

Exam assignment - Language Processing 2

*IT & Cognition, semester F17,
Teachers Jürgen Wedekind and Peter Juel Henriksen*

Write a paper on **Sentiment analysis (SA)**

Your paper must include a practical investigation of the movie review corpus studied in class in April 2017 and available at <http://lab.homunculus.dk/cgi-bin/LangProc2/news.cgi> [link dated May 19 2017].

In your investigation you must utilize at least two machine learning-based approaches to sentiment analysis, using the movie review corpus for training and testing. You may use any elements of the NLTK sentiment package (e.g. <http://www.nltk.org/api/nltk.sentiment.html>). You may, but need not, utilize the “AFINN” lexicon as well (available same place). You may also include materials developed by yourself in connection with the LangProc2 course projects on sentiment analysis (April 2017, presented in class Apr 24). In case you do refer to data, code, or text from your course project, you must clearly mark any recycled material as such.

NB! Please meet *all* requirements below. **Any deviance may affect your grade.**

Formal requirements

- Follow the formal exam requirements stated in the curriculum (see link below)
- http://hum.ku.dk/uddannelser/aktuelle_studieordninger/it_cognition/it_cognition_msc.pdf

Content requirements

- Your paper must be in English (check carefully for spelling errors and broken syntax).
- The paper should have (i) a descriptive title and (ii) a by-line with your name(s) and student ID(s).
- The paper must include the following main sections:
 - **Background** [explaining the purposes and methods of SA in broader terms]
 - **Data generation** [using relevant code (Python 3), written by yourself and/or copied from NLTK (e.g. <http://www.nltk.org/api/nltk.sentiment.html>); this section must report on actual test-runs utilizing the movie review corpus and applying at least two machine learning algorithms (Naïve-Bayes being one example)]
 - **Quantitative evaluation** [demonstrating how SA results can be evaluated automatically]
 - **Qualitative evaluation** [evaluating and discussing the quantitative results wrt. validity, reliability and/or relevance in a “real-world” perspective]
 - **Conclusion** [presenting in a condensed form your results and observations, e.g. pointing to strengths, weaknesses, and future directions]
 - References
- You may, but need not, include more sections. One example: **Cognitive Aspects** discussing the relevance of SA in terms of psychological realism, e.g. by relating “sentiments in the brain” to ‘*sentiment*’ in the SA tradition.