PROJECT MUSE®

A Sentiment-Based Digital Library of Movie Review Documents
Using Fedora / Une bibliothèque numérique de documents
critiques de films basée sur les sentiments en utilisant
Fedora

Jin-Cheon Na, Tun Thura Thet, Arie Hans Nasution, Fauzi Munif Hassan

Canadian Journal of Information and Library Science, Volume 35,
Number 3, September/septembre 2011, pp. 307-337 (Article)

Published by University of Toronto Press

DOI: <https://doi.org/10.1353/ils.2011.0018>

➔ *For additional information about this article*

<https://muse.jhu.edu/article/450030>

A Sentiment-Based Digital Library of Movie Review Documents Using Fedora

Une bibliothèque numérique de documents critiques de films basée sur les sentiments en utilisant Fedora

Jin-Cheon Na, Tun Thura Thet, Arie Hans Nasution, and Fauzi Munif Hassan
Wee Kim Wee School of Communication & Information
Nanyang Technological University
31 Nanyang Link, Singapore 637718
tjcna@ntu.edu.sg

Abstract: This study develops a digital library of movie review documents that supports sentiment-based browsing and searching. Firstly, we develop an automatic method for in-depth sentiment analysis and classification of movie review documents to provide sentiment orientations toward multiple perspectives of movies, such as overall opinion about the movie, director, and cast. By utilizing information extraction techniques such as entity extraction, co-referencing, and pronoun resolution, the review texts are segmented into multiple sections where each section contains multiple sentences and discusses a particular aspect of the reviewed movie. For each aspect section, a machine-learning algorithm, Support Vector Machine (SVM), is applied to determine sentiment orientation toward the target aspect. Secondly a prototype digital library is developed with the automatically analysed data to show the usefulness of sentiment-based browsing and searching. Using the system, the user can browse and search movies by sentiment polarity (positive, neutral, or negative) of multiple aspects in the movie. Finally, a usability evaluation is conducted to observe the effectiveness of the sentiment-based digital library.

Keywords: Sentiment Analysis, Sentiment Summarization, User-Generated Content, Digital Libraries, Movie Reviews

Résumé : Cette étude examine le développement d'une bibliothèque numérique de documents critiques de films permettant l'exploration et la recherche par sentiments. Pour commencer, nous développons une méthode automatique pour l'analyse en profondeur des sentiments et la classification des documents critiques de films propres à fournir des orientations à propos des sentiments capables d'offrir des perspectives multiples sur les films, comme par exemple l'opinion générale sur le film, sur le metteur en scène, et sur les acteurs. Grâce à

l'utilisation de techniques d'extraction d'information telles que l'extraction d'entités, le co-référencement, et la résolution de pronoms, les comptes rendus sont segmentés en de multiples sections où chacune contient plusieurs phrases et aborde un aspect particulier du film en question. À chacune de ces sections on applique un algorithme d'apprentissage automatique, *Support Vector Machine (SVM)*, qui détermine l'orientation du ou des sentiments pour cette section. Ensuite, nous développons un prototype de bibliothèque numérique en utilisant les données analysées automatiquement afin de montrer l'utilité de l'exploration et de la recherche par sentiments. En utilisant ce système, l'utilisateur peut explorer et faire des recherches dans les films selon les polarités des sentiments (positif, neutre, ou négatif) et ce, sur de nombreux aspects des films. Pour finir, nous avons effectué une évaluation d'utilisabilité afin de vérifier l'efficacité d'une bibliothèque numérique basée sur les sentiments.

Mots-clés : Analyse par sentiments, rédaction de résumés par sentiments, contenus générés par l'utilisateur, bibliothèques numériques, critiques de films

Introduction

The emergence and wide use of digital libraries impose many new challenges. It is increasingly important to innovate ways of organizing information and supporting the user to find relevant documents easily. Digital libraries also provide good opportunities as test beds for exploring these issues with various kinds of digital documents. Therefore, researchers are developing diverse digital libraries to explore novel ways of organizing, browsing, and searching digital documents, not only by standard meta-data fields but also by aspects of documents through automatic content analysis. Thus this study develops a prototype sentiment-based digital library to explore the usefulness of automatically analysed sentiment information in the digital library.

Research in automatic text classification seeks mostly to develop models (text classifiers) for assigning category labels to new documents based on a set of training documents. For classification, documents are represented as sets of features from their content and style, called document vectors. Most studies of automatic text classification have focused on either "topical classification" classifying documents by subject or topic (e.g., education vs. entertainment), or "genre classification" classifying documents by document styles (e.g., fiction vs. non-fiction). A detailed introduction to automatic text classification has been provided by Sebastiani (2002).

In recent years, we have witnessed tremendous growth of online discussion groups and review sites, where an important characteristic of the posted articles is their sentiment or overall opinion about the subject matter. Researchers are turning their attention to sentiment classification, a kind of non-topical classification. Though machine-learning techniques have long been used in topical text classification with good results, they are less effective when applied to sentiment classification (Pang and Lee 2008). Sentiment classification is a more difficult task compared to traditional topical classification, which classifies articles by comparing individual words (unigrams) in various subject areas. A challenging aspect of sentiment classification that distinguishes it from traditional topic-based classification is that while topics are often identifiable by keywords alone, sentiment can be expressed in a more delicate expression. For example, the sentence “Who would vote for this presidential candidate?” contains no single word that is obviously negative. Sentiment classification requires more understanding than the usual topic-based classification.

Recently researchers have been working beyond binary classification of positive and negative sentiments which predicts an overall sentiment of a review document (Blair-Goldensohn et al. 2008; Hu and Liu 2004; Zhuang, Jing, and Zhu 2006). They perform more in-depth sentiment analysis of review texts. For instance, the reviews of a pop song cover not only overall sentiment but also many specific aspects such as vocals, lyrics, recording quality, creativity, and so on. For sentiment analysis of review documents such as music and movies, more in-depth sentiment analysis is essential, and various aspects in the review document should be considered in addition to the binary classification of overall sentiment. For example, a reviewer may like some aspects, such as director’s performance, of a movie but not all.

We investigate movie review documents in this study mainly because there is a large amount of review data available on the Internet and such reviews are challenging. Movie reviews are believed to be more challenging than other reviews such as product reviews (Turney 2002). Separate consideration for different aspects of movie reviews, such as opinions on cast, director, storyline, and animation, provides better sentiment analysis of the movie.

In order to classify sentiments about multiple perspectives of a movie, sentences in a review text are tagged into different movie aspects (such as director or cast) so that the sentences in each aspect can be analysed

and processed independently. For this purpose, information extraction techniques such as entity annotation, co-referencing, and pronoun resolution are employed. Then the automatically tagged sentences in each aspect section are also reviewed by two manual coders in order to verify the effectiveness of our automatic sentence-tagging approach. The independent coders also read the tagged sentences in each aspect section and manually code a sentiment orientation toward the target aspect. Inter-coder reliability is verified, since the manually coded information is used as the training data for the supervised machine learning. This initial work is described in Thet, Na, and Khoo (2008), and also summarized in the section “Sentiment analysis and classification” below.

After the sentiment orientations of the movie reviews are automatically analysed, a prototype digital library is developed to show the usefulness of sentiment-based browsing and searching. The sentiment information is stored as the metadata of movie digital objects. Then the system allows sentiment-based search of movie digital objects by specifying sentiment polarity (positive, neutral, or negative) toward the overall opinion about the movie, director, or cast. An example would be a search of movies where directors have received positive sentiment reviews. Sentiment-based browsing allows the user to browse movies, directors, and cast (actors/actresses) alphabetically. While searching and browsing, the system displays sentiment-categorized results. In this way, users can focus on movie objects in their preferred sentiment category only. We expect that the sentiment-based browsing and searching would be a standard feature in future digital libraries of social media content (e.g., expert reviews, user reviews, blog postings, and discussion board postings) as they enhance the usability of the digital library.

In the following sections, “Related works” introduces related approaches for sentiment analysis and opinion mining, and “Sentiment analysis and classification” describes our sentiment analysis method for the movie review documents. The section “A sentiment-based movie-review digital library” gives details of a developed prototype digital library and its user evaluation results. Finally, the section “Discussion and conclusion” discusses future works and concludes the paper.

Related works

Many researchers have carried out studies of automatic sentiment analysis and opinion mining (Pang and Lee 2008). For instance, Pang, Lee,

and Vaithyanathan (2002) examined the effectiveness of three machine-learning methods (Naïve Bayes, Maximum Entropy, and Support Vector Machines [SVM]) for the sentiment classification of movie reviews. They used mainly features based on unigrams (with negation tagging) and bigrams. SVM returned the best results (82.9% accuracy), using unigrams with binary weighting indicating the presence or absence of a feature. Yi et al. (2003) proposed a method of sentiment analysis using natural language-processing techniques to extract positive and negative sentiments for specific subjects from a document, instead of classifying the whole document into positive or negative. They used semantic analysis with a syntactic parser and sentiment lexicon. The prototype system achieved good results for web pages and news articles. Another work by Pang and Lee (2004) proposed a machine-learning method by applying text-categorization techniques to the subjective portions of the document only. They examined the relation between subjectivity detection and polarity classification, showing that subjectivity detection can compress reviews into much shorter extracts that still retain polarity information at a level comparable to that of the full review. They argued that utilizing contextual information via the minimum-cut framework could lead to statistically significant improvement in polarity-classification accuracy.

Some researchers have worked on sentiment summarization of multiple documents. Hu and Liu (2004) mined and summarized the customer reviews of electronic products, such as digital camera, cellular phone, and MP3 player. They extracted the features or aspects (such as picture quality and screen size) of the product on which the customers have expressed their opinions, and predicted whether each opinion sentence is positive or negative. Zhuang, Jing, and Zhu (2006) proposed a method for opinion mining and summarization of movie reviews. The method found mainly feature-opinion pairs from text reviews. For example, in the sentence “The sound effects are excellent,” the feature term is *sound effects* and the opinion word is *excellent*. Some of the feature classes for summarization include overall, screenplay, producer, director, screenwriter, and actor and actress. The goal of this approach was to find feature-opinion pairs and to identify their polarity and the feature classes of the opinions in order to produce a structured sentence list as the summary. Blair-Goldensohn et al. (2008) presented a system that summarizes the sentiment of reviews for a local service such as a restaurant, department store, or hotel. The set of service reviews returned with a local service result on Google Maps were used as input data. The system extracted relevant aspects of a service, such as service, ambiance, or

value, aggregated the sentiment per aspect, and showed aspect-relevant text with sentiment polarity values (positive or negative). Compared to these methods, our approach uses document-level (or paragraph-level) sentiment classification of multiple aspect sections without a lexicon of sentiment words necessary for sentence-level sentiment classification.

Some researchers have developed classification or clustering tools to categorize web search results to help users locate relevant and useful information on the World Wide Web. For the classification or clustering they generally use the snippets from the search engine to provide reasonable response time to the user. Chen and Dumais (2000) designed a user interface that automatically groups web search results into predefined topical categories such as *automotive* and *local interest*, using a machine-learning algorithm, SVM. The method devised by Zeng et al. (2004) provides clustering of web search results and uses salient phrases extracted from the ranked list of documents as cluster names. For instance, with a query input *Jaguar*, the generated cluster names are *Jaguar Cars*, *Panthera onca*, *Mac OS*, *Big Cats*, *Clubs*, and *Others*. Vivisimo is an example of an operational clustering tool for web search results. These tools, however, focus mainly on topical categorization—categorizing documents by subject or topical area.

In our previous work (Na et al. 2005), a prototype meta search engine providing automatic sentiment classification was developed, and it used both snippets and their full-text documents in the classification. It allows the user to specify a product name and subsequently categorizes the search results by the polarity of the desired reviews: *recommended* or *not recommended*. It can help the user to focus on web articles containing either positive or negative comments. For instance, a user who is interested mainly in the negative aspects of a product (e.g., a digital camera) can look at web articles under *not recommended* review category. Since the prototype is a meta search engine, it is tightly dependent on existing search engines and has less control for managing original documents. Thus we have developed a digital library of our own documents (i.e., movie reviews) to investigate the usefulness of sentiment-based browsing and searching.

Sentiment analysis and classification

We conducted sentiment analysis with a data set of 876 movie review documents—438 positive and 438 negative. The movie review docu-

ments were harvested from the movie review site ReelViews, using a web crawler. The movie reviews have been written by a movie critic, and most of them comprise five to seven paragraphs of text. An average length of the review documents in our data set is thirty sentences and they discuss several aspects of the movie. In general, these sentences contain overall opinion about the movie as well as specific opinions about director, cast, or other aspects. According to Ruppenhofer, Somasundaran, and Wiebe (2008), the identification of sources (opinion holders) and targets (topics or aspects) of opinions is an important sub-task for automatic opinion-analysis systems. Our approach focuses on targets for identifying the various aspects of movie reviews. The metadata of review documents such as rating, movie title, and director and cast names are also extracted by parsing the HTML codes of the review documents.

Sentence tagging

We apply information-extraction techniques to tag sentences in a movie review. The cast and director names of the movie are first annotated, since they are important indicators in determining specific aspects of the movie in current sentences. Co-referencing is used to determine which named entities have the same referent (Bontcheva et al. 2004). For example, it determines whether *Jim Carrey* and *Carrey* refer to the same entity. Pronoun or anaphora resolution is also used to address the problem of resolving what a pronoun or a noun phrase refers to in the text. For example, it is used to determine which entity the pronoun *he* in the sentence “He has been quite disappointing” is referring to. It could be referring to a director or an actor. In this study, a simple approach is used to support pronoun resolution. For instance, if a director name is mentioned in previous sentences, *he* is tagged with the director name.

A rule-based approach is used for automatic tagging of sentences into different aspect sections: overall, cast, director, storyline, or others. We use our developed sentence-tagging tool in Java, which performs entity annotation by utilizing the extracted metadata, and then sentence tagging, as illustrated in figure 1, which also shows the distribution of automatically tagged sentences in the review texts.

The rules for automatic sentence tagging are shown in figure 2. Each sentence is compared with conditions in the rules, and if there is a match, the corresponding tag is assigned to the sentence. For instance, the rule in lines 3–4 fires if a sentence matches one of the predefined

Figure 1: Overall process for sentence tagging

```

1  FOR each review document
2    FOR each sentence in a review document
3      IF specific patterns recognized
4        // e.g., a pattern for storyline is found
5        sentence tag = corresponding tag // e.g., assign storyline tag
6      ELSE IF sentence contains director names OR co-references
7        sentence tag = director
8      ELSE IF sentence contains cast names OR co-references
9        sentence tag = cast
10     ELSE IF sentence contains director pronouns
11       sentence tag = director
12     ELSE IF sentence contains cast pronouns
13       IF sentence contains "This Movie" OR "Reviewer-Reader" OR "Casting"
14         terms
15         sentence tag = cast
16       ELSE
17         sentence tag= storyline
18     ELSE IF sentence contains "This Movie" OR "Reviewer-Reader" terms
19       sentence tag = overall
20     ELSE
21       sentence tag = others
22   END FOR
23 END FOR

```

Figure 2: Rules for automatic sentence tagging

specific patterns for sentence tagging. The specific patterns are defined for various tags and tend to be domain specific and depend on the reviewer's writing style. In our movie reviews, the sentence having an actor/actress name in parentheses after a character name represents storyline. For instance, "Katie Burke (Katie Holmes) is bringing to a close a highly satisfactory run at a prestigious college" is tagged as storyline. The rules also use semantic information defined by feature lists, and co-referencing and pronoun resolutions for identifying the same entity, such as a director and an actor/actress.

Table 1: Sample entries of feature lists

"This Movie" Terms	"Reviewer-Reader" Terms	"Casting" Terms
this movie	I	excel perform
this film	me	transform
the movie	my	tribute
the film	you	has/have created
movies of this kind	your	subtle
(9 entries)	(9 entries)	(45 entries)

Table 1 shows sample entries of feature lists, which are used by the rule engine for the conditions in lines 12 and 16. The terms for "This Movie" and "Reviewer-Reader" are prepared manually by scanning through the movie reviews. The terms for "Casting" are prepared automatically using cast and storyline sentences in the training data. Firstly, the mutual information (MI) of n -gram term t is computed through two categories (cast and storyline) as follows:

$$MI(t) = \sum_{i=1}^2 \Pr(c_i) \times MI(t, c_i) \quad (1)$$

$$MI(t, c_i) = \log \frac{\Pr(t \wedge c_i)}{\Pr(t) \times \Pr(c_i)} \quad (2)$$

where $\Pr(c_i)$ is the percentage of sentences in category c_i (either cast or storyline category) over all training sentences, $MI(t, c_i)$ is the mutual information between term t and category c_i , $\Pr(t)$ is the percentage of sentences containing term t over all training sentences, and $\Pr(t \wedge c_i)$ is the joint probability of term t and category c_i both occur. Then n -gram terms with high mutual information (MI) values are collected, and among them the terms appearing more in cast sentences than storyline are selected as "Casting" terms, which are used as indicators to separate cast sentences from storyline sentences. The terms appearing in the storyline sentences usually have low MI values because the storylines cover a variety of topics. The selection of terms could be more refined by using more training data and adjusting the threshold level (i.e., cut-off point for MI values) to extract terms.

Table 2: Inter-coder agreement between two coders

	Coder A					Row Totals
		Positive	Negative	Neutral	Not Applicable	
Coder B	Positive	630	40	42	22	734
	Negative	54	600	31	13	698
	Neutral	16	22	27	13	78
	Not Applicable	18	16	9	199	242
Column Totals		718	678	109	247	1752 (overall total)

To verify accuracy of the automatic tagging, two manual coders are asked to read the automatically tagged sentences in cast and director sections, and highlight sentences that are wrongly tagged as cast or director. The manual and automatic sentence tagging have about 90% agreement, and thus automatically tagged sentences are used for the experiments.

The two coders also manually classify cast and director sections in each movie review into one of the following sentiment classes: *positive*, *negative*, *neutral*, or not *applicable*. The *not applicable* class indicates that the current section is not related to the cast or director aspect at all. Table 2 shows inter-coder agreements between two coders. The numbers in bold indicate agreement between two coders, and the other numbers indicate disagreement.

We use Cohen's kappa coefficient in order to measure agreement between the two independent coders. The equation for Cohen's kappa (CK) is:

$$CK = \frac{\sum_{i=1}^m a_{ii} - \sum_{i=1}^m \frac{\text{row}_i \text{ total} \times \text{column}_i \text{ total}}{\text{overall total}}}{\text{overall total} - \sum_{i=1}^m \frac{\text{row}_i \text{ total} \times \text{column}_i \text{ total}}{\text{overall total}}} \quad (3)$$

where *overall total* is the total number of sections in the cast and director sections, m is the number of rows or columns (i.e., 4, the number of target categories), $\sum_{i=1}^m a_{ii}$ is the total number of agreements by summing the values in the diagonal cells of the table, and $\sum_{i=1}^m \frac{\text{row}_i \text{ total} \times \text{column}_i \text{ total}}{\text{overall total}}$ is the sum of the expected frequencies of agreement by chance.

In our experiment, the percentage of agreement (total number of agreements / overall total) is 0.83, and the inter-coder agreement using Cohen's kappa coefficient is 0.74, which is considered as a good agreement (Byrt 1996). The conflicting labels by the two coders were reviewed and reclassified by one of the authors, and these manually classified sentiment labels for cast and director sections were used as answer keys for a supervised machine-learning approach, which will be discussed in next subsections.

Overall sentiment

The star ratings given by the reviewer are used as overall sentiment orientation for the movies. The reviews are rated with a scale of five stars. Those with two stars and fewer are considered negative reviews, while those with three stars and more are considered positive reviews.

For the experiment, a Support Vector Machine SVM (Joachims 1998) is used as a supervised machine-learning algorithm. Each review text is converted into a bag of words (called a document vector), which are stemmed using Porter's stemming algorithm (Jones and Willet 1997) after removing ineffective stop words. For handling the negation of sentiment/subjective words, n -gram terms are used in addition to a bag of words in the document vector. Our approach uses a list of negation phrases, and these phrases are matched with the review text to negate associated words. A simple example is the sentence "This movie is not bad." As we can see, the sentiment word *bad* is negated. Thus, our approach uses the negation term *not bad*, instead of using two separate words, *not* and *bad*. The negation list includes not only single words such as *not*, *never*, and *no*, but also phrases such as *do not think so*, *never think*, and so on. In another example sentence "I do not think this is a good movie," the phrase *do not think* negates the adjective *good*, and thus the negation term *not good* is used.

Table 3: Accuracy of sentiment classifications for Overall aspect

ID	Term Weighting		Adjective Term Only	Stemming	Negation	Removal of Stop Words	Accuracy (%)	Sentiment	Precision (%)	Recall (%)	F-score (%)
	Presence	Frequency									
1	Y						88.31	Pos.	87.01	90.16	88.56
								Neg.	89.79	86.54	88.14
2		Y					82.01	Pos.	83.06	80.54	81.78
								Neg.	81.11	83.57	82.33
3	Y		Y				76.38	Pos.	75.08	81.25	78.04
								Neg.	79.57	73.03	76.16
4	Y			Y			88.88	Pos.	88.79	89.01	88.90
								Neg.	88.98	88.76	88.87
5	Y			Y	Y		89.00	Pos.	88.82	89.24	89.03
								Neg.	89.19	88.77	88.98
6	Y			Y	Y	Y	90.48	Pos.	90.79	90.15	90.47
								Neg.	90.22	90.85	90.54

According to automatic sentence tagging, about 27% of sentences from the review texts are tagged as *overall* and they are used for predicting overall sentiment orientations of the movie reviews. We conduct a three-fold cross-validation for this experiment. Although the method uses only 27% of the original review texts, it can predict overall sentiment orientations of the movie reviews with relatively high accuracy.

Table 3 shows the results of our experiments. The first column ID indicates an identifier for a unique feature combination, and the remaining columns represent document feature options for the experiment: using presence or frequency for term weighting, adjective terms only from the full texts, word stemming, handling of negation terms, and removal of stop words. We experimented with all combinations of document feature options, and only the accuracy of promising combinations of document feature options is reported in the paper.

As shown in table 3, when stemmed and negation terms are used, it improves the accuracy and *F*-score of overall sentiment classification (IDs 5 and 6 in table 3). It is also observed that using a standard list of

stop words for feature reduction is not ideal, since some words such as pronouns, prepositions, and subjective words carry useful meanings for sentiment classification. In fact, the accuracy of sentiment classification drops noticeably when the standard list of stop words is used. As discussed, our approach uses pronoun resolution to identify discussing the director and cast. Thus pronouns are useful information and they cannot be discarded. Words like *better*, *best*, and *unfortunately* are obviously subjective words that are useful for sentiment analysis. Thus, they should not be included in the list of stop words. Some prepositions such as *above* and *below* can also carry sentiments. For example, the word *above* in the sentence “This is well above my expectation” is useful information, and if it is replaced by *below*, the sentiment orientation would be completely reversed. Therefore, we remove some useful terms for sentiment classification from the standard list of stop words for our experiments (ID 6 in table 3). We also notice that using frequency as term weighting or adjective terms only as features reduces accuracy by a few percentages (IDs 2 and 3 in table 3).

Sentiment toward director

The manually tagged sentiment labels are used as answer keys for Director sections in the movie reviews. The experiment shows that sentiment orientation toward the director can be different from overall sentiment orientation toward the movie. A sentiment toward the director could be positive or neutral while overall sentiment toward the movie is clearly negative.

For example, the movie *Bad Girls* is rated half a star out of five stars, but the review contains positive sentiment about the director, as it says, “At least director Jonathan Kaplan had the good sense to employ a competent cinematographer.” In some reviews, a rating value shows positive sentiment toward the movie overall, but sentiment toward the movie director is clearly negative or neutral.

We conduct a threefold cross-validation, and the results show accuracy of up to 75.54% (see ID 6 in table 4). Like the overall sentiment classification, using customized stop words for feature reduction and using stemmed and negation terms as features improves the accuracy of sentiment classification toward the director, but not when using frequency as term-weighting and only adjective terms as features.

Table 4: Accuracy of sentiment classifications for Director aspect

ID	Term Weighting		Adjective Term Only	Stem- ming	Negation	Removal of Stop Words	Accuracy (%)	Sentiment	Precision (%)	Recall (%)	F-score (%)
	Presence	Frequency									
1	Y						73.25	Pos.	83.540	62.600	71.570
								Neg.	70.096	87.666	77.902
2		Y					69.10	Pos.	81.230	54.470	65.211
								Neg.	65.752	87.413	75.051
3	Y		Y				69.81	Pos.	78.930	58.540	67.223
								Neg.	67.052	84.373	74.722
4	Y			Y			74.82	Pos.	84.720	64.770	73.414
								Neg.	71.485	88.318	79.015
5	Y			Y	Y		75.25	Pos.	84.900	65.580	74.000
								Neg.	71.961	88.336	79.312
6	Y			Y	Y	Y	75.54	Pos.	85.140	65.850	74.263
								Neg.	72.158	88.507	79.501

Sentiment toward cast

Similarly we use the manually tagged sentiment labels as answer keys for Cast sections in the movie reviews. According to automatic sentence tagging, only 15% of sentences from the review texts are about cast, and they are used for predicting sentiment orientation toward the cast in each movie review. We conduct a threefold cross-validation, and results show accuracy of up to 78.74% (see ID 6 in table 5). Again the accuracy of sentiment classification toward cast is improved when using customized stop words for feature reduction and using stemmed and negation terms as features. In some cases, a Cast sentence can contain additional sentiment orientations toward different aspects, such as overall opinion about the movie. For example, the sentence “Actress Julie Delpy is far too good for this movie” carries positive sentiment toward the cast while there is some negativity toward the movie.

Error analysis

We analyse the errors encountered in both our automatic sentence tagging and sentiment classifications. There are about 10% errors in the

Table 5: Accuracy of sentiment classifications for Cast aspect

ID	Term Weighting		Adjective Term Only	Stem- ming	Neg- ation	Remov- al of Stop Words	Accu- racy (%)	Senti- ment	Perci- sion(%)	Recall (%)	F-score (%)
	Pres- ence	Fre- quency									
1	Y						76.40	Pos.	79.64	71.33	75.26
								Neg.	74.04	81.76	77.71
2		Y					70.38	Pos.	77.16	58.31	66.42
								Neg.	66.50	82.74	73.73
3	Y		Y				73.17	Pos.	80.82	60.67	69.31
								Neg.	68.52	85.60	76.11
4	Y			Y			77.71	Pos.	80.09	74.58	77.24
								Neg.	76.22	81.46	78.75
5	Y			Y	Y		77.86	Pos.	80.15	74.88	77.43
								Neg.	76.43	81.46	78.86
6	Y			Y	Y	Y	78.74	Pos.	80.60	76.06	78.26
								Neg.	77.34	81.69	79.46

automatic tagging of Director and Cast sentences and about 10–25% errors in the sentiment classifications for Overall, Director, and Cast sections with the best features combination option (ID 6 in tables 3, 4, and 5).

The following is an example sentence tagged incorrectly as a Director sentence by our automatic tagging tool. The sentence does mention the director name, Peter Howitt, but it is obviously discussing a member of the cast.

Meanwhile, Douglas McFerran, who gave a brilliant supporting performance in Peter Howitt’s Sliding Doors, plays the tough head of NURV security.

The following is an example sentence about the director that is classified incorrectly as a negative sentiment by the machine-learning approach. The words *worst* and *pedestrian nature* may appear negative but they represent positive meaning with contextual words. For instance, the word *justice* changes the sentiment orientation of *pedestrian nature* to

positive. This is one of the limitations of the machine-learning approach, which uses just bags of words and simple negation terms. Deeper linguistic analysis in addition to the bag-of-words approach will be helpful in addressing such problems.

Todd Graff's script is television-quality writing at its worst, and the direction by Ken Kwapis, who made the better *Dunston Checks In*, does justice to the screenplay's pedestrian nature.

In the following example sentence, there is a mixture of positive and negative sentiments about two actors: Tony Danza and Joseph Gordon-Levitt. This can be a difficult case not only for the machine-learning approach but also for human coders to decide whether it is positive, negative, or neutral toward the cast.

Suffice it to say that Tony Danza gives one of the most impressive performances, and young Joseph Gordon-Levitt has serious credibility problems.

As observed, some of the cases encountered are rather complex for the automatic approaches to tag or classify them with high accuracy, unless more advanced natural-language processing techniques are employed.

A sentiment-based movie-review digital library

Using the results of the sentiment analysis and classification discussed in the previous section, a prototype movie-review digital library is developed to support sentiment-based browsing and searching of movie, director, and cast digital objects. After the development of the digital library, we evaluate the effectiveness of the user interface in helping the user to find information of interest. The following subsections discuss details of the developed digital library and its user evaluation results.

System design

The high-level architectural design of the movie-review digital library is shown in figure 3, where the digital library consists of two main modules: a front-end web application and a back-end repository. We use Fedora (the Flexible Extensible Digital Object Repository Architecture) 3.0 as repository software, and J2EE (Java 2 Platform Enterprise

Figure 3: High-level architecture of movie-review digital library

Edition) to build the web application that serves as the front-end. Fedora is open-source digital repository software (Staples, Wayland, and Payette 2003), capable of serving as a digital content repository for a wide variety of uses, such as digital libraries, institutional repositories, digital archives, and content-management systems. It can store digital content items such as documents, videos, data sets, computer files, and images, and can also store metadata about the content items in various formats. In our digital library, the back-end Fedora repository is fed with the movie review data.

The front-end web application is developed using J2EE 5 technology and calls the web services provided by Fedora for accessing and managing digital objects stored in Fedora. The calling process is using Simple Object Access Protocol (SOAP)–based and RESTful type of call. As shown in figure 3, upon receiving a request from the user, the web application calls the web services. Fedora then returns data from its repository to the web application, which will process the data and show them in an appropriate way to the user.

Fedora digital objects can be related to other Fedora objects in many ways. For example, there may be a Fedora object that is considered a part of another object, a derivation of another object, a description of

Figure 4: Relationships of movie's digital objects

another object, or even equivalent to another object. The movie review data stored in Fedora make use of the object relationship capability in Fedora. We have three main types of digital objects in the system: Movie, Director, and Cast types. The digital objects relationship is depicted in figure 4. For instance, a digital object of Movie type has exactly one director and at least one cast.

An object in Fedora stores its digital resources and metadata in data streams. These digital resources can be text, image, video, or audio, while the metadata include relevant attributes that describe an object, such as title, date of publication, author, and so on. Fedora supports Dublin Core metadata (Dublin Core Metadata Initiative) for describing an object. Fedora, in fact, creates a Dublin Core data stream by default upon creation of an object. Hence, we use the Dublin Core to store some metadata related to Movie, Director, and Cast objects.

Dublin Core metadata provide only placeholders for storing basic metadata of an object. Thus, we need to create another data stream to store other types of information, especially an object's digital resources and sentiment data. The Info data stream is created to store movie review text, rating, and sentiment data. Since the data stream is implemented as Extensible Markup Language (XML) in Fedora, these data are stored as text within XML tags.

Object relationships in Fedora are asserted from the perspective of one object to another object, as in the following general pattern:

<SourceObject> — <RelationshipProperty> — <TargetObject>

Object-to-object relationship is stored using a special data stream in a digital object. This data stream is known by the reserved data-stream identifier of RELS-EXT (which stands for Relationships-External). Each digital object can have one RELS-EXT data stream that is used exclu-

Table 6: Object relationships in RELS-EXT data stream

Source Object	Relationship Name	Target Object	Multiplicity Constraint
Movie	isDirectedBy	Director	Exactly one (1)
Movie	isCastedBy	Cast	At least one (1 . . . *)
Director	isDirectorOf	Movie	At least one (1 . . . *)
Cast	playsIn	Movie	At least one (1 . . . *)

Table 7: Data-stream usage by each object type

Object Type	Dublin Core Data Stream	Info Data Stream	RELS-EXT Data Stream	Image Data Stream
Movie	Yes	Yes	Yes	Yes
Director	Yes	No	Yes	Yes
Cast	Yes	No	Yes	Yes

sively for asserting digital object relationships. In our digital library, the RELS-EXT data stream is used to specify the relationship between Movie and Director and that between Movie and Cast object. In the RELS-EXT data stream of Movie object, the digital object identifier of Director or Cast is specified.

Object relationships operate in one direction only, so we need to specify the relation of movie to director, director to movie, movie to cast, and cast to movie. This two-way referencing enables movies to be traced from the Director or Cast objects. This will be helpful, for example, when we need to search the list of movies directed by a certain director, or movies in which a particular cast member plays. In the RELS-EXT data stream of director and cast, we also specify the digital object identifier of Movie object directed by, or played in, the director and cast respectively. The object relationships in our digital library are specified in table 6.

In the digital library, we also store images of movie, director, and cast. These images are stored in the data stream named Image of type “externally referenced content.” In the Image data stream of a digital object, Fedora does not store the actual content, but rather a reference to an external content source or Uniform Resource Locator (URL). Table 7 shows the data streams used by each object type.

Movie Rev - Fed Fedora 3.0

The trends are here ... Search Categories: ☒ Movie ☐ Director ☐ Cast ☐ All

Movie Review Digital Library

HOME **MOVIES** **DIRECTORS** **CASTS** **ADVANCED SEARCH** **ABOUT US**

Movies

[Lightning Jack](#), [Leave it to Beaver](#), [Krippendorfs Tribe](#), [A Knight's Tale](#), [The King and I \(1999\)](#), [Killer Condom \(Kondom des Grauens\)](#), [One, Two, Once Upon a Forest](#), [On the Line](#), [Junior](#), [Kinship](#), [Just the Ticket](#), [The Juror](#), [Johnny Mnemonic](#), [Johnny English](#), [John Tucker Must Die](#), [Indian Summer](#), [In the Soup](#), [Ill Be Home for Christmas](#), [Jeepers Creepers 2](#), [Jason Goes to Hell: The Final Friday](#), [In Losing You](#), [The Haunting \(1999\)](#), [Home Alone 3](#), [Holy Smoke](#), [Hitcher](#), [The \(2007\)](#), [House Arrest](#), [Heart and Souls](#), [Hostel Part II](#), [Godsend](#), [Housequest](#), [House on Haunted Hill \(1999\)](#), [Alien Vs. Predator](#), [Gossip](#), [Ghosts of Mars](#), [Goodbye 20th Century](#), [The Glimmer Man](#), [Get Carter \(2000\)](#), [Ella Enchanted](#), [The Firm](#), [The Astronaut's Wife](#), [Fire in the Sky](#), [Eye For an Eye](#), [Father Hood](#), [Fatal Instinct](#), [Fast and the Furious](#), [The Tokyo Drift](#), [Assen Extreme](#), [The Faculty](#), [Eye of the Beholder](#), [Flawless](#), [The First Wives Club](#), [Fido](#), [101 Dalmatians \(1996\)](#), [Duke Where's My Car?](#), [Ace Ventura](#), [The 8th Man](#), [Life with Mikey](#), [Kung Pow: Enter the Fist](#), [Kull the Conqueror](#), [Knight Moves](#), [Barb Wire](#), [Jungle 2 Jungle](#), [Judgment Night](#), [Lake Placid](#), [Kickin it Old Skool](#), [Inspector Gadget \(1999\)](#), [The In Crowd](#), [Bad Girls](#), [Ace Ventura When Nature Calls](#), [Into the Blue](#), [Jade](#), [Idle Hands](#), [The Jar](#), [Hush](#), [High School High](#), [Hideaway](#), [Hannibal Rising](#), [Gitter](#), [Garfield: The Movie](#), [Getting Away With Murder](#), [August Rush](#), [Final Destination](#), [Fifty/Fifty](#), [The Fan](#), [Eurotrip](#), [AnfTrust](#), [Angels in the Outfield](#), [Bad Moon](#), [I Woke Up Early the Day I Died](#), [Hexed](#), [Grandmas Boy](#), [Excessive Force](#), [Even Cowgirls Get the Blues](#), [Fair Game](#), [3000 Miles To Graceland](#), [Hole in My Heart](#), [A Halloween](#), [The Curse of Michael Myers](#), [Freddie Got Fingered](#), [Ghost in the Machine](#), [Ernest Rides Again](#), [more...](#)

Directors

[Dean Semler](#), [Sydney Pollack](#), [Tony Scott](#), [Tim Story](#), [Tommy O'Haver](#), [Jason Friedberg](#), [Aaron Seltzer](#), [Stefan Fangmeier](#), [John Cherry](#), [James Wong](#), [David Richard Ellis](#), [Stephen T. Kay](#), [Harvey Miller](#), [Rachel Talalay](#), [John Moore](#), [Peter Hewitt](#), [Daniel Sackheim](#), [John Gray](#), [Vondie Curtis-Hall](#), [Tom Green](#), [Aleksander Popowski](#), [Darko Mitrevski](#), [Christopher Cain](#), [John Carpenter](#), [Mikael Salomon](#), [Davis Guggenheim](#), [Nicholas Goossen](#), [Rick Rosenthal](#), [Diane Keaton](#), [Peter Webber](#), [Alan Spencer](#), [William Malone](#), [Nick Hamm](#), [Roland Emmerich](#), [Elia Roh](#), [Tom Brady](#), [Mark S. Waters](#), [Ron Underwood](#), [Harry Winler](#), [Daisy von Scherler Mayer](#), [Chris Koch](#), [Joe Chappelle](#), [Martin Weisz](#), [Dave Meyers](#), [Lukas Moodysson](#), [Jane Campion](#), [Jan de Bont](#), [Bruce Wagoner](#), [Chris Sivertson](#), [Danny Cannon](#), [Aris Iliopoulos](#), [John Polson](#), [Brett Leonard](#), [Hart Bochner](#), [Jonathan Darby](#), [Ebrahim Forouzesh](#), [Adam Marcus](#), [James Isaac](#), [Victor Salva](#), [Rodman Flender](#), [Arlene Sanford](#), [Alexandre Rockwell](#), [Mike Binder](#), [William Friedkin](#), [John Stockwell](#), [Andrew Bergman](#), [Robert Longo](#), [Deborah Kaplan](#), [Harry Elfont](#), [Mary Lambert](#), [Brian Gibson](#), [Mark Walters](#), [Richard Wenk](#), [David Kellogg](#), [Bobby](#), [Paul M. Glazer](#), [Harvey Glazer](#), [M. Night Shyamalan](#), [Steve Miner](#), [Stephen Hopkins](#), [John Pasquin](#), [Eric Bross](#), [Charles Groves](#), [Eric Valette](#), [Martin Weitz](#), [Richard Rich](#), [Antoine Fuqua](#), [Carl Schenkel](#), [Brian Helgeland](#), [Tsai Hark](#), [Todd Holland](#), [John Nicheola](#), [Andy Cadiff](#), [James Lapine](#), [John Schultz](#), [Tom Ropelewski](#), [Keenen Ivory Wayans](#), [Troy Bever](#), [Wally White](#), [Mike Newell](#), [more...](#)

Casts

[Katie Holmes](#), [Benjamin Bratt](#), [Charlie Hunnam](#), [Zoey Deschanel](#), [Melanie Lynskey](#), [Jim Carrey](#), [Courtney Cox](#), [Noble Willingham](#), [Udo Kier](#), [Sean Young](#), [Tone Loc](#), [Dan Marino](#), [Simon Callow](#), [Ian McEwan](#), [Maynard Eziashi](#), [Bob Gunton](#), [Sophie Okonedo](#), [Christopher Lambert](#), [Nakasha Herndria](#), [Elizabeth Berondes](#), [Xavier Decle](#), [Norbert Weisser](#), [Craig Davis](#), [Kevin Zeigars](#), [Wendy Makkena](#), [Michael Jeter](#), [Bill Cobbs](#), [Eric Christmas](#), [Shane McDermott](#), [Seth Green](#), [Britney Powell](#), [Sagourney Weaver](#), [Charles Dutton](#), [Charles Dance](#), [Paul McGann](#), [Brian Glover](#), [Ralph Brown](#), [Danny Webb](#), [Lance Henriksen](#), [Chris Farley](#), [Matthew Perry](#), [Eugene Levy](#), [Kevin Dunn](#), [David Packer](#), [Lisa Riebuscia](#), [Colin Farrell](#), [Scott Caan](#), [Ali Larter](#), [Timothy Dalton](#), [Gabriel Macht](#), [Will McCormack](#), [Gregory Smith](#), [Kathy Bates](#), [Terry O'Quinn](#), [Harris Yulin](#), [Tom Everett Scott](#), [Julie Deloy](#), [Vince Vieluf](#), [Phil Buckman](#), [Julie Bowen](#), [Pierre Cosso](#), [Tom Novembre](#), [Thierry Lhermitte](#), [Joseph Gordon-Levitt](#), [Danny Glover](#), [Christopher Lloyd](#), [Brenda Fricker](#), [Tony Danza](#), [Ryan Phillippe](#), [Tim Robbins](#), [Rachael Leigh Cook](#), [Claire Forlani](#), [Tyah Runyan](#), [Red Belfrage](#), [Douglas McFerran](#), [Paul Gross](#), [Peter Berg](#), [Teri Polo](#), [Finola Hughes](#), [Johnny Depp](#), [Charlize Theron](#), [Nick Cassavetes](#), [Blair Brown](#), [Joe Morton](#), [Donna Murphy](#), [Tom Noonan](#), [Clea DuVal](#), [Freddie Highmore](#), [Keri Russell](#), [Jonathan Rhys Meyers](#), [Robin Williams](#), [Terrence Howard](#), [Leon G. Thomas III](#), [Janina Simone Nash](#), [William Sadler](#), [Ralph Fiennes](#), [Uma Thurman](#), [Sean Connery](#), [Jim Broadbent](#), [Fiona Shaw](#), [more...](#)

[About the Project](#) | [Resources](#) | [Fedora Digital Library](#) | [Disclaimer](#)

© Copyright 2009

Figure 5: Main page

Main and movie pages

The main page of the digital library is shown in figure 5. The search functionality is located in the top right corner. Users can also select which information is to be retrieved: Movie, Director, Cast, or All. Users can specify this option by selecting the radio button, located just above the Search Box. Below global navigation links, the lists of movies, directors, and cast are displayed in descending order of their review ratings

The screenshot shows the 'Movie Rev - Fed' website interface. At the top, there's a header with the site name and 'Fedora 3.0'. Below it, a navigation bar includes 'HOME', 'MOVIES', 'DIRECTORS', 'CASTS', 'ADVANCED SEARCH', and 'ABOUT US'. A search bar is also present. The main content area is titled 'Movie Review Digital Library' and features a breadcrumb trail: 'Home > Movies > B > Before Sunrise'. The 'Full Review' tab is selected, displaying a review for 'Before Sunrise' by Richard Linklater. The review text describes the film as a modern-day romance. Below the review, there's an 'Overall' section with a movie poster, a 'Movie Score' of 10, a thumbs-up icon, and the word 'Good'. A callout box points to the progress bar, stating 'Progress bar with colour code'. The 'Overall Review' section provides a summary of the film's quality.

Movie Rev - Fed Fedora 3.0

The trends are here ... Search Categories: [Movie](#) [Director](#) [Cast](#) [All](#)

Movie Review Digital Library

[HOME](#) [MOVIES](#) [DIRECTORS](#) [CASTS](#) [ADVANCED SEARCH](#) [ABOUT US](#)

[Home](#) > [Movies](#) > [B](#) > [Before Sunrise](#)

Full Review [Overall](#) [Director](#) [Casts](#) [Others](#)

Full Review

Richard Linklater, the creator of *Slackers* and *Dazed and Confused*, has succeeded where many before him have failed -- in fashioning a modern-day romance that is both original and enthralling. *Before Sunrise* is nothing short of movie magic, and the kind of film that deserves to be remembered one long year from now when 1996's Oscar nominations are handed out. Even the best romantic comedy/dramas tend to be formula-driven, frequently relying more upon actor chemistry than plot. Surprises are abo ...

[- show more -](#)

Overall

Before Sunrise

Movie Score: 10

Good

Progress bar with colour code

Overall Review

Before Sunrise is nothing short of movie magic, and the kind of film that deserves to be remembered one long year from now when 1996's Oscar nominations are handed out.

Director

Figure 6: Movie page—Full Review

so that users can see the best movies, directors, and actors/actresses. Users can click on a particular Movie, Director, or Cast link to be directed to the Movie, Director, or Cast page respectively for more detailed information.

The Movie page is displayed upon selecting one of the movies from the main page (figure 6). Five tabs display different kinds of information in the Movie page: Full Review, Overall, Director, Cast, and Others. The Overall tab contains only overall opinions about the movie, the Director tab contains the director information, the Cast tab contains the cast information, and the Others tab contains any additional information that does not fit under the Movie, Director, or Cast category such as storyline.

Figure 7: Movie page—Director

The full review is still available, so users can still choose to have all information displayed in a single page. Experienced users usually prefer to have all information displayed in one page, while novice users prefer to have the information segregated so that they will not experience information overload. By offering the Full Review tab, we can incorporate both novice and expert. Under the Full Review tab, the review is placed under its corresponding field set. For example, full review in the Full Review field set, overall review against movie in the Overall field set, and review against director in the Director field set.

A progress bar with colour code illustrates the score of the movie. When the score is low, the progress bar has only partially filled, and the colour is set to red to indicate the movie has a bad review. In contrast, if the movie is good, the progress bar shows nearly a full gauge, and the colour is green. When the quality of the movie is mediocre, the colour is amber.

Figure 7 shows the Director page when the user clicks the Director tab. The page displays the name of the director and its link to the Director page, the score of the movie in the progress bar, and sentiment toward

the director using a “thumb” icon to the right of the director image (a legend explains this icon in figure 8). The review of the director is displayed under the director image, inside a field set.

Searching movie reviews

Most digital libraries provide a search interface to easily locate digital objects stored in the digital library. In our digital library, sentiment-based search supports two different types of search: basic and advanced. In the basic search, the user can search title or name in three types of digital objects: Movie, Director, or Cast. Figure 8 shows a search result page that displays lists of movies based on the movie title query entered by the user (here, “Blue”).

A tab interface separates movies by sentiment values—positive, neutral, and negative—to assist users in viewing their sentiment of interest. For each item in the search results, a snippet of the movie is displayed. The data include:

- Truncated movie review. If users want to see full data, they can click the “[more details on this movie]” link.
- Director name and sentiment toward the director in this movie.
- Cast names and sentiment toward the cast in this movie.

The tabs give access to four sections: All, Positive, Neutral, and Negative. The All tab displays all search results clustered under positive, neutral, and negative categories. The Positive, Neutral, or Negative tab displays only reviews categorized under positive, neutral, or negative sentiment respectively.

The advanced search interface gives users more options in a sentiment-based search of movies. Similar to results in the basic search, the search result page displays sentiment-categorized search results. The advanced search input page is accessible by clicking the “Advanced Search” button in global navigation links. Figure 9 shows the advanced search interface.

The Advanced Search input page contains input controls for “Sentiment Movie,” “Sentiment Cast,” and “Sentiment Director,” which allow users to specify the sentiment of movie, cast, or director to be searched. By default, each sentiment is set to the “All” value, which includes all senti-

Search Result :

Search keyword: blue

Number of items found: 6

Search results
tabbed panel

Figure 8: Search movie result page

ments (positive, neutral, and negative). This advanced search interface enables sentiment-based searching such as:

- Searching for movies with a movie title and a certain sentiment value for the movie: for example, if users want to find good movies that contain the word *love* in the title, they can type the word in the movie title textbox, then under “Sentiment Movie” they can select the dropdown value of “positive.”

Movie Rev - Fed Fedora 3.0

The trends are here ... Search Categories: ☐ Movie ☐ Director ☐ Cast ☒ All

Movie Review Digital Library Search

HOME MOVIES DIRECTORS CASTS ADVANCED SEARCH ABOUT US

Advanced Search

Keyword

Movie Rating

Cast

Director

Sentiment

Sentiment Movie

Sentiment Cast

Sentiment Director

Search Options

Movie Genre

Ignore case? ☒

Rating sort direction

Search

[About the Project](#) | [Resources](#) | [Fedora Digital Library](#) | [Disclaimer](#)

© Copyright 2008

Figure 9: Advanced search interface

- Searching for movies with a director or cast name and a certain sentiment value of the director or cast: for example, if users may want to find movies successfully directed by “Richard Linklater,” they can fill in the director name text box, then under “Sentiment Director” they can select the dropdown value of “positive.”

Browsing movie reviews

Sentiment-based browsing allows users to browse movies, directors, and actors/actresses alphabetically. During browsing, the system displays sentiment-categorized browse results, so users can browse objects in a particular sentiment category only. To maintain viewing consistency of movie objects, we synchronized the look and feel of the movie listing view in both searching and browsing. Figure 10 shows the user interface for director browsing.

The clickable alphabet list in figure 10 allows users to jump to a set of directors whose names begin with a selected letter. The last option, “List All,” allows users to list all directors. Browse results are displayed in a tabbed panel, and the results are divided into four sentiment tabs: All (i.e. List of Directors), Positive, Neutral, and Negative. This helps

Figure 10: Browse Director page

users focus on the sentiment in which they are interested. Each tab also tallies the number of items it contains, so users easily can see which category has the most or least items.

Each item in the browse result for director or cast consists of director or actor/actress name, the number of movies in which this director or actor/actress is rated positive, neutral, and negative, and the colour-coded computed score of the corresponding director or actor/actress. The director or actor/actress name is also embedded with hyperlinks to the director or cast pages.

To calculate the sentiment score for a director, the digital library system finds all movies directed by the director, then checks how the director's sentiment is classified for each one. A weight is assigned on the basis of the sentiment value. If the sentiment is positive, a weight of 1 is added, if sentiment is neutral, a weight of 0.5 is added, while negative sentiment earns 0 weight. The accumulated weight is divided by the number of movies directed to obtain the score that reflects the director's overall per-

formance. The score is then normalized by multiplying by 10 (to get a value between 0 and 10). The score formula is given as follows:

$$Score = \left(\frac{NumOfPositive + (0.5 \times NumOfNeutral)}{NumOfPositive + NumOfNeutral + NumOfNegative} \right) \times 10 \quad (4)$$

If the scores of two directors are equal, the number of movies will be compared. The director who has directed more positive movies is deemed better than the one who has directed fewer positive movies. On the other hand, a director who has directed more negative movies is deemed to be worse than the one who has directed fewer negative movies. This formula is used similarly when computing the score of the cast. Alternatively, the scores of movies are also available by the expert reviewer and hence no computation is required.

User interface evaluation

A small heuristic usability evaluation was conducted with seven users to assess the effectiveness of the sentiment-based interface design in helping users find relevant movie objects. Users were introduced to the basic concepts and terms used in the digital library, such as movie reviews and sentiment orientations toward various aspects. Then they were asked to do the following:

- Warm up for one or two minutes, to try the look and feel of the digital library.
- Browse the movie page directly from the main page and then assess whether they understand the movie page contents.
- Search one good movie and one bad movie.
- Search a particular actor page using an actor name.
- Search directors who are rated mostly as good in the movies they have directed (directors who have high sentiment scores).

While users were performing these actions, we observed their understanding of the digital library. We asked what the icons and the progress bar represent, and whether users became aware of the Card Stack (tabs). We also asked them to use the basic and advanced search functions. Then we asked users to fill in a three-part questionnaire that evaluated

Figure 11: Functionality evaluation for directors

functionality, system performance, and user interface (aesthetics). This evaluation also included open-ended questions to explore which aspects of our digital library they considered good or bad, and also improvements that could be made.

Specifically, functionality measures the usefulness of the sentiment-based browsing and searching features. We divide functionality into Movie, Director, and Cast. As an example, the results of functionality evaluation for directors are shown in figure 11.

In the functionality evaluation results, 90.5% of the users *strongly agree* or *agree* that the sentiment-based browsing and searching features in our digital library help users find relevant movie objects (movie, director, and cast) with a desired sentiment (see table 8). Overall, our sentiment-based digital library gained users' interest. However, users also highlighted problematic aspects of our digital library. For instance, one user sug-

Table 8: Functionality evaluation results: Easiness to search/find relevant movie objects with a desired sentiment

Object Type	Easiness to Search/Find Relevant Movie Objects with a Desired Sentiment			
	Strongly Agree	Agree	Disagree	Strongly Disagree
Movie	3 (42.9%)	4 (57.1%)	0 (00.0%)	0 (00.0%)
Director	3 (42.9%)	3 (42.9%)	1 (14.2%)	0 (00.0%)
Cast	3 (42.9%)	3 (42.9%)	1 (14.2%)	0 (00.0%)
Average	3 (42.9%)	3.33 (47.6%)	0.67 (9.5%)	0 (00.0%)

gested that the search could be improved by allowing users to search and browse by movie genre as well as by sentiment.

Discussion and conclusion

Sentiment analysis of a review document should consider several perspectives when there are multiple sentiments toward different aspects of the entity under review. Our proposed method segments a review text into several sections based on its target aspects before applying the supervised machine-learning approach for sentiment classification. Our experimental results show that using customized stop words for feature reduction and using stemmed and negation terms as features improves the accuracy of sentiment classification, but not when using frequency as term weighting or only adjective terms are used as features. The best accuracy of the sentiment classification for overall movie, director, and cast is 90.48%, 75.54%, and 78.74% respectively.

One limitation of our study is that it is conducted with movie review documents harvested from only one movie review site and it seems to be relatively domain specific. More evaluations and experiments are to be carried out with larger data sets, a wider range of reviewers, and different domains. Our sentiment analysis approach might also be appropriate where thorough research is needed to make a more sustentative decision, such as a stock investment. Another limitation is that we do not employ advanced information-extraction techniques, which could produce higher accuracy in sentence tagging. For future work, advanced information extraction tools, such as General Architecture for Text Engineering: A Nearly New Information Extraction System (GATE-ANNIE) (Cunningham et al. 2002), are to be employed to analyse multiple perspectives of the movie reviews using in-depth syntactic and semantic information-extraction patterns.

In this study we also developed a prototype sentiment-based digital library using the results of the sentiment analysis and classification of the movie reviews. To measure the effectiveness of the digital library, we performed a user-interface evaluation with seven users. The result indicates that most users agree that the sentiment-based browsing and searching features are useful to find relevant movie objects with a desired sentiment. According to Nielsen and Molich (1990), to conduct heuristic evaluation, the most favourable number of evaluators is five, and at least

three. They argue that the ratio of benefits to costs decrease after reaching ten evaluators, and no significant usability problems are found by adding additional evaluators. Even so, for future work, a further user evaluation with a larger sample size of evaluators could give us more representative feedback on the usability of the sentiment-based searching and browsing interface. Another possible area of improvement is to add a social computing aspect to the digital library by allowing users to contribute comments and reviews, and then have the digital library perform sentiment value computation based on community inputs. It will allow users to have the option of viewing opinions of domain experts as well as the community.

References

- Blair-Goldensohn, S., K. Hannan, R. McDonald, T. Neylon, G. Reis, and J. Reynar. 2008. Building a sentiment summarizer for local service reviews. Paper presented at WWW 2008 Workshop: NLP Challenges in the Information Explosion Era (NLPiX2008), Beijing, China.
- Bontcheva, K., V. Tablan, D. Maynard, and H. Cunningham. 2004. Evolving GATE to meet new challenges in language engineering. *Natural Language Engineering* 10 (3–4): 349–73.
- Byrt, T. 1996. How good is that agreement? *Epidemiology* 7 (5): 561.
- Chen, H., and S.T. Dumais. 2000. Bringing order to the web: Automatically categorizing search results. *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'00)*, 145–52. New York: ACM.
- Cunningham, H., D. Maynard, K. Bontcheva, and V. Tablan. 2002. GATE: A framework and graphical development environment for robust NLP tools and applications. *Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics*, 168–75. Philadelphia: ACL.
- Dublin Core Metadata Initiative. 2011. Dublin Core Metadata Initiative: Making it easier to find information, <http://dublincore.org/>.
- Hu, M., and B. Liu. 2004. Mining and summarizing customer reviews. *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 168–77. New York: ACM.
- Joachims, T. 1998. Text categorization with support vector machines: Learning with many relevant features. *Proceedings of 10th European Conference on Machine-learning*, Chemnitz, Germany, 21–24 April, 137–42. Berlin: Springer.
- Jones, K.S., and P. Willet, eds. 1997. *Readings in information retrieval*. San Francisco: Morgan Kaufmann.
- Na, J.-C., C. Khoo, S. Chan, and N.B. Hamzah. 2005. Sentiment-based search in digital libraries. *Proceedings of JCDL (Joint Conference on Digital Libraries) 2005*, Denver, Colorado, June, 143–44. New York: ACM.

- Nielsen, J., and R. Molich. 1990. Heuristic evaluation of user Interfaces. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'90)*, Seattle, WA, 249–56. New York: ACM.
- Pang, B., and L. Lee. 2004. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics ACL*, Barcelona, Spain, 271–78. Stroudsburg, PA: Association for Computational Linguistics.
- Pang, B., and L. Lee. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval* 2 (1–2): 1–135.
- Pang, B., L. Lee, and S. Vaityanathan. 2002. Thumbs up? Sentiment classification using machine-learning techniques. *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing*, 79–86. Stroudsburg, PA: Association for Computational Linguistics.
- Ruppenhofer, J., S. Somasundaran, and J. Wiebe. 2008. Finding the sources and targets of subjective expressions. *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC 2008)*, Marrakech, Morocco, 2781–88. Marrakech: European Language Resources Association.
- Sebastiani, F. 2002. Machine-learning in automated text categorization. *ACM Computing Surveys* 34 (1): 1–47.
- Staples, T., R. Wayland, and S. Payette. 2003. The Fedora Project: An open-source digital object repository management system, *D-Lib Magazine* 9 (4).
- Thet, T.T., J.-C. Na, and C. Khoo. 2008. Sentiment classification of movie reviews using multiple perspectives. *Proceedings of ICADL (International Conference on Asian Digital Libraries) 2008*, Bali, Indonesia, December, 184–93. Berlin: Springer-Verlag.
- Turney, P.D. 2002. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, Philadelphia, 417–34. Stroudsburg, PA: Association for Computational Linguistics.
- Yi, J., T. Nasukawa, R. Bunescu, and W. Niblack. 2003. Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques. *Proceedings of the Third IEEE International Conference on Data Mining (ICDM'03)*, 427–34. Washington DC: IEEE Computer Society.
- Zeng, H.-J., Q.-C. He, Z. Chen, W.-Y. Ma, and J. Ma. 2004. Learning to cluster web search results. *Proceedings of the 27th Annual International ACM SIGIR Conference*, Sheffield, UK, 210–17. New York, ACM.
- Zhuang, L., F. Jing, and X.-Y. Zhu. 2006. Movie review mining and summarization. *Proceeding of the 15th ACM Conference on Information and Knowledge Management*, 43–50. New York, ACM.