

Heroes and Heroines, Villains and Villainesses

Notification:

~~We noticed that there's suddenly been an issue in reading the corpus from the code, and we're working on fixing it. In the mean time, please download the corpus from the link provided and develop your solution locally. We will update participants when the issue is fixed.~~

You will be provided with a large corpus of text which contains one or more stories or snippets, which could either be present in full, summarized or in partial form. This text will contain the thoughts, actions, dialogues and interactions between various characters. It will be segmented into paragraphs which may or may not have a continuity of thought and logic between them. **Your task is to create an intelligent program that guesses the gender of certain characters in this text, whose first names will be specified to you.**

There might be multiple characters who share the same first-name. If a test contains such a name, all such characters will share the same gender.

Text Corpus

This file [corpus.txt](#) needs to be read by your program.

Input Format

The first line is an integer N. This is the number of characters from this text, for whom you need to detect the gender.

This is followed by the first names of N people, from the corpus text, each on a new line.

Constraints and Nature of the Corpus

As mentioned previously, it will be segmented into paragraphs which may or may not have a continuity of thought and logic between them. You do not need to code defensively for obscure or ambiguous cases. Tests will be set in such a way that the genders of the characters will be reasonably direct and simple, for a human reader of the text to infer, based on the few sentences before or after the position where that name occurs in the text. Text has been taken from sources such as project Gutenberg, Wikipedia and publicly available news snippets, biographies and stories. To replicate the difficulty of dealing with much of the text available on the web, parts of it might be somewhat unstructured, and there might even be a small percentage of text which is not in English.

Number of Lines in Corpus = 65443

Number of Tests(N) will be such that $1 \leq N \leq 100$.

Tests will contain first names, where only the first letter is capitalized. Names will only be from the sections of the corpus which are in English.

Output Format

N lines. Each line contains just one word: Male or Female. The i-th line contains the gender detected for the i-th first-name in the tests.

Corpus

The corpus will be provided to you [here](#). Your program can read in the provided corpus by assuming that the file "corpus.txt" lies in the current folder of your program.

Example Input

3
John
Sherlock
Mary

Example Output

Male
Male
Female

Explanation and a few hints

On reading the text, if we need to infer the gender of John, we find several portions of the text which provide us with sufficient hints to figure out the gender with reasonable accuracy.

*"Mr. **John** Turner," ried the hotel waiter, opening the door of our sitting-room, and ushering in a visitor. "My nme," said **he**, "is **John** Openshaw, but my own affairs have, as far as I can understand, little to do with this awful business. It is a hereditary matter; so in order to give you an idea of the facts, I must go back to the commencement of the affair. "There is one thing," sid John Openshaw. **He** rummaged in **his** coat pocket,*

Scoring

The Max score for this problem is 50.

Score = MaxScore for a test case * Max(C-W,0)/N

C = Number of tests for which you guess the gender correctly

N = total number of tests (first-names) in the input file.

W = Number of tests for which you guess the gender incorrectly

ML Libraries are not enabled for this challenge because many of them have the potential to provide a near-direct solution for this challenge.