

# Apples And Apple's

## Problem Statements:

The word "Apple" could generally refer to one of these two:

(a) Apple Inc., the great Computer giant.

(b) Apple, the fruit

You are provided two text files, with a number of lines. Each line contains either a sentence or a paragraph or a text snippet which could either be related to Apple, the computer company, or the apple, the fruit. Your task is to perform disambiguation between these two groups and identify which one is being referred to. It is possible that the plural or the possessive form of Apple might exist in some of the tests (apples, Apple's).

## Training Data:

You are provided with two text files, which contain near-complete text from the Wikipedia for Apple Inc. as well as apple the fruit. You have to use these two text file for the training purpose. For offline inspection and access, you could access these two files here:

Text from Wikipedia entry on Apple-Computers: "**apple-computers.txt**"

Text from Wikipedia entry on Apple the fruit: "**apple-fruit.txt**"

## Input Format

An Integer N, no more than 100.

line\_1

line\_2

line\_3

line\_4

...

...

line\_N

## Output Format

computer-company

fruit

computer-company

fruit

..

..

..

N lines of output

## Sample Input

10

Apple already plans to buy back \$100 billion in shares, including \$16 billion worth last quarter. Icahn probably pounded the dinner table he and Cook shared recently for their much-reported bread-breaking at Icahn's New York apartment. Apple's cash stash currently sits at a whopping \$145 billion but only \$43 billion is in the U.S., which is why Icahn wants to float bonds to cover a buyback.

Fortunately, there are “low-chill” apple varieties for temperate climates. (Chilling hours are defined as nonconsecutive hours of winter temperatures below 45 degrees.) As a general guide, if you live on or near the coast, your garden gets only 100 to 200 chilling hours. Inland San Diego gardens get about 400 to 500 chilling hours — still considered “low chill.”

If this seems a bit like déjà vu, you’ll recall that Apple just held an event to unveil two new iPhone models – the 5c and 5s – back on September 10.

.

.

.

.

.

## Sample Output

computer-company

fruit

computer-company

## Explanation

In the first chunk of text, Apple Inc. is being referred to.

In the second chunk, Apple the fruit is being referred to.

In the last chunk of text, again, it is Apple Inc. which is being referred to.

## Deliverable

1. Submit your complete code and the saved model. We will execute your code considering both the text file in the same directory with our custom test dataset. You are allowed to implement the solution in python only. You can use any built-in library for this purpose. We encourage you to use scikit-learn and other functions from nltk. Along with code you have to submit a written document about the method/algorithms used in the solution and your justification why that method will work for this problem.
2. Along with the code, saved model and documentation you have to build a simple GUI (using any web technology) where we can play with the model. In the GUI there will be a simple search/text box. Given a sample text in the search/text box, upon submitting it will show the desired results ( computer-company or fruit) in the GUI.
3. To interact with the saved model from GUI you need to build a REST API in python. The API will take a sample text as a parameter and will return the desired output as a response. You need to handle all possible edge cases for this API.
4. Need to host a database to store all the input text in the GUI and the response returned by the model.
5. Need to submit the GUI code, API code and a simple document about the deployment and architecture of GUI and API.

Your submission will be assessed based on the performance of the model, GUI and DB architecture and design of the API.