each processed quotation is labeled as either neutral, positive, or negative. For both tasks, subjectivity and polarity classification, we train separate Support Vector Machine (SVM) classifiers [8, 50] with a different feature set and with different hyperparameters. We choose a radial basis kernel for both SVMs and select the hyperparameters γ and C by performing tenfold cross-validation on the dataset described in Sect. 1.4.5. We represent the quotations as vectors of diverse features (Sect. 1.4.4). Among others we include the part-of-speech tags and sentiment words as features that turned out to be essential for sentiment

[19] http://dare.dfki.de/.

[20] Christlich Demokratische Union Deutschlands (Christian Democratic Union (Germany)).

[21] Sozialdemokratische Partei Deutschlands (Social Democratic Party of Germany).

analysis. We weight the features by different weighting schemes ranging from simple counts to enhanced weighting schemes like tf-idf and tf-delta-idf, a sentiment-based tf-idf value, as proposed in [30].

Opinion Target Extraction. We also provide a supervised approach for the extraction of opinion targets, which may be reasonably applicable as long as the targets are explicitly mentioned within the quotations and can be localized by text anchors. In our target extraction approach we first select a set of candidates and then classify each candidate with a binary classifier to predict whether it is the wanted target or not. The opinion target candidates are represented as feature vectors of POS tags surrounding each candidate in a window of two words before and two words after the candidate. We perform a multistage decision process to prefer specific candidates over other candidates if more than one candidate was classified as opinion target. We check the environment nearby the candidate and accept only candidates conforming specific predefined POS patterns.

1.4.4 Sentiment Features

Finding an appropriate representation of the data at hand is a crucial task since the performance not only depends on the chosen machine learning algorithms but also to a large extent on the selected features [12]. In this section we present the features explored for our sentiment analysis approach. Besides primitive features we also exploit derived lexical and linguistic features. Following [37] we include position information for each feature. We encode whether the features were calculated based on the beginning, the end, or the middle part of the text or whether the entire text was considered.

Bag-of-Words. A standard representation of documents for natural language processing tasks is the bag-of-words model [29]. It represents a document as a vector of weighted terms from a dictionary. We built up a dictionary with uni- and bigrams and calculated the idf and delta-idf based on German news articles from a time period of three months of 2012, the same time period as used for creating the evaluation corpus described in Sect. 1.4.5. The lexicons with both uni- and bigrams were limited to 10,000 entries each. We included bigrams because they encode word order, which adds meaningful sentence structure information to the feature vector representation. Previous work shows that bigrams help in the task of sentiment analysis [51]. In order to compile a feature vector for a document we remove stop words and lowercase and stem each term using Apache's German Analyzer.[22] Then we weight each term (uni- and bigram) using one of four different schemes: occurrence flag (0/1), tf (term frequency), tf-idf (term frequency x inverse document frequency), tf-delta-idf (term frequency x delta inverse document frequency).

[22] https://lucene.apache.org/core/3_6_2/api/all/org/apache/lucene/analysis/de/GermanAnalyzer.html.

Parts-of-Speech. Part-of-speech tags (POS tags) serve as the basis for various natural language processing tasks. Especially, in sentiment analysis the part of speech information is widely exploited. The presence of certain parts of speech like adjectives [19] and word phrases corresponding to certain part-of-speech patterns [49] correlates with opinion-oriented language and is therefore a strong predictor of subjectivity [36]. We assembled a dictionary of POS tag uni- and bigrams and calculated the idf and delta-idf analogously to the dictionary of term uni-and bigrams. Based on this dictionary we form the feature vector as a bag-of-pos-tags. Again, we assign the following values to each items: occurrence flag, tf, tf-idf, tf-delta-idf.

Sentiment Words. Sentiment lexicons like SentiWordNet [14], WordNet-Affect [47] or MicroWNOp [7] often are a central resource for many sentiment analysis approaches [3]. Sentiment bearing words have been applied for lexicon-based solutions [4] and as features for supervised classification approaches. For German the SentimentWortschatz (SentiWS) has been well established [42]. SentiWS is a lexicon covering affective words along with weights for their polarity that range between -1 (very negative) and $+1$ (very positive). In addition to the sentiment value, the part of speech and a set of inflections is available for each entry. SentiWS contains 1,650 positive and 1,818 negative adjectives, adverbs, nouns, and verbs. We use SentiWS for compiling term and aggregated features. First, we represent a quotation as a vector of SentiWS terms with either an occurrence flag, the term frequency, or the term frequency multiplied with the polarity weight of the term in SentiWS. Second, we aggregate features by grouping positive, negative, or all SentiWS words and count and weight their occurrences. Given quotations like *"Das ist **nicht gut**", sagte Angela Merkel.* ("That's **not good**", Angela Merkel said.) we must consider negation to avoid erroneous feature values. Simply regarding SentiWS word's polarity value as it is, we would assign a positive value to the term 'gut' and label the quotation probably as POSITIVE although a negative sentiment is expressed. For a more precise feature calculation we again make use of POS information and identify five POS tag patterns starting with PTKNEG (negation particle 'nicht' (not)) or PIAT (attributive indefinite pronoun without determiner 'kein' (no)), both inverting the polarity value of the following word. Table 1.3 shows the set of five patterns capturing phrases commonly used in German for negating adjectives, verbs, and nouns. We apply the

Table 1.3 The list of POS patterns used for shifting polarity weights of quotation terms

POS tag pattern	Examples
PTKNEG ADJD	Nicht richtig (not correct), nichtüberzeugend (not convincing)
PIAT NN	Kein Gegner (no opponent), kein problem (no problem)
PTKNEG VVPP	Nicht gelungen (not succeeded), nicht vorbestraft (not previously convicted)
PTKNEG VVINF	Nicht tolerieren (not tolerating), nicht bessern (not becoming better)
PIAT ADJA	Kein gutes (no good), kein unzumutbares (no unacceptable)

The examples are taken from our evaluation corpus described in Sect. 1.4.5

patterns to the quotation's text and augment our feature vector by adding a negated form of the SentiWS term in form of NOT_SENTIWS_TERM to the vector. We also invert the SentiWS value for the negated term, if the weighting scheme requires this. In the example above we add NOT_GUT (not good) to our term vector and regard the term as negative for calculating our aggregated features.

Valence Shifters. Valence shifters are words or phrases like "nicht" (not), "höchst" (extremely), or "weniger" (less), that change the intensity or polarity of other lexical items. We distinguish between three types of valence shifters: negations, diminishers and intensifiers. Previous work examined the effect of valence shifters in sentiment classification of movie reviews and concluded that incorporating valence shifters slightly increases classification accuracy [22]. In order to create our features, we exploit a list of around 100 valence shifters derived from the MLSA Corpus, a multi-layered reference corpus for German-language sentiment analysis [9]. The corpus consists of three layers with sentiment annotations at different granularity levels. Layer 2 provides polarity related annotations for words and phrases. At phrase-level the text spans are labeled as positive, negative, bipolar, and neutral. Words are labeled in addition as diminishers, intensifiers, and shifters (negations). With the aid of these annotations we compile feature vectors in the form of bag-of-valence-shifters and derived features that accumulate the three types of valence shifters.

Discourse Markers. Discourse markers are words or phrases that connect sentences or sentence parts and thereby express the semantic relations between them. Examples are "weil, aber, abgesehen davon dass, sogar, dennoch…" (because, but, apart from this, even, however…). The usage of discourse markers may influence the orientation or intensity of sentiments like in the quotation "*Wir sind zufrieden mit dem Stand der Dinge, **aber** wir wollen mehr*", *sagte Vettel.* ("We are happy with the situation, but we want more", Vettel said.) In our approach we search quotations for discourse markers from a predefined list. The list is derived from the online lexicon for German grammar[23] of the "Institut für Deutsche Sprache"[24](IDS, Institute for German Language). It contains around 350 discourse markers of different types. The resulting feature vector encompasses all discourse markers making no distinction between the types. We assign each marker a value (occurrence flag and term frequency) and encode whether the quotation contains discourse markers and how many.

All Features. Table 1.4 provides an overview of all feature groups that we use in our sentiment analysis approach. A feature vector containing all feature combinations consists of 160 K entries. If we include text position information into the feature vector the number of entries rises to almost 650 K. The relative big dictionaries, in comparison to the short quotations, result in very sparse feature vectors that we have to deal with.

[23] http://hypermedia.ids-mannheim.de/index.html.

[24] http://www1.ids-mannheim.de/start/.

Table 1.4 Overview of the sentiment analysis features and values

Feature Name	Feature Value
Bow-of-Words (uni- and bigrams)	Occurrence flag, tf, tf-idf, tf-delta-idf
Parts-of-Speech (uni- and bigrams)	Occurrence flag, tf, tf-idf, tf-delta-idf
SentiWS words	Occurrence flag, tf, tf x polarity-value
SentiWS pos/neg/all	Occurrence, count, count x polarity-value
Valence shifters	Occurrence flag, tf
Valence diminishers/intensifiers/shifters	Occurrence, count
Discourse markers	Occurrence flag, tf

1.4.5 Corpus

The corpus for evaluating our sentiment analysis approach consists of 851 quotations extracted from a dataset of German news articles dated from February 23, 2012 to May 21, 2012. The manually annotated quotations of the quotation extraction corpus described in Sect. 1.3.4 served as the basis for the sentiment corpus. We asked four annotators to tag each quotation as either NEUTRAL, POSITIVE, NEGATIVE, or MIXED and if possible to mark the opinion target.[25] In order to obtain a consistent corpus we defined a set of annotation rules, which we describe in detail below. In general, personal attitudes of the annotators, their moral perceptions, or political views must not influence the tagging. The task was to determine the opinion of the quotation speakers. If a quotation appeared to be incomplete, the annotators also had the possibility to tag a quotation as DON'T KNOW.

1.4.5.1 Neutral Quotations

A quotation should be regarded as neutral if the statement serves solely to transport facts. Information, announcements, or intentions of the speaker without any personal assessment by the speaker are considered to be neutral. It is not relevant whether the fact itself is positive or negative from a moral, political, or any other point of view. In the remainder of this section, we provide examples of different types of quotations. The content of the quotation in Example 1.1 solely reports fact:

> *Example 1.1* Alisade berichtete, dass bald 500 weitere KFC-Filialen landesweit eröffnen würden. (Alisade reported, that soon 500 more KFC-stores would open countrywide.)

[25] We do not use the annotated opinion targets yet, but describe them here for the sake of completeness.

The speaker announces an event without any personal assessment, so that the quotations in Example 1.2 has to be marked as neutral as well:

Example 1.2 Es werde Sicherheitskontrollen an den Einlässen geben, sagte ein Sprecher. (There will be security checks at the entries, a speaker said.)

1.4.5.2 Subjective Quotations

Subjective quotations could be marked as POSITIVE, NEGATIVE, or MIXED. To make their decision the annotators should answer the question whether the speaker supports or dislikes the topic or the expressed intention or announcement. In the following quotation the speaker explicitly expresses a negative opinion by using the term "Unverschämtheit" (impertinence). Annotators should rate it as NEGATIVE:

Example 1.3 "Der Fakt alleine ist eine absolute Unverschämtheit gegenüber dem Klub und dem Team", sagte Horstmann. ("The fact alone is an absolute impertinence toward the club and the team", Horstmann said.)

The phrase "verdient gewonnen" (corresponds to "deserved to win") indicates a positive opinion of the speaker in the following quotation so that this quotation should be classified as POSITIVE:

Example 1.4 Claus Finger war zufrieden: "Das Team hat schnell ins Spiel gefunden und verdient gewonnen." (Claus Finger was happy: "The team quickly got into the game and deserved to win".)

1.4.5.3 Corpus Overview

The final corpus exclusively contains quotations annotated by at least two annotators. The gold standard answers were determined by majority voting. We discarded quotations where the annotators predominantly disagreed or where the majority of the annotators marked a quotation as DON'T KNOW. The inter-annotator agreement amounts to 79 %. Figure 1.6 shows the distribution of NEUTRAL, POSITIVE, NEGATIVE, and MIXED quotations. The majority of the quotations, namely 86 %, are NEUTRAL, whereas only 8 % of the quotation are POSITIVE and only 5 % NEGATIVE. Thus, we

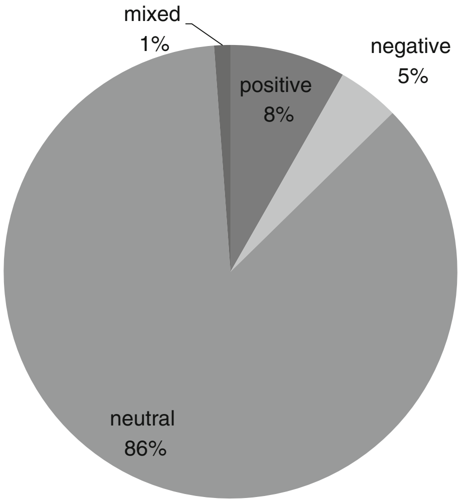


Fig. 1.6 The corpus for evaluating the sentiment analysis approach is highly unbalanced. It consists of 742 neutral, 71 positive, 38 negative, and 10 mixed quotations

have to deal with a highly unbalanced corpus. We discard the 1 % MIXED quotations, because we do not aim at the classification of such quotations.

1.4.6 Evaluation

We conduct our experiments on a human-annotated corpus of 851 quotations tagged as POSITIVE, NEGATIVE, or NEUTRAL. We first evaluate each classifier of our two-stage approach separately and then assess the performance of the overall sentiment classification. In our experiments we first examine our sentiment features individually and then if combining them helps to solve the task of subjectivity and polarity classification.[26] We measure the effectiveness of our approach according to the precision, recall, and harmonic mean between precision and recall, the F1-score. We consider all classes equally important, determine the evaluation scores for each class separately, and then macro-average the scores across the classes. Our evaluation is performed as tenfold cross-validation where 90 % of the data is used to train the classifier and the remaining 10 % to test it in each evaluation run. Within the 10 folds the distribution of quotations is pertained. We normalize the feature values to fit into the interval of [0, 1]. In each run we perform a nested tenfold cross-validation to find

[26] We skip the evaluation of our target extraction solution because of the lack of text anchors for targets in our corpus. The corpus contains only a few target annotations because in most cases the targets are abstract topics or expression of sentiments rather than entities or nouns and therefore the annotators could not mark them within the quotation.

Table 1.5 Results for the subjectivity classification part

	Neutral			Subjective			Macro-averaged			Accuracy
	Pre	Rec	F1	Pre	Rec	F1	Pre	Rec	F1	
NB baseline (all)	0.914	0.856	0.884	0.314	0.450	0.370	0.614	0.653	0.627	0.804
All	0.921	0.925	0.923	0.472	0.459	0.465	0.696	0.692	0.694	0.865
All-bow	0.918	0.923	0.921	0.457	0.440	0.449	0.688	0.682	0.685	0.861
All-postags	0.924	0.903	0.913	0.429	0.495	0.460	0.676	0.699	0.687	0.851
all-bow-postags	0.925	0.942	0.933	0.547	0.477	0.510	0.736	0.710	**0.722**	0.883
Bow	0.888	0.973	0.929	0.474	0.165	0.245	0.681	0.569	0.587	0.870
Discourse markers	0.881	0.725	0.795	0.150	0.330	0.206	0.515	0.528	0.501	0.675
Postags	0.906	0.867	0.886	0.298	0.385	0.336	0.602	0.626	0.611	0.805
Sentiws	0.918	0.939	0.929	0.511	0.431	0.468	0.715	0.685	**0.698**	0.874
Valence shifters	0.874	0.795	0.833	0.136	0.220	0.168	0.505	0.508	0.501	0.722

Isolated, the SentiWS features are most suitable for subjectivity classification with a SVM. Our approach outperforms the Naive Bayes baseline. The best performing feature set achieves an F1-score of 0.72. It includes SentiWS terms, discourse marker and valence shifters

the best hyperparameter γ and C for our SVM by employing a grid search. As our corpus is highly unbalanced we set the penalty for class SUBJECTIVE 7 times larger than for class NEUTRAL and the penalty for class POSITIVE 2 times larger than for class NEGATIVE.

Subjectivity Classification. For the assessment of our subjectivity classifier we make use of all 851 annotated quotations. We consider all quotations tagged as positive or negative being SUBJECTIVE. By doing so we prepare a corpus of 109 subjective and 742 neutral quotations. The first experiments evaluate our sentiment features individually to examine their impact on the subjectivity classification task. Table 1.5 shows the results. Isolated, the SentiWS features achieve the best F1-score of 0.698. We determine the best-performing feature set, containing the SentiWS term, valence shifter, and discourse-marker-based features, by conducting a feature ablation study. We leave out one feature type in each experiment that does not improve or even worsens the classification result. Using the best-performing features we achieve a *macro-averaged F1-score of 0.72* and succeed in improving the F1-score by 0.095 over the Naive Bayes baseline and by 0.024 over the F1-score exploiting solely the SentiWS features. Regarding the classes NEUTRAL and SUBJECTIVE retrieving neutral quotations works notably better than the retrieval of subjective quotations.

Polarity Classification. We evaluate our polarity classification approach on the 109 subjective quotations of the entire sentiment corpus. As with subjectivity classification the most relevant features for polarity classification are the SentiWS features (Table 1.6). Including only SentiWS features our approach already achieves an F1-score of 0.82. We are able to improve our results by adding POS tags and discourse markers as features. Together, the three feature types achieve a *macro-averaged F1-score of 0.86*. The score is 0.062 higher than the F1-score that results by including all features into the set and 0.169 higher than the F1-score achieved by the Naive Bayes baseline with all features. We find that polarity classification is more difficult for the NEGATIVE class.

Sentiment Classification. Our sentiment classification approach aims to solve a three-class classification problem. We evaluate our overall approach by putting together the results of our two SVM classifiers. Starting with the results of the subjectivity classifier we pass all quotations classified as subjective to our polarity classifier for further separation into POSITIVE and NEGATIVE. The overall sentiment classification results in a macro-average *F1-score of 0.51*. Remember, the dataset is unbalanced and contains 86 % neutral citations. We also conduct experiments providing Gold Standard answers for one of the tasks in order to evaluate the impact of the second classifier on the performance of the overall system. First, we simulate the output of the subjectivity classifier by taking the Gold Standard answers as input of the polarity classifier (Table 1.7, Gold Subjectivity Answers + Polarity Classification) and, second, assign Gold Standard answers to quotations marked as subjective by the subjectivity classifier (Table 1.7, Subjectivity Classification + Gold Polarity Answers). We then measure the effect of each classifier when integrated in an optimal system. With Gold Subjectivity answers our overall system achieves a *F1-score of 0.86*. Looking at it the other way round, using the subjectivity classifier's output and the Gold polarity answers, our system achieves a *F1-score of 0.61*, which is around 0.1 higher than the overall system result but around 0.25 lower than the system grounded on the Gold subjectivity answers. These results correlate with the results obtained when testing the classifiers separately. The main error source is the subjectivity classifier.

Table 1.6 Results for the polarity classification part

	Positive			Negative			Macro-averaged			Accuracy
	Pre	Rec	F1	Pre	Rec	F1	Pre	Rec	F1	
NB baseline (all)	0.763	0.859	0.808	0.655	0.500	0.567	0.709	0.680	0.688	0.734
All	0.825	0.930	0.874	0.828	0.632	0.716	0.826	0.781	0.795	0.826
All-bow	0.889	0.901	0.895	0.811	0.790	0.800	0.850	0.845	0.848	0.862
All-bow-valence	0.890	0.916	0.903	0.833	0.790	0.811	0.862	0.853	**0.857**	0.872
All-bow-valence-postags	0.863	0.887	0.875	0.778	0.737	0.757	0.820	0.812	0.816	0.835
Bow	0.693	0.986	0.814	0.875	0.184	0.304	0.784	0.585	0.559	0.706
Discourse markers	0.663	0.775	0.714	0.385	0.263	0.313	0.524	0.519	0.513	0.596
Postags	0.743	0.775	0.759	0.543	0.500	0.521	0.643	0.637	0.640	0.679
Sentiws	0.863	0.887	0.875	0.778	0.737	0.757	0.820	0.812	**0.816**	0.835
Valence shifters	0.711	0.831	0.766	0.539	0.368	0.438	0.625	0.600	0.602	0.670

Our polarity SVM classifier outperforms the Naive Bayes baseline with an F1-score of 0.86 achieved on all 109 subjective quotations in our corpus. It uses a feature set consisting of SentiWS terms, POS tags and discourse markers. As with subjectivity classification the most appropriate features are the SentiWS features

Table 1.7 Results for the overall sentiment classification

	Gold subj. answers + Pol. classification			Subj. classification + Gold pol. answers			Subj. classification + Pol. classification		
	P	R	F1	P	R	F1	P	R	F1
Positive	0.849	0.873	0.861	1.0	0.521	0.685	0.642	0.479	0.548
Negative	0.75	0.711	0.730	1.0	0.105	0.191	0.077	0.053	0.063
Neutral	1.0	1.0	1.0	0.916	1.0	0.956	0.912	0.949	0.93
Macro-avg	0.866	0.861	**0.864**	0.972	0.542	**0.611**	0.543	0.493	**0.514**
Accuracy	0.977			0.920			0.870		

Our approach achieves a macro-averaged F1-score of 0.51. While the polarity classifier performs reasonable, the subjectivity classifier introduces a large error. Many negative quotations are marked as neutral and therefore are not further examined by the polarity classifier. Given correct subjectivity labels the overall performance rises to an F1-score of 0.86

1.4.7 Conclusion

We solve the problem of sentiment classification of quotations in news articles by employing a two-stage approach where we first separate subjective from neutral quotations and, second, categorize the subjective quotations as either positive or negative. Our approach performs the best for both tasks with only a subset of the presented sentiment features. In either case SentiWS features strongly contribute to an efficient sentiment classification. Leaving them out decreases the F1-score considerably. In contrast to the SentiWS features, leaving out simple bag-of-word features (uni- and bigrams) increases the classification quality so that we exclude them from the final feature sets. The relatively low overall F1-score of 0.51 mainly results from the output of the subjectivity classifier. The subjectivity classifier introduces a large error in the first step. It misses many subjective quotations which the polarity classifier would tag correctly. Particularly, the majority of negative quotations is filtered out by the subjectivity classifier. Generally speaking, separating objective from subjective quotations is especially challenging in our scenario. It is easier to classify quotations as subjective if they are positive. If quotations are negative the algorithm classifies them more often as neutral. The polarity classification quality for negative and positive quotations is comparable. As Pang et al. [37] we find that incorporating position information into the feature vectors hardly influences sentiment classification effectiveness and therefore can be excluded from the feature vectors.

Inspired by Polanyi and Zaenen [40] we intend in our future work to imply more contextual shifters and patterns for German to calculate contextual feature weights instead of only encoding the presence and frequency of valence shifters. At the same time we plan to consider discourse markers for feature weight calculation following Mukherjee and Bhattacharyya [32]. In addition, appraisal groups may serve as supplementary information for the feature vectors [53]. Considering sentiment targets our future work will include on the one hand the extension of our corpus by more text-anchored sentiment targets and on the other hand we will shift our work toward topic-oriented sentiment target detection that aims at determining supporting or opposing statements [34].

To extract even more opinionated passages from news articles in the future work we not only want to consider quotations but also include other news article parts. Furthermore, a context-dependent sentiment analysis requiring world knowledge or interpretation could retrieve additional subjective text [27].

1.5 Application

We present the results of our news aggregator to users via a web interface. The home page of our *news'd* demonstrator provides an overview of the currently most important events discussed in the news. In order to tackle the enormous amount of news material, our interface implements various ranking scores. Besides sorting news clusters by actuality or size, users may navigate news clusters ordered by their "hotness". The hotness measure combines different cluster characteristics. It weights appropriately a cluster's total growth since, its creation time, and its recent growth in a sliding time window to calculate one score that indicates how "hot" the news cluster is. Figure 1.7 shows the main page of our *news'd* demonstrator. As in other commercial news portals our interface also allows browsing news events by categories like "Politics", "Economy", etc. Each department page is organized

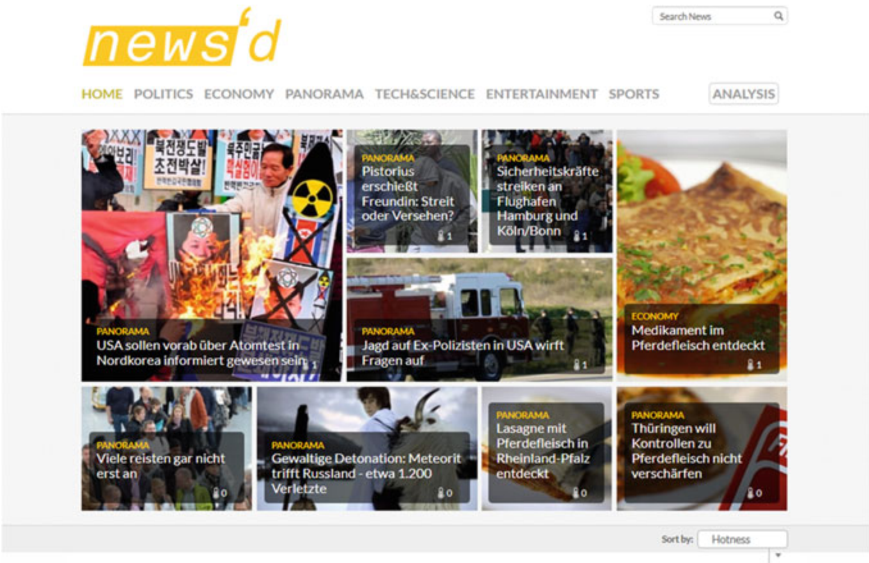


Fig. 1.7 The main page of the web application presenting the current top events

analogously to the main page except that it filters the news clusters corresponding to the category.

For a deeper, more structured insight into single events the interface organizes the view for each news cluster by dividing the page into specific sections presenting different types of information instead of simply listing all contained news articles. A news cluster view starts with its leading article, followed by the most recent news articles grouped by their age. On the right-hand side the users are provided with meta-information like a visualization of the cluster's development or the key concepts and named entities the news cluster is dealing with.

An extra analysis page offers the search for quotations and sentiment tags. Users have the possibility to specify whether to search for quotations or sentiment tags and to select how to order the results. At this point our interface allows sorting by relevance and actuality. The resulting quotations or sentiment tags, respectively, are presented at the top of the page and the corresponding news articles thereunder. Figures 1.8 and 1.9 show a preliminary presentation of the analysis search results.

The future work includes an extension of the analysis page by offering additional statistical information on quotations and their opinions and the implementation of a view that directly compares opinionated quotations according to a topic or a target entity. The comparison view allows users to easily grasp, e.g., the most opposing or the most frequent/important comments. We aim to create a service with full archive

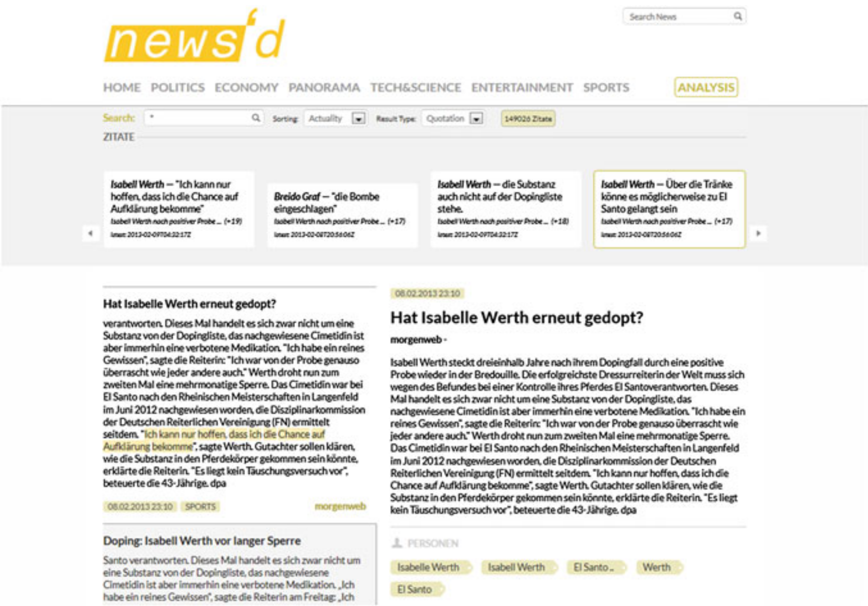


Fig. 1.8 The analysis page of the web application presenting quotations by and about "Isabelle Werth"

Fig. 1.9 Sentiment tags evaluating the perception of "Telekom" in comparison to related items

support, both to research topics and newsworthy entities as well as their perception in the press.

1.6 Conclusion

We presented a system for aggregating and analyzing German news articles collected from a wide range of news sources. In addition to topic detection and tracking of news stories, and building a dynamic topic hierarchy based on the current news situation, the central component implements methods for quotation extraction and sentiment analysis. As a stylistic means, quotations are used to underline significant information and can be regarded as a trustworthy piece of information, because they reflect statements of cited people and organizations in an original and genuine manner. Our approach to quotation extraction is rule-based and exploits text parts surrounded by quotation marks for retrieving direct quotations and the presence of reporting verbs and phrases for detecting indirect quotations. Each quotation is assigned a quotation speaker by applying a set of predefined rules and including a crude form of co-reference resolution. Evaluated on a manually created dataset with German-language quotations, the method yields convincing results, in particular for direct quotations. We assume quotations being the most subjective parts of news articles and base our sentiment analysis approach on quotations previously extracted by our quotation extraction component. We decompose the task into subjectivity detection and polarity classification and train a separate SVM classifier for each subtask. In order to find the most appropriate feature set for both tasks, we assess a range of text classification features according to their applicability to sentiment analysis. The results suggest that sentiment words are most suitable for both tasks and that the classifiers perform best using a slightly different feature set. While part-of-speech information positively effects polarity classification, incorporating valence shifters and discourse markers improves subjectivity detection. Overall, the subjectivity detection in news articles appears more challenging than the polarity classification. To evaluate our work, we have created two corpora. The first corpora aims to support developing and assessing methods for quotation extraction. It contains direct and indirect quotations attributed with a quotation speaker and a reporting verb or clue if available. The second corpus bases on our quotation corpus and provides a sentiment label (positive, negative or neutral) for each quotation and an opinion target, if it is explicitly mentioned in the quotation. Both corpora are freely available for research purposes upon request.

In the future work, we plan to improve the recall of indirect quotations by automatically detecting reporting verbs instead of using a predefined list. Concerning the extraction of quotations speakers, we intend to incorporate a sophisticated approach to co-reference resolution. The future work on our sentiment analysis approach will include incorporating additional information during feature vector calculation to represent the text more precisely. We plan to shift our work toward topic-related and context-dependent opinion retrieval and allow also other text parts than quotations for sentiment analysis. In order to benefit from our results we plan to implement an extended view on newspaper quotations. The users will be presented a direct comparison of quotations expressed by different speakers according to a topic or entity, and a timeline of opinions to facilitate monitoring developments and estimating trends.

Acknowledgments We would like to thank Neofonie GmbH for providing news articles and the infrastructure for our demonstrator and developing important components of the news aggregator. In particular, the topic detection and tracking component and the component mining a dynamic topic hierarchy were designed and implemented by Neofonie GmbH. We would also like to thank Neofonie GmbH for regular communication, many helpful discussions, and valuable suggestions. Our thanks also go to Sascha Narr, Kerstin Schütt, Michael Hülfenhaus, Jonas Katins, Xenofon Chatziliadis and Leonhard Hennig. This work was funded by the Federal Ministry of Economic Affairs and Energy (BMWi) under funding reference number KF2392309KM1.

References

1. A. Akbik, M. Schenck, *QuoteMine: A Repository of Newsworthy Quotes* (Darmstadt, Germany, 2013)
2. A. Balahur, R. Steinberger, Rethinking sentiment analysis in the news: from theory to practice and back, in *Proceeding of WOMSA'09* (2009)
3. A. Balahur, R. Steinberger, E. van der Goot, B. Pouliquen, M. Kabadjov, Opinion mining on newspaper quotations, in *Proceedings of the 2009 IEEE/WIC/ACM International Joint*

Conference on Web Intelligence and Intelligent Agent Technology, WI-IAT'09, vol. 3 (IEEE Computer Society, Washington, 2009), pp. 523–526
4. A. Balahur, R. Steinberger, M. Kabadjov, V. Zavarella, E. van der Goot, M. Halkia, B. Pouliquen, J. Belyaeva, Sentiment analysis in the news, in *Proceedings of the Seventh International Conference on Language Resources and Evaluation* (*LREC'10*)*, Valletta, Malta, May 2010*, ed. by N. Calzolari (Conference Chair), K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, D. Tapias (European Language Resources Association (ELRA), 2010)
5. L. Barbosa, J. Feng, Robust sentiment detection on Twitter from biased and noisy data, in *Proceedings of the 23rd International Conference on Computational Linguistics: Posters, COLING'10.* (Association for Computational Linguistics, Stroudsburg, (2010) pp. 36–44
6. S. Brants, S. Dipper, P. Eisenberg, S. Hansen-Schirra, E. König, W. Lezius, C. Rohrer, G. Smith, H. Uszkoreit, Tiger: linguistic interpretation of a German corpus. Res. Lang. Comput. **2**(4), 597–620 (2004)
7. S. Cerini, V. Compagnoni, A. Demontis, M. Formentelli, G. Gandini, Micro-WNOP: a gold standard for the evaluation of automatically compiled lexical resources for opinion mining, in *Language Resources and Linguistic Theory: Typology, Second Language Acquisition, English Linguistics*, ed. by A. Sanso (Franco Angeli Editore, Milano, 2007)
8. C.-C. Chang, C.-J. Lin, LIBSVM: a library for support vector machines. ACM Trans. Intell. Syst. Technol. **2**(3), 27:1–27:27 (2011)
9. S. Clematide, S. Gindl, M. Klenner, S. Petrakis, R. Remus, J. Ruppenhofer, U. Waltinger, M. Wiegand, MLSA—a multi-layered reference corpus for German sentiment analysis, in *Proceedings of the 8th International Conference on Language Resources and Evaluation* (*LREC*) (2012)
10. L.A.D.F. de Morais Sarmento, S.S. Nunes, Automatic extraction of quotes and topics from news feeds, in *4th Doctoral Symposium on Informatics Engineering* (2009)
11. É. de La Clergerie, B. Sagot, R. Stern, P. Denis, G. Recourcé, V. Mignot, Extracting and visualizing quotations from news wires, in *Human Language Technology. Challenges for Computer Science and Linguistics*, Lecture Notes in Computer Science, vol. 6562, ed. by Z. Vetulani (Springer, Berlin, 2011), pp. 522–532
12. P. Domingos, A few useful things to know about machine learning. Commun. ACM **55**(10), 78–87 (2012)
13. D.K. Elson, K. McKeown, Automatic attribution of quoted speech in literary narrative, in *AAAI*, ed. by M. Fox, D. Poole (AAAI Press, 2010)
14. A. Esuli, F. Sebastiani, SentiWordNet: a publicly available lexical resource for opinion mining, in *Proceedings of Language Resources and Evaluation* (*LREC*) (2006)
15. M. Faruqui, S. Padó, Training and evaluating a German named entity recognizer with semantic generalization, in *Proceedings of KONVENS 2010* (Saarbrücken, Germany, 2010)
16. J.R. Finkel, T. Grenager, C.D. Manning, Incorporating non-local information into information extraction systems by Gibbs sampling, in *Proceedings of the 43nd Annual Meeting of the Association for Computational Linguistics* (*ACL 2005*) (The Association for Computer Linguistics, 2005), pp. 363–370
17. J.G. Fiscus, G.R. Doddington, Topic detection and tracking evaluation overview, in *Topic Detection and Tracking*, ed. by J. Allan (Kluwer Academic Publishers, Norwell, 2002), pp. 17–31
18. A. Go, R. Bhayani, L. Huang, Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford (2009) pp. 1–6
19. V. Hatzivassiloglou, J.M. Wiebe, Effects of adjective orientation and gradability on sentence subjectivity, in *Proceedings of the 18th Conference on Computational Linguistics, COLING'00* (Association for Computational Linguistics, Stroudsburg, 2000), pp. 299–305
20. L. Hennig, D. Ploch, D. Prawdzik, B. Armbruster, H. Düwiger, E.W. De Luca, S. Albayrak, SPIGA—a multilingual news aggregator, in *Proceedings of GSCL'11* (2011)
21. M. Hu, B. Liu, Mining and summarizing customer reviews, in *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD'04* (ACM, New York, 2004), pp. 168–177

22. A. Kennedy, D. Inkpen, Sentiment classification of movie reviews using contextual valence shifters. Comput. Intell. **22**, 2006 (2006)
23. S.-M. Kim, E. Hovy, Extracting opinions, opinion holders, and topics expressed in online news media text, in *Proceedings of the Workshop on Sentiment and Subjectivity in Text, SST'06* (Association for Computational Linguistics, Stroudsburg, 2006), pp. 1–8
24. R. Krestel, S. Bergler, R. Witte, Minding the source: automatic tagging of reported speech in newspaper articles, in *Proceedings of the Sixth International Language Resources and Evaluation (LREC 2008)* (European Language Resources Association (ELRA), 2008), pp. 2823–2828
25. W. Lezius, Morphy—German morphology, part-of-speech tagging and applications, in *Proceedings of the 9th EURALEX International Congress* (2000), pp. 619–623
26. J. Liang, N. Dhillon, K. Koperski, A large-scale system for annotating and querying quotations in news feeds, in *Proceedings of the 3rd International Semantic Search Workshop, SEMSEARCH'10* (ACM, New York, 2010), pp. 7:1–7:5
27. H. Li, X. Cheng, K. Adson, T. Kirshboim, F. Xu, Annotating opinions in German political news, in *8th ELRA Conference on Language Resources and Evaluation. International Conference on Language Resources and Evaluation* (*LREC-2012*)*, 8th, 23–25 May, Istanbul, Turkey* (European Language Resources Association (ELRA), 2012). Accepted for publication
28. B. Liu, L. Zhang, A survey of opinion mining and sentiment analysis, in *Mining Text Data*, ed. by C.C. Aggarwal, C.X. Zhai (Springer, US, 2012), pp. 415–463
29. D.C. Manning, P. Raghavan, H. Schütze, *Introduction to Information Retrieval* (Cambridge University Press, Cambridge, 2008)
30. J. Martineau, T. Finin, Delta TFIDF: an improved feature space for sentiment analysis. Artif. Intell. **29**, 258–261 (2009)
31. S. Momtazi, Fine-grained German sentiment analysis on social media, in *Proceedings of the Eight International Conference on Language Resources and Evaluation* (*LREC'12*) (European Language Resources Association (ELRA), Istanbul, Turkey, May 2012)
32. S. Mukherjee, P. Bhattacharyya, Sentiment analysis in Twitter with light weight discourse analysis, in *Proceedings of COLING 2012*, Mumbai, December (2012). The COLING 2012 Organizing Committee, pp. 1847–1864
33. T. Nakagawa, K. Inui, S. Kurohashi, Dependency tree-based sentiment classification using CRFs with hidden variables, in *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics* (Association for Computational Linguistics, Los Angeles, 2010), pp. 786–794
34. T. O'Keefe, J.R. Curran, P. Ashwell, I. Koprinska, An annotated corpus of quoted opinions in news articles, in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (Association for Computational Linguistics, Sofia 2013), pp. 516–520
35. T. O'Keefe, S. Pareti, J.R. Curran, I. Koprinska, M. Honnibal, A sequence labelling approach to quote attribution, in *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning* (Association for Computational Linguistics, Jeju Island, (2012), pp. 790–799
36. B. Pang, L. Lee, Opinion mining and sentiment analysis. Found. Trends Inf. Retr. **2**(1–2), 1–135 (2008)
37. B. Pang, L. Lee, S. Vaithyanathan, Thumbs up?: sentiment classification using machine learning techniques, in *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing, EMNLP'02*, vol. 10 (Association for Computational Linguistics, 2002), pp. 79–86)
38. S. Pareti, T. O'Keefe, I. Konstas, J.R. Curran, I. Koprinska, Automatically detecting and attributing indirect quotations, in *EMNLP* (ACL, 2013), pp. 989–999
39. W. Paulo, D. Fernandes, E. Motta, R.L. Milidiú, in *Proceedings of the 8th Brazilian Symposium in Information and Human Language Technology*, Chapter Quotation extraction for Portuguese (2011)
40. L. Polanyi, A. Zaenen, Contextual valence shifters, in *Working Notes—Exploring Attitude and Affect in Text: Theories and Applications (AAAI Spring Symposium Series)* (2004)

41. B. Pouliquen, R. Steinberger, C. Best, Automatic detection of quotations in multilingual news, in *Proceedings of the International Conference Recent Advances in Natural Language Processing* (Borovets, 2007), pp. 487–492
42. R. Remus, U. Quasthoff, G. Heyer, Sentiws—a publicly available German-language resource for sentiment analysis, in *Proceedings of the 7th International Language Resources and Evaluation* (*LREC'10*) (2010), pp. 1168–1171
43. E. Riloff, J. Wiebe, W. Phillips, Exploiting subjectivity classification to improve information extraction, in *Proceedings of the 20th National Conference on Artificial Intelligence, AAAI'05*, vol. 3 (AAAI Press, Pittsburgh, 2005), pp. 1106–1111
44. A. Schiller, S. Teufel, C. Thielen, *Guidelines für das Tagging deutscher Textcorpora mit STTS*. Technical Report, IMS-CL, University Stuttgart (1995)
45. N. Schneider, R. Hwa, P. Gianfortoni, D. Das, M. Heilman, A.W. Black, F.L. Crabbe, N.A. Smith, Visualizing topical quotations over time to understand news discourse. Technical report, Technical Report TR CMU-LTI-10-013, Carnegie Mellon University, Pittsburgh (2010)
46. T. Scholz, S. Conrad, L. Hillekamps, Opinion mining on a German corpus of a media response analysis, in *Text, Speech and Dialogue*, vol. 7499, ed. by D. Hutchison, T. Kanade, J. Kittler, J.M. Kleinberg, F. Mattern, J.C. Mitchell, M. Naor, O. Nierstrasz, C.P. Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M.Y. Vardi, G. Weikum, P. Sojka, A. Horák, I. Kopeček, K. Pala (Springer, Berlin, 2012), pp. 39–46
47. C. Strapparava, A. Valitutti, WordNet-affect: an affective extension of WordNet, in *Proceedings of the 4th International Conference on Language Resources and Evaluation* (ELRA, 2004), pp. 1083–1086
48. M. Tsytsarau, T. Palpanas, Survey on mining subjective data on the web, in *Data Mining and Knowledge Discovery* (Springer, New York, 2011)
49. P.D. Turney, Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews, in *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, ACL'02* (Association for Computational Linguistics, Stroudsburg, 2002) pp. 417–424
50. V.N. Vapnik, *The Nature of Statistical Learning Theory* (Springer, New York, 1995)
51. S. Wang, C. Manning, Baselines and bigrams: simple, good sentiment and topic classification, in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Short Papers)*, vol. 2 (Association for Computational Linguistics, Jeju Island, 2012), pp. 90–94
52. St. Weiser, P. Watrin, Extraction of unmarked quotations in newspapers, in *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, ed. by N. Calzolari (Conference Chair), K. Choukri, T. Declerck, M.U. Doğan, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis (European Language Resources Association (ELRA), Istanbul (2012)
53. C. Whitelaw, N. Garg, S. Argamon, Using appraisal groups for sentiment analysis, in *Proceedings of the 14th ACM International Conference on Information and Knowledge Management, CIKM'05* (ACM, New York, 2005), pp. 625–631
54. H. Yu, V. Hatzivassiloglou, Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences, in *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing, EMNLP'03* (Association for Computational Linguistics, Stroudsburg (2003), pp. 129–136