4. 文章抄襲比對 (佔分 10 分)

請寫出一個程式來比對一篇文章 A 是否抄襲另一篇文章 B，並計算出相似度百分比。A 和 B 兩篇文章皆為英文，為簡化起見，兩篇文章的英文單字數量都小於 1000 個單字，且相似度百分比的計算採用下列規則：

(1) 英文單字可由空格、逗號、和句號分開，標點符號和大小寫都不影響相同單字的判斷，例如 the amount of time he jogs. 和 He jogs every morning. 視為重複了 he jogs，不會因為 H 大寫和.句號而視為不同單字。

(2) 單字的不同時態或單複數則視為不同之單字，例如 jog 和 jogs 會視為不重複之單字。

(3) 若文章 A 有連續 7 個英文單字(含)以上與文章 B 重複，即視為相似句子，累計其重複單字的數量，6 個以下的重複單字則不計入數量。

(4) 若文章 A 的連續英文單字與文章 B 中有多處重複，則僅取重複最多單字的一次數量計入，並不會多次計入。

(5) 相似度百分比的計算公式為 (文章 A 中累計重複單字的總數量) / (文章 A 總單字數量)。

以範例 1 的輸入為例，文章 A 中的 high level languages, low level languages are closer to the hardware 與文章 B 中的 high level languages. Low level languages are closer to the hardware 重複了 11 個單字。Unlike high level languages, low level languages 則重複了 7 個單字，但因為其中的 high level languages, low level languages 6 個單字已經計入過重複單字的數量(在 11 個單字中)，只有 Unlike 1 個單字需要再累計數量，所以重複單字的總數量為 12 個單字。文章 A 的總單字數量為 21，因此相似度百分比為 12/21=57.14%。

輸入說明：

第 1 列為文章 A，第 2 列為文章 B（假設文章內容均不分段換行，一篇文章自成一列），比對 2 篇文章的相似度百分比。

輸出說明：

輸出文章 A 與文章 B 相似度百分比(以百分比方式顯示，四捨五入到小數點以下兩位)。

範例輸入：（第 1 列為文章 A 內容，第 2 列為文章 B 內容，以下為配合版面限制換列呈現）

Machine languages and assembly languages are low level languages. Unlike high level languages, low level languages are closer to the hardware.
C languages, C++ languages, and Java Languages are high level languages. Low level languages are closer to the hardware than are high level languages. High-level languages are designed to simplify computer programming. Unlike high level languages, low level languages can be converted to machine code without using a compiler or interpreter.

範例輸出：

57.14%