An online search engine for classical Chinese poems

Xiwei Wu & Ning Yang & Hui Ma

Abstract: This is a group-based project for Web Search & Mining. In this project, we implemented a poetry crawler that crawled more than 10,000 poems from https://www.shi-ci.com/ from pre-Qin to modern times, then implemented a search engine that can support a variety of different search requirements, and finally provided a web front-end that can be easily used.

Keywords: Scrapy, , Chinese poems, search engine

1 Introduction

Outline. In Section 2, we will introduce our crawler based on Scrapy and SQL. In Section 3, we will introduce our design for search engine. In Section 4, we will introduce our design for web front-end. In Section 5, we will give a demonstration of the results of our project.

2 Data Crawling and Preprocessing

3 Search Engine

3.1 Overall Design

3.2 Boolean Search

3.3 Zone-specific Search

3.4 Pinyin-based Tolerant Search

3.5 Ranked Search

4 Web Interface

5 Conclusion