

# Paraphrasing a text using replacement by synonyms

April 1, 2009

## 1 Project description

You are going to implement a program that does a very simple form of paraphrasing a text, by replacing nouns in a given text by synonyms.

**Example:** “adobe” is similar to “cob” (română, traducere aproximativă “paiantă”).

**Main phrase:** Most indian houses in the American Southwest were made of adobe. Adobe is a natural building material made from sand, clay, and water, with some kind of fibrous or organic material (sticks, straw, dung), which is shaped into bricks using frames and dried in the sun. Adobe structures are extremely durable and account for some of the oldest extant buildings on the planet.

**Paraphrase:** Replace “adobe” by “cob” everywhere in the text above. Of course, you should do something like this for other nouns as well.

Your program should work as follows:

- The main text is read from a file. You may assume, for convenience, that the text does not contain punctuation signs, and each phrase starts after a new line.
- A second file is the “dictionary”. It contains a list of words, together with a list of synonyms. You may assume that each line contains a word, followed by a list of all its synonyms.
- You are going to read this dictionary and create a linked list of nodes.
- Each node contains two things: an item (the main word) and a linked list of synonyms.
- When you read a word from the main file implement the following process:
- Search for the word in the linked list representing the dictionary. If it is not found then the word is not a known, and is output in a modified form.
- If the word is found then select one of its synonyms from the list.

- **The same synonym is going to be used consistently through the text.** To do so, you will maintain a list of pairs (noun, synonym). When searching for the replacement word you first search this list, only otherwise you search in the dictionary.
- You should write the paraphrased text in an output file.

## 2 What to turn in

You will turn in a report containing:

- A description of the significant choices/issues in the design of your code and experiments.
- The results of your experiments, followed by a few lines describing your conclusions.
- The listing (source code) of your programs.

You may turn in the document in class, or via email ([gabriel.istrate@gmail.com](mailto:gabriel.istrate@gmail.com)).

## 3 Coding standards

A percentage of your grade will be based on the quality of your code, so pay attention to it. Discuss changes (if any) you made to programs presented in class. Take extra care in documenting the code you are implementing on your own. Properly modularize the code (for instance implement separate functions for significant parts of the program).

## 4 Deadline

Two weeks from now (April 15, 2008, 4PM local time). This is a strict deadline. No credit will be given for homework turned in late.