

以爬蟲技術進行網頁關鍵詞之分析

Use Web Crawler Technology in Keyword Analysis of Web Pages

陳建伯、林聖倫

Jian-Bo Chen

銘傳大學電腦與通訊工程學系

Department of Information and Telecommunications Engineering,
Ming Chuan University

摘要

本文主要是以 Python 程式來進行網路爬蟲，將針對特定的主題將內容擷取下來，再經由斷詞套件將整篇文章斷詞後，找出該篇文章相關的關鍵詞並統計其出現的次數。最後，將我們所需的關鍵詞以文字雲做視覺的呈現。呈現出來的結果能夠很清楚的知道文章所敘述的重點，做一個快速分析與重點整理。

關鍵字：爬蟲、斷詞、文字雲

Abstract

In this paper, we use web crawler technology to get the contents of a specific topic. These contents are tokenized to achieve the keywords and their statistic information. Finally, the word-cloud are used in these keywords to demonstrate their importance.

Keywords: web crawler, tokenize, word-cloud.