Grundlagen der künstlichen Intelligenz
Exam Bonus Programming Exercise 1   26th October 2015

Problem 1.1: Syzygies

This is an simplification of a word puzzle created by the English polymath Lewis Carroll where the player is given a start word (e.g. WHITE) and a goal word (e.g. GREEN). One must construct a chain of words, where, for adjacent words in the chain, either the first two letters of one word are the last two letters of the next word, or the last two letters of one word are the first two letters of the next word.[1] An example:

$$
\begin{array}{cccc}
 & \text{WHI} & \text{TE} & \\
 & & \text{TE} & \text{AMWORK} \\
\text{EN} & \text{UMERA} & \text{TE} & \\
\text{GRE} & \text{EN} & & \\
\end{array}
$$

A list of words[2] is available as a `.txt` and a `.mat` file on Moodle. Using these words, implement, in a language of your choice, a search algorithm of your choice to connect the following words, using the **shortest chain possible** (which may not be unique):

- PEN → PAPER
- ACORN → OAK
- AUTUMN → WINTER
- EARTHQUAKE → TSUNAMI
- STARSHIP → ENTERPRISE

This problem is to be solved individually. Please email a zipped folder to `aaron.pereira@tum.de` under the subject header "Syzygies" with:

1. the answers to the above questions,
2. your source code (so that I can check for uniqueness),
3. an executable file/MATLAB script/shell script to run on command line on MacOSX or Windows, with the start and end words as inputs and the shortest chain of words connecting them as output,
4. a `readme.txt` file telling me how to run the executable/script, which directory to place the word list (if necessary), etc. It may be in English or German, but please make it completely "idiot-proof"!

A pass will be awarded only if:

- I receive correct answers to the questions above,
- your executable also delivers the shortest solution for two word pairs of my choice,
- your executable finds the solution to any of the word pairs mentioned above, in less than 1 hour on my computer (Macbook Pro, 2.8GHz i7 processor 16GB RAM) or equivalent Windows PC.
- the code of the search algorithm is your own work (although you can use open-source code such as `github.com/tinevez/matlab-tree` to support your work, as long as you reference it). I will **not** grade your source code.

Submission will close on **Sunday 29th November at 23:59**. If your executable does not work for technical reasons, I will contact you – please respond within 48 hours, or I cannot consider your solution. **A hint:** the list contains over 100,000 words! To save time when finding the children of nodes, you can pre-compute lists of words that begin/end with pairs of letters, then look them up.

---

[1] This means that CEDAR can follow FACE, and CHASE can follow SEMANTICS. However, FALCON cannot follow FACE and SERIOUS cannot follow SEMANTICS.

[2] adapted from `www-01.sil.org/linguistics/wordlists/english/`, accessed 14.10.2015