a 2 ONLINE

■ HackMD

Prime Full-text search on all paid plans

CHANGED AN HOUR AGO

學號:410821204

姓名:杜昉紜 系級:資工三

• 作業描述

• 程式碼說明

課程:文字探勘_HW2 - 林金龍

■ 程式碼如下:

■ 執行結果如下:

• 1. Download these four books from Project Gutenbergr

■ 1. 接下來進行資料的先前處理,我們將它們分成章節

■ 2. 使用tidytext unnest_tokens將它們分成單詞

Count number of words in each document.

■ 首先我們利用程式碼到 https://www.gutenberg.org/ 去下載我們要的四本書

2. Use tidytext's unnest tokens to divide these four books into chapters and treat them as

■ 3. 刪除停用詞stop_words,並計算每個章節單字量,將每一章視為一個單獨的

□ Like □ Bookmark □ 0 comments