Chapter5_TianFengshou_EN.docx

by Fengshou TIAN

Submission date: 16-Jan-2025 10:18PM (UTC-0800)
Submission ID: 2565795456
File name: Chapter5_TianFengshou_EN.docx (20.3K)
Word count: 345
Character count: 1920

CHAPTER 5

CONCLUSION AND RECOMMENDATIONS

5.1 Summary

This study completed three core tasks: (1) Building a crawler that can crawl Weibo, Xiaohongshu, Zhihu, and Bilibili. (2) Building a data cleaning and conversion system that can merge data from multiple social media. (3) Building a word segmentation and sentiment analysis model to perform word segmentation and sentiment analysis on the acquired data. In this process, an emoji discovery system was also built that can automatically discover emojis based on text format and vocabulary frequency.

In the end, the overall project basically achieved the design goal, and the project can provide effective support for subsequent research and analysis related to Chinese social media.

5.2 Future Works

There are three directions for improvement and enhancement in future work: (1) Optimizing data acquisition capabilities. The current concurrent performance of

crawlers is average. Compared with the hundreds of millions of active users of Chinese social media, the data acquisition capability is too weak and cannot fully acquire real-time hot data. Therefore, the concurrent performance of the current crawlers needs to be further improved in the future. This requires starting from the account pool and proxy pool to build a large-scale distributed crawler system that can acquire a large amount of social media data.

(2) (2) Improving the prediction ability of sentiment analysis models. Currently, mainstream sentiment analysis models are trained on large-scale data sets and then perform transfer learning. However, the data sets used for transfer learning are relatively small. Therefore, in the future, native data sets of Chinese social media should be built to fully explore the learning ability of the model. Secondly, we can start from specific means in the field of sentiment analysis such as sentiment dictionaries to further improve relevant algorithms.

(3) For hot events on social media, there is a propagation mechanism between multiple platforms, but this study did not involve research on related content. In the future, we can combine professional knowledge such as communication to study the mechanism of event propagation, so as to build a more complete event tracking and analysis mechanism.

Chapter5_TianFengshou_EN.docx

ORIGINALITY REPORT

5%	3%	2%	0%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	thecustomwritings.com Internet Source	3%
2	Zidong Jiang, Fabio Di Troia, Mark Stamp. "Chapter 22 Sentiment Analysis for Troll Detection on Weibo", Springer Science and Business Media LLC, 2021 Publication	2%

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		