



# Expressing Opinion Diversity

Andreea Bizău – Babeş Bolyai University,  
Cluj-Napoca, Romania

**Delia Rusu**, Dunja Mladenić - Jožef Stefan  
Institute, Ljubljana, Slovenia



# Motivation: Opinion Diversity



**Guds** Some birds aren't meant to be caged, that's all. Their feathers are just too bright.....The **Shawshank Redemption**  
about 4 hours ago via web



**Guds** Fear can hold you prisoner, hope can set you free. A strong man can save himself, a great man can save another.—The **Shawshank Redemption**  
about 4 hours ago via web



**jameliahdz** The **Shawshank Redemption** .....sooo good!  
about 5 hours ago via ÜberSocial



1168 out of 1381 people found the following review useful:

**Tied for the best movie I have ever seen**, 26 November 2003



Author: [carflo](#) from Texas



**Nancy McGaw** one of the best..  
March 21 at 4:27am



**Aniruddh Tare** gr8est movie of all time..!! :-)  
March 21 at 4:28am · 1 person



**Sharon Stead Vassily** I know every line in this movie.  
March 21 at 5:00am



# Overview

- Terminology
- Related work
- Domain Driven Opinion Vocabulary
- Use Case:
  - Twitter Movie Comments
- Conclusions and Future Work



# Terminology

- ***opinion*** - a subjective expression of sentiments, appraisals or feelings
- ***opinion words*** - a set of keywords/phrases used in expressing an opinion
- ***orientation*** of an opinion word indicates whether the opinion expressed is positive, negative or neutral
- the totality of opinion words forms an ***opinion vocabulary***



# Related Work – Opinion Vocabulary

- Dictionary based
  - Esuli and Sebastiani, 2006 - SentiWordNet: three numerical scores (objective, positive, negative)
- Corpus based
  - Kanayama and Nasukawa, 2006 - context coherency (same polarity tend to appear successively)
  - Jialin Pan et al, 2010 - feature bipartite graph modeling the relationship between domain-specific words and domain-independent words.
- Dictionary and Corpus based
  - Jijkoun et al, 2010 - dependency parsing on a set of relevant documents, extracting patterns of the form (clue word, syntactic context, target of sentiment)

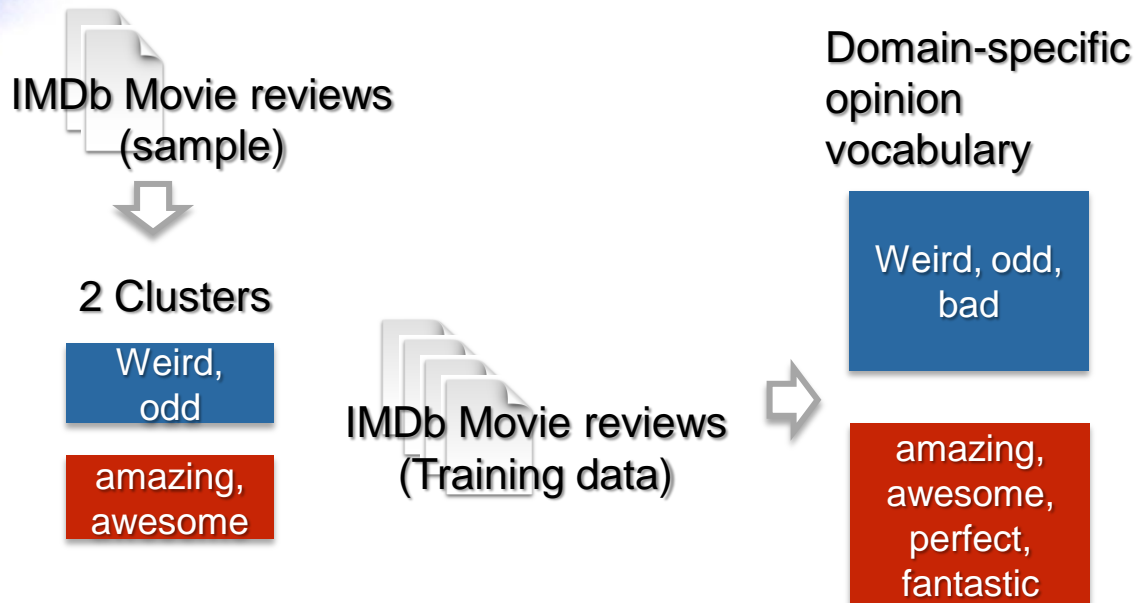


# Related Work – Opinion Analysis

- Approaches:
  - Hatzivassiloglou et al, 1997 – relevance of using connectives (conjunctions: and, or, but, etc)
  - Kim and Hovy, 2004 – use word seed lists and WordNet synsets to determine the strength of the opinion orientation for the identified opinion words
  - Gamon and Aue, 2005 – assign opinion orientation to candidate words, assuming that opinion terms with similar orientation tend to co-occur
- Twitter Applications:
  - Pang and Lee, 2002 – classifying movie reviews by overall document sentiment
  - Asur and Huberman, 2010 – sentiments extracted from Twitter can be used to build a prediction model for box-office revenue



# Our Approach



Vocabulary

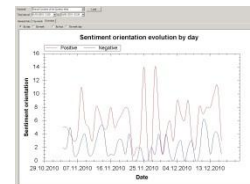


applied to



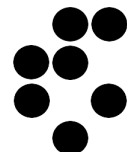
Movie tweets (Test data)

Twitter comments analysis



**RSeligstein** Finally watched **Inception** last night. **Loved it. GREAT** **movie** although I think I have lots of questions. Definitely need to see again!

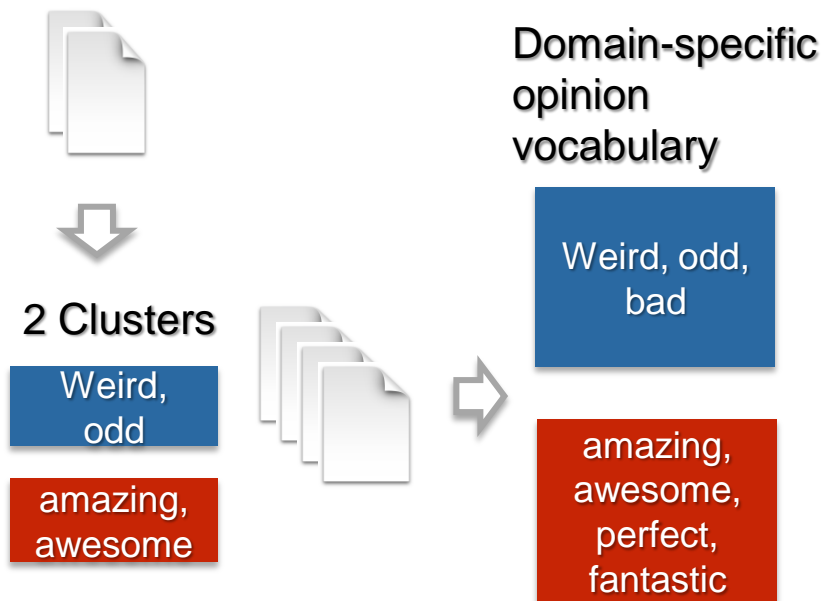
34 minutes ago via Twitter for iPhone





# Domain Driven Opinion Vocabulary

- Given a positive word seed list and a negative word seed list, expand the initial seed lists using synonymy / antonymy relations (WordNet)
  - the initial words will be assigned a score of 1 for positive words and -1 for negative words
- Given a corpus, parse and extract all adjectives and conjunctions – obtain a **graph** with two types of relationships between nodes:
  - same context orientation (words connected by and, or, nor) or
  - opposite context orientation (words connected by but, yet)
- Clean the resulting set of words and relationship graph by removing stop words and self-reference relations







# Domain Driven Opinion Vocabulary

Some of the characters are *fictitious*, **but not grotesque**

synonym

fictional

$$s_{\text{syn}} = -f * 0.43046$$

real

$$s_{\text{ant}} = f * 0.3089$$

fictitious

$$\text{score}(\text{"fictitious"}) = \max(s_{\text{syn}}, s_{\text{ant}})$$

$$\text{score}(\text{"fictitious"}) = s_{\text{syn}}$$

$$\text{score}(\text{"fictitious"}) = -0.3874 \text{ if } f = 0.9$$

antonym

fictitious **but** grotesque – **ContextOpposite** relationship

fictitious **but not** grotesque – **ContextSame** relationship



# Domain Driven Opinion Vocabulary

4. Propagate scores through the graph, by determining for each word  $w$ 
  - Positivity score **sPos**
  - Negativity score **sNeg**

---

**if**  $rel_{ij}$  **is a ContextSame relation**

$sPos(w_i) += weight(rel_{ij}) * prevSPos(w_j)$

$sNeg(w_i) += weight(rel_{ij}) * prevSNeg(w_j)$

**else if**  $rel_{ij}$  **is a ContextOpposite relation**

$sPos(w_i) += weight(rel_{ij}) * prevSNeg(w_j)$

$sNeg(w_i) += weight(rel_{ij}) * prevSPos(w_j)$

---



# Use Case: Twitter Movie Reviews

- domain specific document corpus of 27,886 IMDb movie reviews
- domain specific opinion vocabulary:
  - 9,318 words: 4,925 have a negative orientation and 4393 have a positive orientation

Inception (2010)	Meet the Spartans (2008)
<b>Positive words:</b> good, great, awesome, amazing, favorite, fantastic, incredible, thrilling, different, speechless	<b>Positive words:</b> funny, awesome, great
<b>Negative words:</b> bad, confusing, weird, stupid, dumb, boring, predictable, horrible, disappointing	<b>Negative words:</b> bad, stupid, dumb, weird, silly, common, ridiculous, terrible



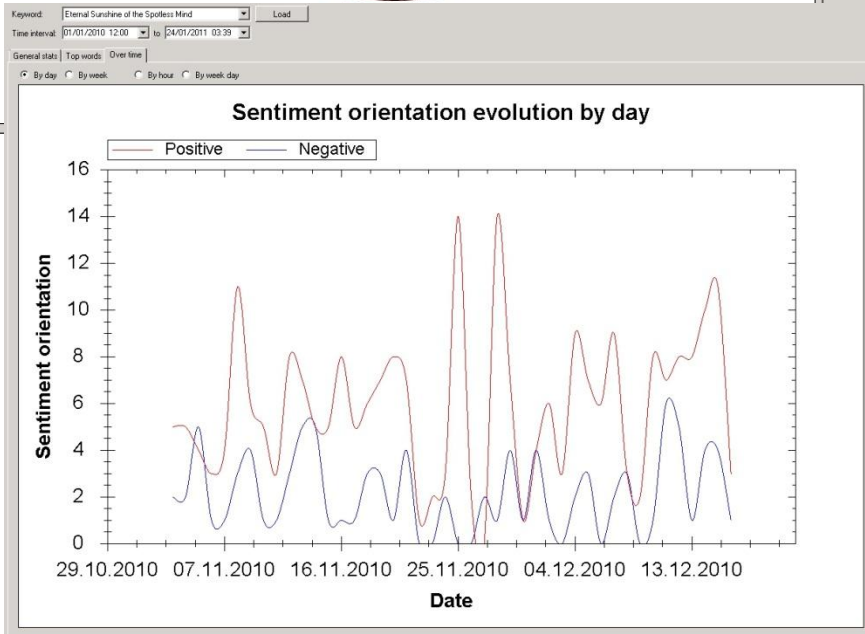
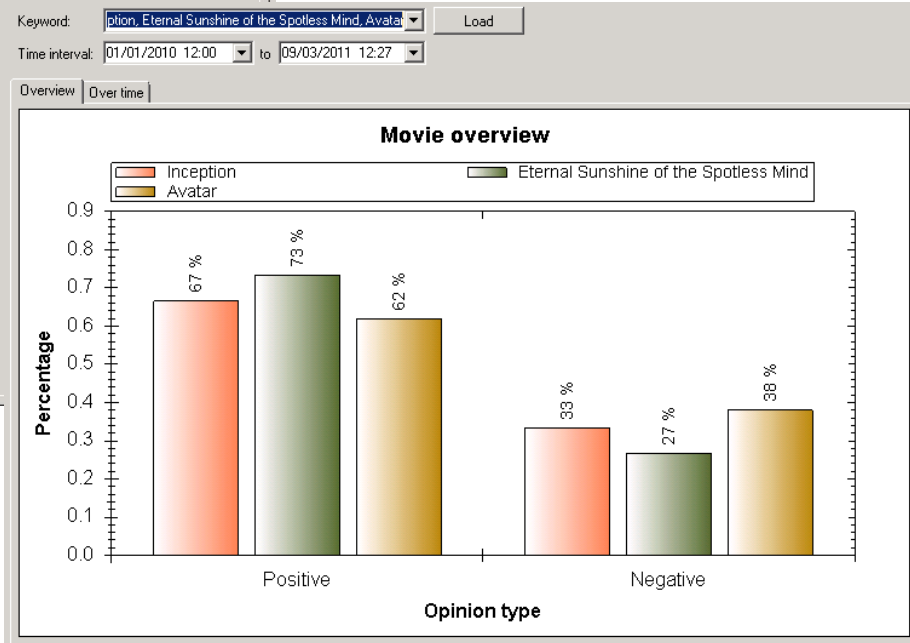
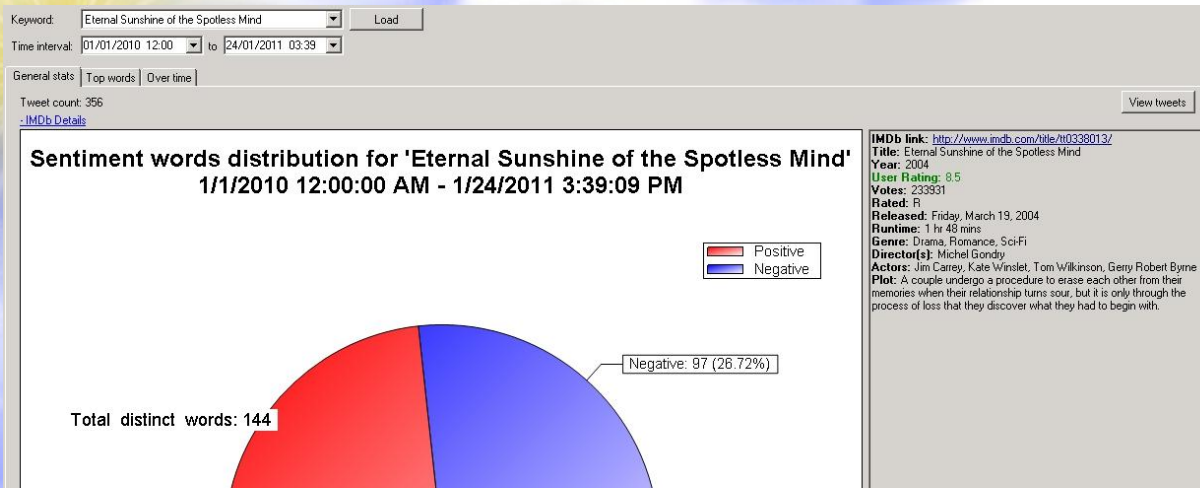
# Use Case: Twitter Movie Reviews

- 220,387 tweets crawled over a two month interval, keyed on 84 movies

Movie	Genre	Tweets	IMDb score	Our score
<b>Inception (2010)</b>	mystery, sci-fi, thriller	19,256	8.9	66.52
<b>Megamind (2010)</b>	animation, comedy, family	8,109	7.3	67.71
<b>Unstoppable (2010)</b>	drama, thriller	15,349	7	63.67
<b>Burlesque (2010)</b>	drama, music, romance	1,244	6.2	70.78
<b>Meet the Spartans (2008)</b>	comedy, war	44	2.5	40.67

# Twitter comments analysis

- Sentiment words distribution for a movie
- Sentiment orientation evolution per week, day, hour
- Movie comparison



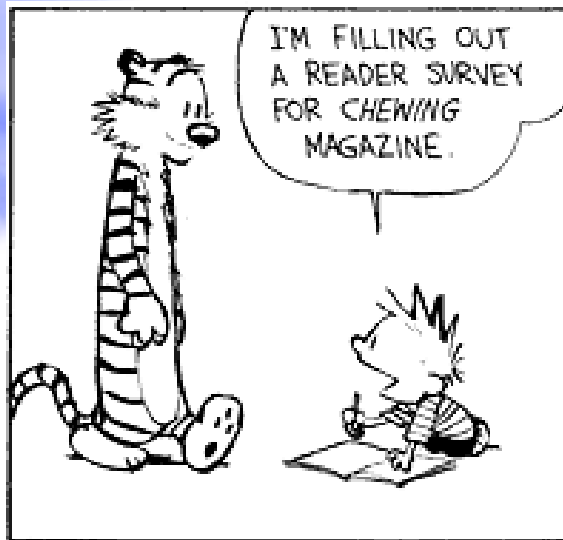


# Conclusions and Future Work

- identifying opinion diversity expressed within text, with the aid of a domain-specific vocabulary
- processing a corpus of IMDb movie reviews, generated an opinion lexicon and analyzed a different opinion source corpus, i.e. a tweet collection
- further extend our algorithm to include opinion words expressed by verbs and adverbs, as well as more complex expressions
- conduct experiments in order to
  - determine the correlation between positive opinion words for a given movie and the IMDb movie rating
  - evaluate the opinion lexicon directly



# Thank You for Your Attention!



SEE, THEY ASKED HOW MUCH MONEY I SPEND ON GUM EACH WEEK, SO I WROTE, "\$500." FOR MY AGE, I PUT "43," AND WHEN THEY ASKED WHAT MY FAVORITE FLAVOR IS, I WROTE "GARLIC / CURRY."

