Stitch the Torn Wiki

Text blocks which are approximately 500 to 1000 words in length are picked up from N different Wikipedia articles. Every block of text has been picked up from a unique Wikipedia article, about a well known person or place.

Each of these text blocks is split into two parts of roughly equal length.

The first (starting) part obtained after splitting is placed in **Set A** which will hold all the starting blocks. The second part of the block, is placed in **Set B** which will contain the second part for all the text fragments which we selected. Both the Sets A and B are shuffled up, and the ordering of elements is lost.

Your task is to identify, for each text fragment (*a*) in **Set A**, which is the correct, corresponding text fragment (*b*) in **Set B**, such that both *a* and *b* were in the same text block initially.

Getting started - Think about using the vector space model

Here's a wonderful Youtube video of Professor Christopher Manning from Stanford, explaining the vector space model and cosine similarity , which you could consider using as a starting point.

Input Format

An Integer N on the first line. This is followed by 2N+1 lines.
Text fragments (numbered 1 to N) from Set A, each on a new line (so a total of N lines).
A separator with five asterisk marks "*" which indicates the end of Set A and beginning of Set B.
Text fragments (numbered 1 to N) from Set B, each on a new line (so a total of N lines).

Output Format

N lines, each containing one integer.
The i-th line should contain an integer j such that the i-th element of **Set A** and the j-th element of **Set B** are a pair, i.e., both originally came from the same block of text/Wikipedia article.

Constraints

1 <= N <= 100
No text fragment will have more than 10000 characters.

Sample Input

(*Please note that the real inputs used will be much longer, and generated with text blocks with 500-1000 words. This is for explanatory purposes only*)

> 3
> Delhi (also known as the National Capital Territory of India) is a metropolitan region in India that includes the national capital city, New Delhi. With a population of 22 million in 2011, it is the world's second most populous city and the largest city in India in terms of area. The NCT and its urban region have been given the special status of National Capital Region (NCR) under the Constitution of India's 69th amendment act of 1991. The NCR includes the neighbouring cities of Baghpat, Gurgaon, Sonepat, Faridabad, Ghaziabad, Noida, Greater Noida and other nearby towns, and has nearly 22.2 million residents.
> Seattle is a coastal seaport city and the seat of King County, in the U.S. state of Washington. With an estimated 634,535 residents

as of 2012, Seattle is the largest city in the Pacific Northwest region of North America and one of the fastest-growing cities in the United States. The Seattle metropolitan area of around 4 million inhabitants is the 15th largest metropolitan area in the nation.[6] The city is situated on a narrow isthmus between Puget Sound (an inlet of the Pacific Ocean) and Lake Washington, about 100 miles (160 km) south of the Canada–United States border. A major gateway for trade with Asia, Seattle is the 8th largest port in the United States and 9th largest in North America in terms of container handling.

Martin Luther OSA (10 November 1483 – 18 February 1546) was a German monk, Catholic priest, professor of theology and seminal figure of a reform movement in 16th century Christianity, subsequently known as the Protestant Reformation.[1] He strongly disputed the claim that freedom from God's punishment for sin could be purchased with money. He confronted indulgence salesman Johann Tetzel, a Dominican friar, with his Ninety-Five Theses in 1517. His refusal to retract all of his writings at the demand of Pope Leo X in 1520 and the Holy Roman Emperor Charles V at the Diet of Worms in 1521 resulted in his excommunication by the Pope and condemnation as an outlaw by the Emperor.

*****

The Seattle area had been inhabited by Native Americans for at least 4,000 years before the first permanent European settlers. Arthur A. Denny and his group of travelers, subsequently known as the Denny Party, arrived at Alki Point on November 13, 1851. The settlement was moved to its current site and named "Seattle" in 1853, after Chief Si'ahl of the local Duwamish and Suquamish tribes.

Although technically a federally administered union territory, the political administration of the NCT of Delhi today more closely resembles that of a state of India, with its own legislature, high court and an executive council of ministers headed by a Chief Minister. New Delhi is jointly administered by the federal government of India and the local government of Delhi, and is the capital of the NCT of Delhi.

Luther taught that salvation and subsequently eternity in heaven is not earned by good deeds but is received only as a free gift of God's grace through faith in Jesus Christ as redeemer from sin and subsequently eternity in hell. His theology challenged the authority of the Pope of the Roman Catholic Church by teaching that the Bible is the only source of divinely revealed knowledge from God and opposed sacerdotalism by considering all baptized Christians to be a holy priesthood. Those who identify with these, and all of Luther's wider teachings, are called Lutherans.

Sample Output

```
2
1
3
```

Explanation

The first, second and third text fragment of **Set A** are about Delhi, Seattle and Martin Luther respectively.
In set B, the paragraph on Delhi, is the second text fragment.
The paragraph on Seattle is the first text fragment in **Set B**.
The paragraph on Martin Luther Kind is the third text fragment in **Set B**.
So, the expected output is 2, 1, 3 respectively.

Scoring

A sample test case with twenty paragraphs is provided to you when you Compile and Test.
Extensive training data is not required for this challenge. The weightage for a test case will be proportional to the number of tests (Articles) which it contains. This works out to a ratio of 1:2 (Sample Test: Hidden Test).
Score = M * (C)/N Where M is the Maximum Score for the test case.
C = Number of correct answers in your output.
N = Total number of Wikipedia Articles (which were split into 2N fragments and divided into Set A and Set B respectively).

Note:Submissions will be disqualified if it is evident that the code has been written in such a way that the sample test case answers are hard-coded, or similar approaches, where the answer is not computed, but arrived at by trying to ensure the code matches the sample answers.

Timelimits

Timelimits can be seen here. ML libraries are enabled for this contest.