In this exercise, you will perform aspect-based sentiment analysis using methods of information extraction, following the idea of Hu & Liu 2004 (see lecture slides 28 - 31). In more detail, you will first extract frequent nouns, which are supposed to be important features for the product that is rated in the review. Then, you will search for the closest adjective to this noun, which represents the opinion phrase expressing a sentiment on this product feature. Finally you will look up the polarity of this adjective in a sentiment lexicon.

**Code**
You can find the completed code as well as a result file in the uploaded Eclipse Project (Exercise4_AspectBasedOpinionMining_Project_Solution.zip). Import it just like the original project.

**Experiments and Discussion**
Have a look at the results opening data/output.csv with Excel.

- What kind of errors does the extracted data contain?
  - Errors caused by the POS tagger: extracted terms are no nouns / opinion phrases are no adjectives
  - The extracted opinion phrases do not really refer to the noun
  - The polarity derived by the lexicon is wrong or no polarity was found

- Play around with the frequency threshold for nouns. Does it improve the results to change it? Adjust the search window size for the closest adjective. What is a good size?
If the frequency threshold is too low, nouns that are no features will be extracted. If it is too high, it will miss a lot of feature. Depending on whether one prefers a high precision of the extracted results or rather a high recall, an appropriate threshold has to be chosen for each particular problem.

If the search window is too small, an appropriate adjective might be missed. If it is too large, the algorithm will extract adjectives for a noun that do not refer to it. It could be considered as well using different window sizes for the right and the left sight by investigating typical grammar structures.

- How could the results be evaluated properly?
  - First of all, there are two tasks that have to be evaluated. The first one is whether the feature extraction worked, the second is whether the extracted opinion phrase is correct.
  - Evaluation measures that are appropriate to tell something about the quality of the results are precision and recall. Precision is the amount of correctly extracted instances out of all extractions (= the quality), and recall refers to the amount of extracted information out of all information that was in the corpus and should have been extracted (= the completeness of the results).
  - Precision and recall can be measured by comparing the extractions against a gold standard. The gold standard refers to result set that contains the “truth”, i.e. correct results. It is typically created manually. By comparing the extracted results to the gold standard, true and false extractions can be counted and thereof precision and recall can be calculated.