###  CMPE 480 INTRODUCTION TO ARTIFICIAL INTELLIGENCE 14.11.2016

### APPLICATION PROJECT

In this project, you will write a Prolog program for the problem of “guessing the rating of a movie review”.

We have a corpus (data file) that includes movie reviews and ratings corresponding to the reviews. Each entry in the corpus has the form

*w1 w2 ... wn rating*

where *w*’s are the words in the review and *rating* (between 1 and 10) is the score given by the user to the movie. For instance, an example entry may be as follows:

*Bu filmi çok beğendim ama süresi uzundu. Herkese tavsiye ederim. 8*

You will use this corpus to learn a score for each word. The score for a word *w* will be calculated as follows:

$$\frac{\sum\_{i=1}^{10}\left(number of times w occurs in reviews with rating i\right)\*i}{total number of times w occurs in reviews}$$

This score indicates how much positive or negative effect a word has. For instance, the word “terrible” will mostly occur in reviews with low ratings, thus its average score will tend to be close to 1.

We also have a dictionary file that contains words and their grammatical categories. Some example entries may be: (film noun), (recently adverb), (the determiner), (wonderful adjective).

As a third file, we have a category file that shows the importance of each category. The weight (importance) is between 0 and 1. For instance: (adjective 0.8) (noun 0.5).

The program will use these three files to learn the scores of words, the categories of words, and importance of categories.

The program will be given a goal to calculate the rating of a movie review. The review will be given, but no rating will be given. The rating of the review will be calculated as follows: For each word in the review, its score calculated from the corpus as stated above will be multiplied with the weight of the category the word belongs to. These scores will be added and divided by the number of words in the review. The result is the final rating of the review.

Design and develop your MovieRater system. Test the system comprehensively and show the execution with different sets of data. The corpus should include at least 50 reviews. Use different dictionaries, give different weights to categories, etc. in the tests.

**Notes:**

* Your project will be graded on correctness, readability, testing strategy, efficiency, documentation, and the use of logic programming style.
* You must use *pure* logic programming style. That is, you may not use the operators and predicates like ; (“or”), -> (“if-then”), and “repeat”. An exception is that you may use the ! (cut) predicate.
* Prolog program (SWI Prolog) with some manuals and tutorials are on the course website.
* Read the *Programming Projects* section in the document General Information for Students link in the *Syllabus*. You should obey the rules for electronic submission (e-mail subject, etc.) and documentation. You must prepare a document about the project, which is an important part of the project. Explain clearly the program (predicates, parameters of predicates, etc.) and include several test cases in the document.
* The project will be done individually, not as a group.
* **If your code includes any predicates not written by you, you must explicitly emphasize them inside the source code.**
* The strict deadline is 20.12.2016. There will be a demo; the exact date and time will be arranged later.