

# VISA: A Visual Sentiment Analysis System

Dongxu Duan   Weihong Qian

IBM Research – China  
19 Zhongguancun Software Park  
Beijing, 100193 China  
{duandx, qianwh}@cn.ibm.com

Lei Shi\*

Institute of Software  
Chinese Academy of Sciences  
Beijing, 100190 China  
shil@ios.ac.cn

Shimei Pan

IBM T. J. Watson Research Center  
19 Skyline Drive  
Hawthorne, NY 10578 USA  
shimei@us.ibm.com

Chuang Lin

Department of Computer Science and Technology  
Tsinghua University  
Beijing, 100084 China  
chlin@tsinghua.edu.cn

## ABSTRACT

Sentiment plays a critical role in many information-centric business scenarios. The opinion mining methods proposed in the recent decade have formed a solid foundation to investigate the sentiment analysis tasks, but are often too complicated and scattered to serve the needs of real customers. We introduce the VISA system in this paper, which applies the visualization technology to synthesize the sentiment analysis results and present to the end user in an interactive manner. VISA builds on the generic sentiment tuple based data model and consumes the different facets of sentiment data with coordinated multiple views, hence is scalable to work with most of existing sentiment analysis engines on various application domains. We showcase the usage of VISA in two real world examples and also demonstrate the system's effectiveness through user experiments comparing to two baseline systems in both objective and subjective metrics.

## Categories and Subject Descriptors

H.5.2. [Information interfaces and presentation (e.g., HCI)]: User Interfaces. I.3.6. [Methodology and Techniques]: Interaction Techniques

## General Terms

Design, Human Factors.

## Keywords

Text visualization, sentiment analysis, opinion mining.

## 1. INTRODUCTION

Sentiment analysis is becoming more and more important as the online opinion data, such as micro-blogs, product reviews, social news, grow tremendously. There have been a lot of efforts in automated sentiment analysis (or opinion mining [1]). To name a few, Pang et al [2] studied the document-based sentiment

classification and evaluated three machine learning algorithms in completing the task. Hu et al [3] proposed a more fine-grained opinion summarization method in product feature level. In most cases, visualization was employed to examine and analyze the mining results. Gamon et al [4] constructed a tree-map to visualize the sentiment clusters. Oelke et al [5] proposed a 2-D matrix to show the sentiment overview of customer feedbacks and a radial graph to analyze the correlations among the numerical ratings, sentiment orientations and product features of the reviews. While existing visualization methods are impressive in visually illustrating the sentiment analysis result, there are still big gaps in helping users in their real business tasks, e.g. making decisions based on a huge amount of text. Rather than only showing which document or feature tends to be positive or negative, the users also need to know how the positives/negatives are described in documents (*context*); how the sentiment evolves over time (*trend*) and how one user's sentiment compares to others' (*compare*). Furthermore, in most existing systems, there is a lack of the integrated approach to visualize both the sentiment analysis results and the structured facet data (e.g. profile of the reviewer) to give a more out-of-box context of the sentiments.

Designing visualizations to meet the above goals is a nontrivial task. For example, sentiment analysis methods differ from one to another. They also use different output format, which makes it difficult to build a generic sentiment visualization tool to consume the results of different sentiment analysis algorithms. Visualizing both the sentiment results and their contexts together is challenging. It's easy to fall into some counter-intuitive complicated view piling up with all kinds of information.

In this paper, we present VISA - the Visual Sentiment Analysis system. VISA evolved from a generic text visualization tool called TIARA [21, 22, 23] which is designed to demonstrate the overview of a text corpus in temporal trends and topics. To cope with the new elements in sentiment analysis scenario, we optimize the basic trend design and the content panel with several sentiment-centric enhancements. We also introduce the dashboard view to interpret the important facet information and finally wrap up all the visualizations in a coordinated multiple view approach. To name a few of our innovations, we have experimented with the two-sided trend view to compare sentiments and the fine-grained feature/opinion-based trend summarization for more precise sentiment perception.

\* Corresponding Author

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

VINCI'12, September 27–28, 2012, Hangzhou, China.

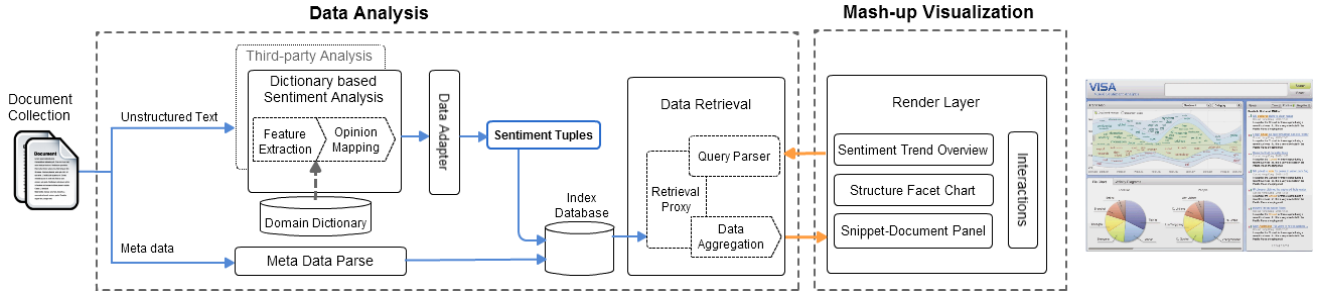


Figure 1: The system overview of VISA composed of the data analysis backend and mash-up visualization frontend.

VISA also employs a novel sentiment data model to support the finer-grained sentiment analysis. At the core of the data model lies the *sentiment tuple*, which is composed of four elements: *feature*, *aspect*, *opinion* and *polarity*. This data model serves as a data interface between the backend sentiment analysis system and the visualization frontend. With such an architecture design, it is possible to plug-in an external sentiment analysis engine in VISA if its analysis results can be transformed into sentiment tuple model. We also introduce a default implementation of the sentiment analysis engine based on domain dictionaries. Based on the sentiment tuple model, we further developed a time and topic summarization method to generate the data directly for visual mapping.

We showcase the usage of VISA through two real world examples: the visual sentiment comparison for hotel’s online customer reviews, and the sentiment summarization for the NFL-related twitter data. We also conducted an experiment to compare VISA to two baseline systems from user’s perspective. Results show that VISA outperforms the baseline systems significantly in both the objective benchmarks in term of the average completion time, correctness and the subjective results in terms of user perceived usability and satisfaction.

The reminder of this paper is organized by first reviewing the related works in Section 2. Then, we give an overview to the VISA system in Section 3. The details of the data analysis process and the visualization design and implementation are further explained in Section 4 and 5. We describe two case studies and present the user experiment results in Section 6 and finally conclude the paper in Section 7.

## 2. RELATED WORK

We first summarize existing sentiment analysis literatures, especially those related to visualization. Then we give an overview of existing sentiment visualization techniques.

Literatures on sentiment analysis mainly focus on product or movie reviews [3, 6], which are convenient for evaluation. The other document genres, such as web pages and news, have also been explored [7]. Most of these work employed document-level or sentence-level sentiment analysis to extract general opinions [2, 8]. Since both the positive and negative information may co-exist in the same document, making it hard to indicate the overall sentiment of a document with only one score. To understand the detailed sentiment context, feature-level sentiment analysis has been proposed in [9, 10]. The sentiment analysis framework of VISA employs the feature-level sentiment analysis. In addition, the topic-level sentiment analysis was also investigated. Some

attempted to incorporate the sentiment into topic models such as the Probabilistic Latent Semantic Indexing (PLSI) and the Latent Dirichlet allocation (LDA) to explain the opinion generation process [11, 12]. However, it is hard to accurately classify documents with both sentiment polarities and topics simultaneously. In VISA system, we take a more conservative approach. First we classify documents using existing categories or automatically derive topics. Then we divide each topic/category into positive and negative parts by feature-level sentiment analysis.

The sentiment visualization is attracting growing attentions because it is effective as both an information access technology for end users and an evaluation technology for the opinion mining community. In both cases, visualization creates summaries from opinion mining results. Pulse [4] clustered reviews by topics and displayed the overall sentiment of these topics in a treemap. Each treemap node drew one keyword to summarize the corresponding topic. Morinaga et al. [13] aggregated key phrases of the group of positive and negative sentences into a 2D scatterplot. Gregory et al. [16] proposed an adaptive rose plot to display both sentiment polarity and aspects such as pleasure, plain, and conflict. The BLEWS system [17] represented the number of documents related to a specific news article as a bar, and drew a glow to encode the sentiment strength. Rohrdantz et al. [18] adopted the 2-D matrix to compare the linguistic features across multiple documents and discovered the linguistic phenomenon called Vowel Harmony.

Some other visualization works involve temporal sentiment analysis. Wanner et al. [14] created a bar chart-based trend visualization to analyze the RSS data of US presidential election in 2008. Chen et al. [15] drew conflicting movie reviews of The Da Vinci Code in a two-sided trend chart. Diakopoulos et al. [19] combined a sentiment timeline with a topic timeline to help journalists extract valuable news from a huge collection of social media content around broadcasting events. Narratives [20] assisted users place news stories in their historical and social context by understanding how the relevant topics changed over time.

## 3. SYSTEM OVERVIEW

As Fig. 1 illustrates, VISA takes a raw document collection as the input and builds a visual sentiment summary for the data as the output. VISA is composed of two main components: a data analysis component and a mash-up visualization component. The data analysis backend extracts sentiment information from text. It also structures the results for effective presentations. The mash-up visualization frontend takes pre-formatted sentiment data as input and creates a suite of visualizations to help users understand sentiment analysis results.

### 3.1 Analysis

The data analysis component first preprocesses the data by separating them into unstructured text and structured metadata. Then the two data sources are processed in two parallel paths.

In the first path, the unstructured text is analyzed by the sentiment analysis/annotation engine which can be either the default dictionary-based engine in VISA or an external 3<sup>rd</sup> party engine. Then the sentiment analysis results are transformed into a generic data model, namely the sentiment tuples, which consists of four elements: *aspect*, *feature*, *opinion*, and *polarity*. Details of the sentiment tuple will be described in Section 4. In another data path, the structured metadata is parsed into a list of *field-value* pairs for further usage.

After these processing, the output sentiment tuples and the metadata *field-value* pairs are stored and indexed using the Lucene indexing system. The indexed content later can be retrieved by a data retrieval component. Later, a data aggregation component combines the sentiment tuples and the associated metadata into a data format that can be directly consumed by the visualization frontend.

### 3.2 Visualization

Built over the sentiment analysis results, the mash-up visualization is designed to give a comprehensive overview of sentiment in a document collection. Three interactive widgets are snapped together as coordinated multiple views: the sentiment trend view showing the temporal sentiment dynamics and sentiment comparisons among different categories/topics; the chart visualization view illustrating the associated structured facets, and the snippet/document panel providing details of documents and context of sentiment. Mash-up capabilities among the three views allow the user to navigate the data set using optimal interactions.

## 4. DATA ANALYSIS

In this section, we present the details of our data analysis process. We start from an introduction to our data model, followed by a brief description of our sentiment analysis, facet analysis and data retrieval process.

### 4.1 Data Model

Data models define the format of data used by the main components of VISA. There are two data models in VISA: an index data model and a visualization data model.

The index data model defines the format of the data analysis results stored in the index database. The index data model consists of a set of documents. In addition to the text body, each document is associated with a set of structured fields such as category, time, facet, and sentiment tuples. Among the structured fields, some are native fields that come with the raw data (e.g. time stamp); others are derived automatically from the data set (e.g. category, sentiment tuples). For example, the *category* information can be automatically derived based on topic analysis. It can also be inferred based on human-provided tags. In the following, we want to focus more on one of the derived structured fields, the sentiment tuple, since it is the focus of this paper.

A sentiment tuple provides a finer-grained representation of sentiment. Currently, it is defined as  $\{Aspect, Feature, Opinion, Polarity\}$ . An *aspect* is a sub-topic which is shared by all the documents, and it's represented by a set of features. In our hotel

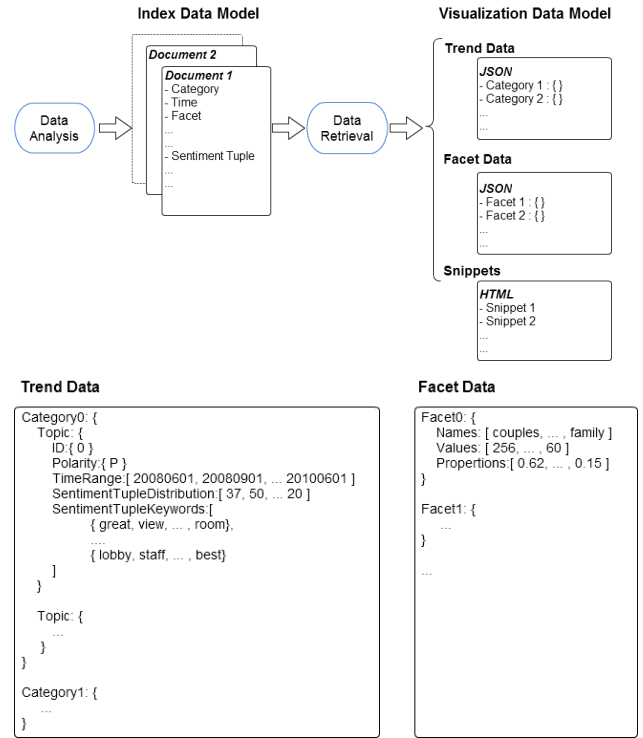


Figure 2: Data models in VISA backend analysis part.

data set, an aspect can be a room, a view, or a service. It represents a middle ground for summarizing the sentiment information. A *feature* is a specific object users are commenting on. It could be an entity, person, location, and abstract concept etc. Each feature is assigned to an aspect. In addition, an *opinion* is a particular word or phrase describing a feature. Previous works on sentiment analysis mostly ignored opinion words but instead using only a metric to measure it. In many applications however, opinion words are irreplaceable. First, opinion words may differ in tone strength which can be hard to accurately measure; second, opinion words, when highlighted in context, will help user pinpoint the opinion in the original document easily. A *polarity* is the sentiment orientation of an opinion word, positive or negative. Since the polarity of an opinion word may change when there is a preceding privative or adversative, we add a *not* before the opinion word to indicate that situation. The sentiment tuple is relatively simple. Different sentiment analysis results can be transformed into this representation with limited effort.

The visualization data models define the data formats for sentiment visualization. It consists of three main data types: the trend data, the facet data and the snippet data.

The sentiment trend data consists of one or more categories. Each category is represented as a JSON object. Each object includes all the information needed to display a strip in the sentiment trend view. Figure 2 shows the representation of such an object. A *category* represents the sentiment trends of a collection of documents (e.g. all the positive or negative reviews of a particular hotel). As the positive and negative trends are displayed separately in the trend overview, each category is divided into two topics. This partition is based on the sentiment tuples, not the documents. Each *topic* contains a set of positive or negative sentiment tuples of a category. The trend of a topic is represented

by a series of points. The X coordinates of these points is time, which is in the *TimeRange* attribute. The height of these points is in the *SentimentTuplesDistribution* attribute, which is computed from the count of sentiment tuples in the documents between the times of two points. The features and opinion words of the sentiment tuple in each time range are accumulated, and the top ranked keywords are put into the *SentimentTupleKeywords* attribute.

Similarly, each facet data is also wrapped into a JSON object array. All the facet items a document has are put into the *Names* array and the number of documents in each facet item is put into the *Values* array. *Proportions* are the percentage of documents in a facet. *Proportions* are computed from the *Values* array.

Finally, the snippet data is wrapped into HTML text. As the number of documents retrieved will be large and exceed the display capacity of the snippet-document panel, the data is paged. At the beginning of the panel, a summary of the retrieve results such as the total number of hit documents and the retrieval time are displayed. Then, for each document, the system displays its category name, time stamp, a bar chart summarizing the sentiment of the document and a trimmed content.

## 4.2 Dictionary-based Sentiment Analysis

Our sentiment analysis component uses three domain-specific dictionaries. We use the dictionaries for four purposes. First, we need to define the features frequently mentioned and discussed in a domain. Second, these features need to be clustered into aspects. Moreover, the opinion words describing features should be extracted. Finally, adversative words should be extracted as their usage may change the polarity of the opinion words.

To encode the above information, the *Entity Dictionary* lists all the features and the aspect each feature belongs to. The *Word Dictionary* lists the opinion words. The *Opposite Dictionary* lists all the privative and adversative words.

At the beginning of the data analysis process, a document is split into segments, and each segment is tokenized. Then all the tokens will be searched in the Entity Dictionary to determine their associated features and aspects. If a token is not in the Entity Dictionary or there is no opinion word nearby in the same segment, it is dropped. Otherwise, system will use the Opposite Dictionary to determine the polarity of the opinion words. Based on the analysis result, the system automatically constructs a sentiment tuple. After all the segments have been processed, all the sentiment tuples constructed will be stored as a vector within each document.

## 4.3 Facet Analysis

In addition to the text that is used in sentiment analysis, the structured information in a document is also valuable for both the analysts and users. In the current implementation, we extract structured information including trip type, the visit reason, and the rating in the TripAdvisor scenario, and in the NFL scenario, we extract the user profiles as tags.

## 4.4 Data Retrieval

Once a user submits a query, the data retrieval component retrieves relevant documents and returns the aggregated results.

There are two types of queries a user can perform using VISA: document-based filtering such as time, category, structured facet-

based filters; and sentiment-based filtering, including the aspect, polarity, and sentiment tuple, keyword-based filtering.

These parameters are reorganized into an index query and a data aggregation strategy. The retrieval component uses the index query to search in the index database and returns a collection of matched documents. Then the data aggregation component extracts queried fields within these documents based on the aggregation strategy, and wraps the field values into visualization data results.

## 5. MASH-UP VISUALIZATION

In this section, we cover VISA's overall visualization design and its several sentiment-centric optimizations.

### 5.1 Design

VISA visualization is designed to address the needs of both analysts and users. In an analytics view, VISA consumes different sentiment analysis results and generates comprehensive visual summaries so that it is easy to evaluate a sentiment analysis engine's effectiveness. In a user view, VISA works as a concise visual tool to help the ordinary people understand the sentiments conveyed in a huge amount of textual data.

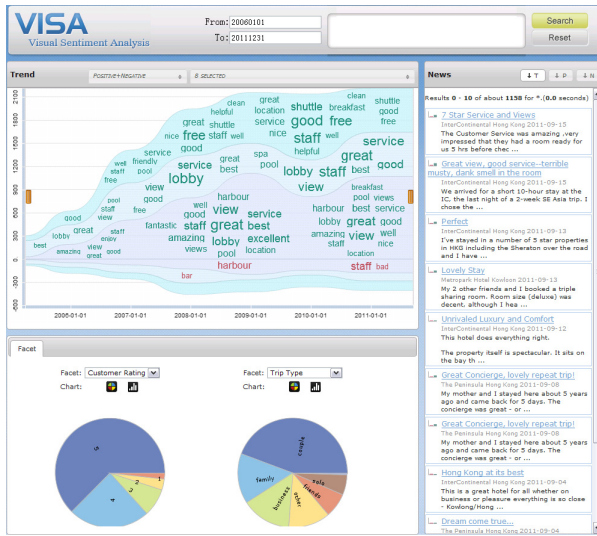
Balancing the above considerations, we adopt the classical coordinated multiple view design that snaps together several existing visualization widgets. On the one hand, such design allows comprehensive analysis and exploration of raw data. It on the other hand also makes each visual widget simple and concise so that it will not frighten novice users.

As shown in Fig. 3, the design includes a sentiment trend visualization, which is the major view of VISA, summarizing the sentiment tuples extracted from the raw text. It also combines a dashboard panel representing the facet information in the documents. The third snippet view displays the textual content including sentiment snippets and original text. A query bar is also an integrated part of the design, with both a keyword search box and a time range controller.

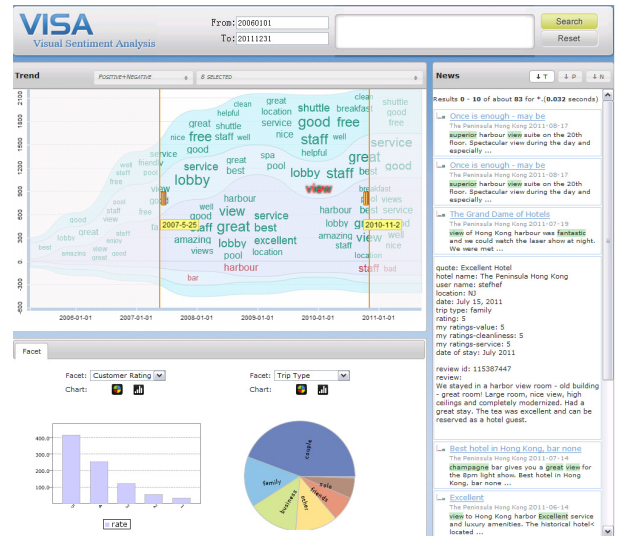
#### 5.1.1 Sentiment Trend View

The sentiment trend view is built on top of the TIARA approach [21, 22, 23], which is designed to visually summarize the content of text collection from both temporal and topical perspectives. In VISA, to handle the sentiment scenario, we map the category info to topics in TIARA and the textual sentiment-tuples to topical keywords in TIARA.

As given in the top-left panel of Fig. 3, the trend view encodes the input sentiment tuples. Specially, 1) The X axis is mapped to the time facet. 2) The stacks along the Y axis are mapped to the category of a document, such as a particular hotel in the customer review scenario and a particular team in the NFL twitter scenario. The height along Y axis in each stack indicates the number of sentiment tuples in each category. Hence each stacked strip in the view corresponds to one category's sentiment trend in the data. 3) The keywords on each stacked strip are summarized from the sentiment tuples of the corresponding category in a time-sensitive manner to reveal the content evolvement over time. The size of each keyword is mapped to the re-ranked keyword weight. The more frequent a keyword appears in the corresponding time frame compared to other time frames, the larger the keyword will be drawn. 4) The color of the keyword is mapped to the polarity of



(a)



(b)

Figure 3. VISA visualizations: (a) the coordinated multiple views. (b) interactions.

the sentiment tuple, with green indicating positive and red indicating negative.

In this design, the category selector is combined with another aspect selector in the same drop-down list. The category selector allows a user to customize the visual sentiment summary by choosing a set of categories to compare. The aspect selector allows a user to only focus on the content related to selected aspects.

Apart from the visual mappings described above, the trend view also invents several sentiment-specific new features including the two-sided trend view and the sentiment tuple mapping. These will be detailed in Section 5.2.

### 5.1.2 Facet Dashboard View

To present the structured facet data in each document, VISA introduces the facet dashboard view. An example is given in the bottom-left panel of Fig. 3, which includes two pie charts showing the distribution of customer rating and trip type in the reviews. The dashboard is designed to be customizable. It allows the user to add a chart, select the corresponding data facet, switch among different chart types (including pie chart, bar chart, line chart etc.) and finally delete one if it is not useful any more. The dashboard view is also connected to the other widgets in VISA and is updated during data brushing and filtering.

### 5.1.3 Sentiment Snippet Panel

This panel lists the snippets corresponding to the sentiment tuples. Each snippet is defined as a block of continuous text in the document surrounding both the corresponding opinion and feature words. In the panel, each snippet is shown by its document title, date and text content. By clicking on the document title, the original document content is shown in context.

Initially, the snippets of all the sentiment tuples in the document collection are shown. Consequently, changes of time-range/category/aspect/polarity/keyword selection will also update the snippet list to display the sentiment tuples in the new document set.

## 5.2 Optimization for Sentiment

### 5.2.1 Trend Visualization

In the trend view, several new features are introduced to better illustrate sentiment analysis results.

First, the trend visualization is extended in VISA to support the positive and negative mega-groups. By default, a two-sided trend chart is drawn to summarize the sentiments in each polarity respectively, with positives in the top and negatives in the bottom. The view could also be switched to a one-sided view by selecting only one polarity. Note that the neutral sentiments, although extractable by the analysis, are not presented in our design due to little interest from users.

Second, VISA evolves to illustrate the sentiment tuple-level trend rather than the straightforward document-level trend. Experiences over several real data sets inform us that many documents contain both positive and negative sentiments. Summing up sentiments within each document and drawing the corresponding trend will be too coarse-grained to reflect the real sentiment distribution. In the hotel customer review case, the general document-level sentiment trend has almost no negative part.

Third, the keyword content in VISA is summarized from the sentiment tuples rather than all the keywords in the document collection. The current approach in VISA aggregates the feature words and opinion words together into a time-sensitive keyword set. VISA then summarizes these time-sensitive keyword lists similar to that used in TIARA [21, 22, 23]. Users may switch between feature keywords, opinion keywords or both by interacting with VISA.

### 5.2.2 Text Content Visualization

The snippet panel in VISA is also customized to support sentiment-analysis. For the text content shown in each snippet, we not only keep the background tags for the selected keywords, but also the keyword tags for their associated opinion/feature words. For example, when the feature word “staff” is clicked in the positive trend, the corresponding opinion words such as “warm”, “friendly”, “helpful” are tagged. When the same word is clicked



in the negative trend, the opinion words such as “unfriendly” and “not welcoming” are highlighted.

The snippet panel also provides new sentiment-related interactions. Except for sorted by time, the snippet list can also be sorted by the number of positive/negative sentiment tuples in their documents. This feature helps the user locate the snippets with the strongest sentiment. To illustrate the exact sentiment distribution of a document, we add a thumbnail bar chart before each snippet based on the number of positive/negative tuples in the document.

## 5.3 Interactions

### 5.3.1 Single-View Interactions

**TIARA INTERACTIONS:** VISA’s trend view keeps most of TIARA interactions which include the temporal navigation by dragging the side bars in the trend, category customization through the category/aspect selector, category reordering by dragging the trend strip up and down, category-based document/tuple quantifying with line-style detector and the keyword query by clicking on a keyword in the trend view.

**SENTIMENT SUMMARY SELECTION:** Besides TIARA original interactions, the most critical new interactions introduced in VISA is to create a sentiment summary of a specific subset of documents. VISA enables the selection of documents based on a facet in time (through time range box), a category and an aspect (through the right drop-down selector), a polarity (through the left drop-down selector), and any combination of these facets.

### 5.3.2 Mash-up Interactions

**BRUSHING:** Like most coordinated multiple view visualizations, VISA provides the brushing interactions to allow user to understand all the facets of a small data set, in the context of the whole data set. In the current version, VISA only supports brushing of the document content based on one or several sentiment keywords. By explicitly selecting these keywords in the trend view, the existences of the keywords in all the time range are also highlighted. In the same time, the facet dashboard is updated to illustrate the facets of the related document sets. The snippet panel also switches to the list of the corresponding snippets containing the keywords.

**FILTERING:** Filtering is handed similarly to the brushing interaction, except that the trend view is updated to only summarize the sentiments of the corresponding document set. In VISA, filtering by keyword is enabled by using the search box in the top-left of the view. Filtering by facet can be achieved by clicking on the values of one of the charts shown in the dashboard.

## 6. EVALUATION

### 6.1 Case Study

Our first case study is in the travelling domain. We collect 11032 reviews of 173 Hong Kong hotels on TripAdvisor.com [www.tripadvisor.com] from 2002 to 2010. Each review consists of the textual comment, the reviewer’s profile (country, visit type and length of stay) and the reviewer’s numerical ratings to the hotel. We classify the features extracted from the textual comment into five aspects, according to the rating aspects in the website, namely value, location, cleanliness, room and service. The other processing follows the default settings of the VISA system. Suppose John, a middle-aged man, plans to take a holiday

with his family in Hong Kong. John logs in to VISA to find a good hotel to stay during their vacation there. In his first access, VISA provides him a summary of 8 hotels’ customer reviews, as shown in Fig. 4(a). These hotels are the system-recommended ones ranked by the number of positive reviews. John would like to receive advices from the other visitors on the family vacation. So he proceeds by selecting the “Family” option in the trip-type chart of the facet dashboard. As their major interest is on sightseeing and shopping, the location of the hotel is their first priority. John then checks the location aspect and de-checks the hotels with fewer comments in the location aspect. To identify the outstanding hotels, he changes the sentiment selector into positive, as shown in Fig. 4 (b). In the generated view, John quickly discovers a lot of relevant keywords such as “view”, “harbor”, “fun” etc. John clicks on these keywords one by one, and examine the related snippets. For example, he clicks the word “free” on the category of Metropark Hotel. He glances over the snippet list and unfolds some snippets into full text to examine (Fig. 4(c)). In the process, he finds useful information such that this hotel has a free shuttle bus to the airport, the hotel has a 30 minutes driving distance away from the harbor! Since the night scene of the harbor is one of their plans, he then excludes the Metropark Hotel. Considering their children, John also put the cleanliness and the room aspect to high priorities. He checks the cleanliness and room aspects and changes the sentiment selector into negative to detect alerts, as shown in Fig. 4(d). John quickly discovers that in the past, the negative comments of the Peninsula Hotel are much more than those of InterContinental Hotel, however, in recent months, the Peninsula improves dramatically, with much less negative comments than the InterContinental, as shown in Fig. 4 (e). Finally, John decides to stay at Peninsula Hotel.

In our second case study, we apply the VISA system to analyze the user’s sentiments in the social media data. We crawled 665360 tweets from twitter [www.twitter.com] on the topic of National Football League (NFL), from late March 2011 to early August 2011. These tweets are from 307973 users, with an average length of 16.8 words. For each user, we extract their bios, locations and the number of followers/followings/tweets from the website.

We conduct three separate analyses on the data collection to compose the sentiment tuples defined in VISA. First, we annotate the collection of tweets with a domain-independent sentiment dictionary, identifying all the opinion keywords associated with each tweet. Second, we classify all the tweets with LDA into 5 topics as data aspects, including “season play”, “player draft”, “lockout bad”, “lockout end” and “football return”. In this step, we notice that most of the tweets in the collection are talking about NFL Lockouts, as we can see that the time period of the data set is in the NFL offseason when the most important event is the 2011 NFL Lockouts. Third, we categorize all the tweets according to the subject of the sentiments – NFL teams. For each of the 32 teams, we manually create a relevant subject keyword list including the team full/nick name, city, stadium, head, owner and super stars. We match all the tweets with these team lists and assign one or more category tag for each possible tweet. Finally, it turns out that out of the 665360 tweets, 196055 tweets have at least one positive/negative annotations, 43258 tweets have NFL-team related subjects, and all tweets have aspect tags as LDA works as a classification algorithm.

Figure 5(a) gives an overview of the sentiments in the NFL tweets, categorized by the aspects, which gives an evolvement map of the

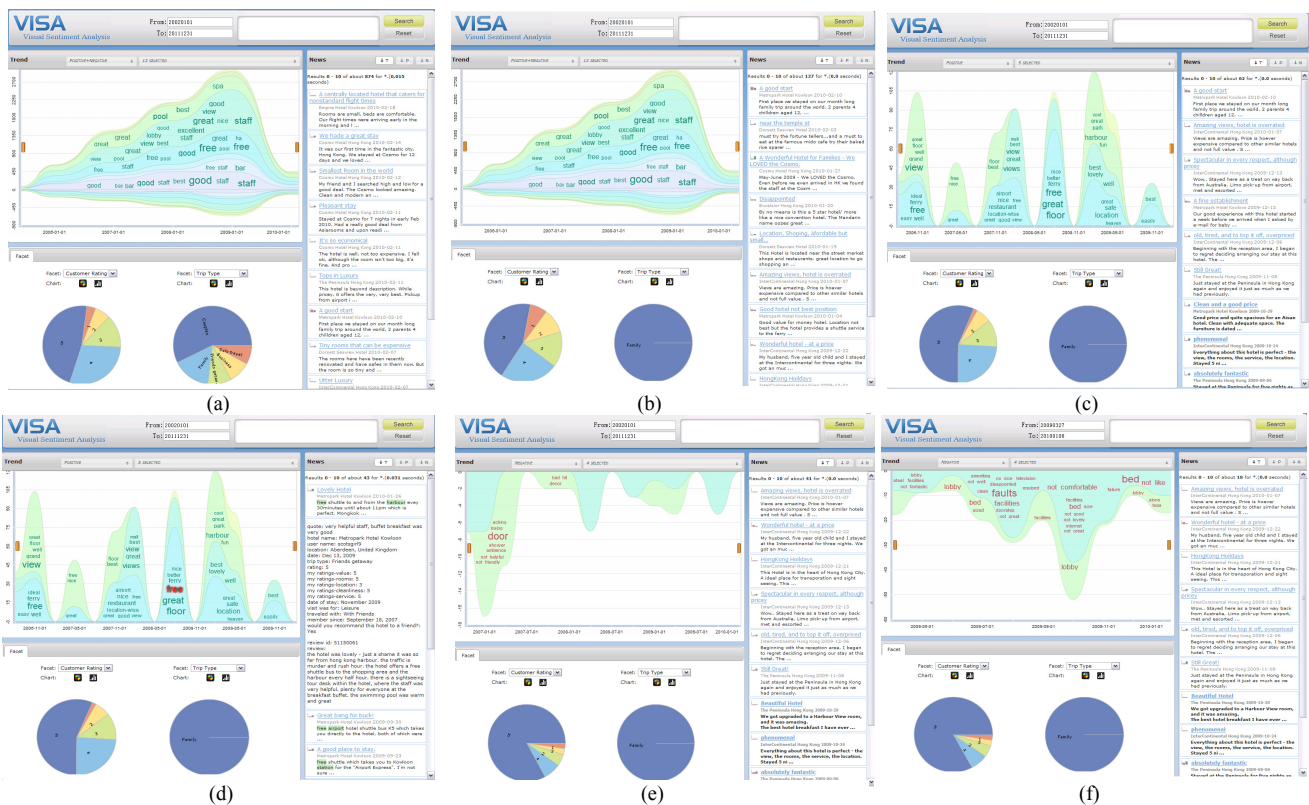


Figure 4: Case study in HK hotel customer review scenario.

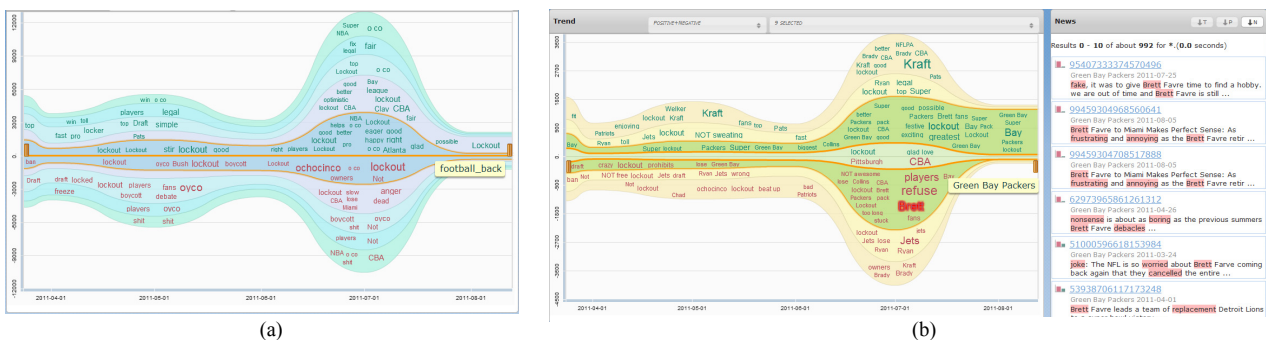


Figure 5: Case study in NFL twitter data collection.

NFL 2011 Lockouts. In the beginning, the fans were talking about the “draft”, “freeze” and “locked” as they did not predict the current lockout to be such a big thing. As the time goes, more explicit “lockout” appeared followed by the “debate” and “boycott”. Then in the July as the new agreement is being made, more attentions are paid, focusing on the “player”, “owner”, “CBA” and the sentiments including the negatives ones “anger”, “dead” and the positive ones “good”, “happy”.

Figure 5(b) shows the team-oriented view where the four popular teams (“Green Bay Packers”, “Pittsburgh Steelers”, “New York Jets”, “New England Patriots”) are selected. We pay our attention to the Packers and the Steelers. Both positive and negative comments reach their peaks in July where the new Collective Bargaining Agreement (CBA) is proposed, voted and approved by both the team owners and players. Drilling down to the major negative event/topic during this time, we notice the outstanding keyword “CBA” in red for the Steelers. By examining the content

through the keyword context in snippets, it is discovered that the Steelers is the only team that votes “No” to the new CBA, leading to the negative comments from a lot of NFL fans. For the Packers, most of the concerns focus on “Brett”, which corresponds to “Brett Favre”, the former NFL all-star quarterback who spent the majority of his career with the Packers. In the related comments, we find that fans are generally impatient about the frustrating debates of the return of Brett in season 2011, given his numerous retire-and-returns in history.

## 6.2 User Experiment

We conducted an experiment to explore how efficiently people could learn and use VISA. We choose the reviews of three Hong Kong hotels from TripAdvisor as our test data set. Both structured (e.g. rating) and unstructured data (e.g. user comments in free text) were extracted from these reviews. We also included two baseline systems in the study: the TripAdvisor website and a text

editor. TripAdvisor is a commercial tool specifically designed for organizing and presenting hotel review data while the text editor allows users to easily browse the reviews saved as local files. Twelve (7 male, 5 female) volunteers with ages from 26 to 35 participated in the study. Each volunteer was given a gift for his/her participation.

### 6.2.1 Task Design

We tested the systems with two sets of tasks. The first set composed of three tasks focuses on looking up specific sentiment-related information of a hotel (e.g. traveler's ratings). The second set of tasks focuses on summarizing opinions on a general aspect of a hotel (e.g. the view of a hotel). The three tasks within each set are similar. For example, the three tasks in the first set require users to find the same information for three different hotels. Each of the three tasks was then performed by each user using one of the three systems in shuffle: VISA, TripAdvisor and a text editor.

### 6.2.2 Procedure

We ran the study as a within-subjects design, with each participant performing all sets of the tasks using all the three systems. The study began with the administrator explaining the dataset, the three systems and the goals of the study. After a demographics survey, participants started to perform tasks using a designated system. The administrator timed all the tasks. At the end of each task, each participant was asked to fill out a survey questionnaire about the usefulness, usability of each system and his/her overall satisfaction. All responses to the survey questions were rated using a five-point Likert scale. At the end of the study, we also asked a few open-ended questions to gather additional information such as which VISA features they liked the most, which VISA features they liked the least and what features they would like to have in VISA.

### 6.2.3 Result Analysis

In this section, we first summarize our quantitative evaluation results, followed by a brief description of the survey results.

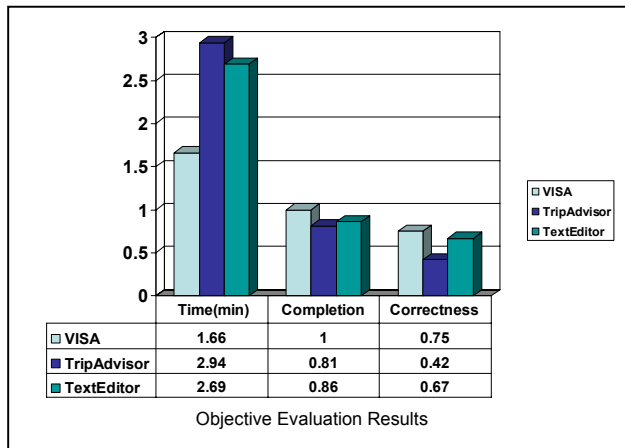


Figure 6: User experiment objective results.

#### Quantitative Evaluation Results

We used three objective metrics to assess how effective users can accomplish their tasks using three different systems: *Elapsed time* (in minutes), *task completion rate* and *task correctness*. As shown in the Fig. 6, on average, users interacted with VISA can finish their tasks faster, with a higher task completion rate and a higher

level of correctness. Based on an independent sample t-test, the differences of the means between VISA and TripAdvisor on all three objective measures were statistically significant (for Elapsed time,  $p < 0.007$ , for task completion  $p < 0.006$ , for task correctness,  $p < 0.004$ ). Moreover, VISA was also consistently performed better than the text editor on all three metrics. The differences of means for elapsed time ( $p < 0.05$ ) and task completion rate ( $p < 0.023$ ) were also statistically significant. Interestingly, people performed their tasks consistently better when they used the text editor than when they used TripAdvisor. The difference of means for task correctness was statistically significant ( $p < 0.034$ ) between TripAdvisor and the text editor. To our surprise, the text editor performed better than TripAdvisor. We expected the text editor to do worse than others because it is not specially designed for travel reviews. We finally found the reason of TripAdvisor's poor performance is it does not support search function on the reviews' content.

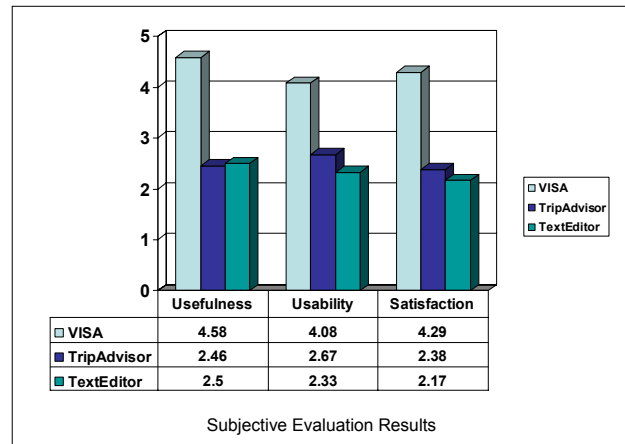


Figure 7: User experiment subjective results.

#### Survey Results

We also asked users to answer three survey questions after they finished each of their tasks using each of the three systems. As shown in the Fig. 7, users rated VISA most favorable with regard to all three survey questions. For example, they strongly agreed that VISA was useful and usable in helping them finish their tasks (mean scores were consistently above 4). They also reported highest overall satisfaction when interacting with VISA. In contrast, users in general had negative views towards TripAdvisor (All its scores were consistently lower than 3, the neutral score). Based on an independent sample t-test, the differences of means between VISA and TripAdvisor on all three subjective evaluation metrics were statistically significant with  $p = 0.000$ . Moreover, even though the Text Editor seemed to help users accomplish their tasks better than TripAdvisor based on the objective evaluation metrics, user's view towards the text editor was equally negative w.r.t. to its usefulness, usability and their overall satisfaction. There was no statistically significant difference of the means between TripAdvisor and the text editor on all three metrics.

#### Discussion and Implication

A post questionnaire contains open-ended questions for the text-based responses to VISA. When asked whether VISA performed better, and if yes, why VISA performed better than TripAdvisor and the text edit tool? Participants made it clear that the "mash-up



visualizations” and “rich interactions” directly contribute to the better performance of VISA: “Mash-up visualizations provide more information and it’s quite intuitive”, “rich interactions make it easy to search what I want to know”. Participants also gave some suggestions on what could be improved for VISA, “it now needs some learning efforts to use VISA”, “It could introduce better UI design and richer interactions”.

## 7. CONCLUSION

In this paper, we present VISA, a visual sentiment analysis system combining a generic sentiment analysis paradigm based backend processing engine and a coordinated multiple view based frontend sentiment visualization tool. The novel concept of sentiment tuple allows the system to take advantage of most cutting-edge opinion mining algorithms. The visualization design differentiates with existing text visualization approaches by several sentiment-specific optimizations. The system is evaluated with applications from both the traditional consumer-oriented website and the modern social media in representing the sentiments within the text collections. User experiments show that VISA significantly outperforms two baseline systems in both objective and subjective metrics.

## 8. ACKNOWLEDGEMENT

This work is supported in part by the National Basic Research Program of China (973 Program) under Grant No. 2010CB328105, National Science Foundation of China (NSFC) under Grant No. 60932003 and 60973144.

## 9. REFERENCES

- [1] B. Pang and L. Lee. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2):1-135, 2008.
- [2] B. Pang, L. Lee and S. Vaithyanathan. Thumbs up? : Sentiment classification using machine learning techniques. In *Conference on Empirical methods in natural language processing*, pp. 79-86, 2002.
- [3] M. Hu and B. Liu. Mining and summarizing customer reviews. In *ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 168 – 177, 2004.
- [4] M. Gamon, A. Aue, S. Corston-Oliver, and E. Ringger. Pulse: Mining customer opinions from free text. In *proceedings of the International Symposium on Intelligent Data Analysis (IDA)*, pp. 121-132, 2005.
- [5] D. Oelke, M. Hao, C. Rohrdantz, D. A. Keim, U. Dayal, et al. Visual Opinion Analysis Customer feedback data. In *IEEE Symposium on Visual Analytics Science and Technology*, pp. 187-194, 2009.
- [6] L. Zhuang, F. Jing, X. Yan Zhu, and L. Zhang. Movie Review Mining and Summarization. In *Proceedings of the 15<sup>th</sup> ACM international conference on Information and Knowledge Management (CIKM)*, pp. 43-50, 2006.
- [7] W. Zhang, C. Yu and W. Meng. Opinion retrieval from blogs. In *Proceedings of the 15<sup>th</sup> ACM international conference on Information and Knowledge Management (CIKM)*, pp. 831 – 840, 2007.
- [8] S.-M. Kim and E. Hovy. Determining the sentiment of opinions. In *Proceedings of international conference on Computational Linguistics*, pp. 1367-1373, 2004.
- [9] M. Hu and B. Liu. Mining opinion features in customer reviews. In *Proceedings of the 19<sup>th</sup> national conference on Artificial intelligence (AAAI)*, pp. 755-760, 2004.
- [10] A. -M. Popescu and O. Etzioni. Extracting product features and opinions from reviews. In *Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pp. 339-346, 2005.
- [11] Q. Mei, X. Ling, M. Wondra, H. Su, and C. Zhai. Topic sentiment mixture: modeling facets and opinions in weblogs. In *Proceedings of the 16<sup>th</sup> international conference on World Wide Web (WWW)*, pp. 171–180, 2007.
- [12] C. Lin and Y. He. Joint sentiment/topic model for sentiment analysis. In *Proceeding of the 18<sup>th</sup> ACM conference on Information and knowledge management (CIKM)*, pp. 375-384, 2009.
- [13] K. Toutanova, D. Klein, C. D. Manning, and Y. Singer. Feature-rich part-of-speech tagging with a cyclic dependency network. In *Proceedings of the conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology (NAACL)*, pp. 173-180, 2003.
- [14] F. Wanner, C. Rohrdantz, F. Mansmann, D. Oelke, and D. A. Keim. Visual sentiment analysis of rss news feeds featuring the US presidential election in 2008. In *Workshop on Visual Interfaces to the Social and the Semantic Web (VISSW)*, 2009.
- [15] C. Chen, F. Juan, E. Juan and C. Weaver. Visual Analysis of Conflicting Opinions. In *IEEE Symposium on Visual Analytics Science and Technology (VAST)*, pp. 59-66, 2006.
- [16] M. L. Gregory, N. Chinchor, O. Whitney, R. Carter, E. Hetzler, and A. Turner. User-directed sentiment analysis: Visualizing the affective content of documents. In *Workshop on Sentiment and Subjectivity in Text*, pp. 23-30, 2006.
- [17] M. Gamon, S. Basu, D. Belenko, D. Fisher, M. Hurst, and A. C. Konig. BLEWS: Using blogs to provide context for news article. In *AAAI Conference on Weblogs and Social Media*, pp. 60-67, 2008.
- [18] C. Rohrdantz, T. Mayer, M. Butt, F. Plank and D. A. Keim. Comparative Visual Analysis of Cross-Linguistic Features. In *IEEE Symposium on Visual Analytics Science and Technology (VAST)*, pp. 1-7, 2010.
- [19] N. Diakopoulos, M. Naaman, F. Kivran-Swaine. Diamonds in the Rough: Social Media Visual Analytics for Journalistic Inquiry. In *Visual Analytics Science and Technology (VAST)*, pp. 115-122, 2010.
- [20] D. Fisher, A. Hoff, G. Robertson and M. Hurst. Narratives: A Visualization to Track Narrative Events as they Develop. In *IEEE Symposium on Visual Analytics Science and Technology (VAST)*, pp. 115-122, 2008.
- [21] F. Wei, S. Liu, Y. Song, S. Pan, M. X. Zhou, W. Qian, L. Shi and Q. Zhang. TIARA: A Visual Exploratory Text Analytic System. In *Proceedings of the 16<sup>th</sup> ACM international conference on Knowledge discovery and data mining (KDD)*, pp. 153-162, 2010.
- [22] S. Liu, M. X. Zhou, S. Pan, W. Qian, W. Cai and X. Lian. Interactive, Topic-based Visual Text Summarization and Analysis, In *ACM conference on Information and knowledge management (CIKM)*, pp. 543-552, 2009.
- [23] L. Shi, F. Wei, S. Liu, L. Tan, X. Lian and M. X. Zhou. Understanding Text Corpora with Multiple Facets. In *IEEE Symposium on Visual Analytics Science and Technology (VAST)*, pp. 99-106, 2010.