

---

# WI - MINI PROJECT 2

---

GROUP: MIKKEL LARSEN & BRUNO THALMANN

Web Intelligence

Autumn Semester 2014



# Indhold

1	Sentiment Analysis	1
---	--------------------	---

# 1 Sentiment Analysis

## 1.0.1 Tokenization and Feature Extraction

The sentiment tokenization is very important because it catches a lot of the significant sentiment clues. Especially catching emoticons can mean a lot, because a lot of people use that for expressing how they feel about something. Capitalization and length of a word is also important to catch, fx 'this book is reeeeeeealy BAD'.

In our implementation we did not take any of the topics mentioned above into account. We wanted to focus on getting a running program before improving its results. So our tokenization basically just splits the words, but the implementation is ready for more in depth tokenization.

## 1.0.2 Feature Extraction

Feature extraction is about which words to use and how to handle negation. We have chosen to use all words and our handling of negation can be seen in the next section.

### Handling of Negation

Is used for identifying when a part of a sentence is negative. Fx 'I didn't like this product vs I really like this product'.