Jiacan Huang

August 6, 2018

CS6200

# Weibo Personal Network Search Engine

## Abstract

Weibo is the second popular social media site in China, which is a place for the young to share and find peers. The Weibo Personal Network Search Engine aims at improve users' search experience to identify a specific post in personal network, and help users strength and expand their personal network when using Weibo. This search engine provides post search and user search functions, based on Elasticsearch and ik Chinese analyzer. Ranking function is modified in response to the situation. The project also include a user interface with search toolbar. In Search Engine Result Page (SERP), summary of each retrieved document with key words highlighted are shown. In future work, improvements can be made by acquiring fresh dataset with more fields, improving phrases dictionary, providing advanced search and modifying ranking function according to user feedback.

## Introduction

Although western social media platform is banned, Chinese people still spend a lot of time on social media. Three most popular social media sites in China are WeChat, Weibo, Tencent QQ.

According to China Internet Watch, monthly active users of Weibo reached 392 million in 2017. Weibo is labeled as 'Twitter of China', but it means more than that for the young in China . QQ is abandoned by most young man because they thought it is a social media for primary school students. Since friends and elder family members are in social network of WeChat, posts' style of young people in WeChat are peaceful and loving. It is Weibo where Chinese young people can reveal their real thinking, find peers having same hobbies.

Both WeChat and QQ relies on strong and two-way connections. Unlike WeChat and Tencent QQ, Weibo is an open social media platform. Users can follow an individual and read their posts, then like, comment and share them without being followed back. Weibo offers users chances to expand their network.

There are some improvements can be made to improve user experience, strength and expand user's personal network on Weibo. Please image following situations when you using Weibo: You want to share a post you read 1 week ago with your friend. How can you find it quickly? You wonder How people in your network think about a trending news, is there a good way to gather their posts about it? You want to expand your network from your followers, how can you find the one share same interests with you? If one of the situation once bother you, Weibo Personal Network Search Engine is a good choice.

## Solution

A posts searching engine indexed all the posts in user's personal network can solve first and second problem mentioned above. As for the third problem, a user search engine is the answer.

For post searching, it includes two part. Indexer and query part. The dataset used contains about a million posts downloaded from the internet. Posts' contents are indexed by Elasticsearch with ik Chinese analyzer. Elasticsearch has its own Chinese analyzer, which basically just split the sentence into single Chinese character. In contrast, ik divide a sentence into different common combination of phrases. Publish time is essential in query part, so four different formats of publish time is translated to one standard format. In the query part, ik Chinese analyzer is used to analyze query. This search engine use Elasticsearch default ranking function BM25 to rank the relevance between document and a certain query. For query independent score, two main elements is considered. The first one is number of shares, comments and likes, which can reflect the importance of the post. Post with large number of those three elements can be considered has higher value. And users want to re-read these post with a higher possibility. Those posts are more representative for users who wants to identify mainstream thinking or find resources. Based on the effort people have to make to complete those actions, comments takes up largest portion, shares second and likes least. The second element is publish time of posts. Query results are sorted by time in descending order. Fresh post might have higher possibility to be noticed by users. Also freshness is vital for people who cares about mainstream thinking.

User query is intended to help user strength current relationships or find potential peers with same hobbies. People tend to put what they mainly interested in in their username. So the user search query provide query on user's username. Indexer of user search part indexed the username of about 36 thousands users using Elasticsearch and ik analyzer. For query independent score, number of followers can show the importance of the person. But there are some exceptions. Some people might follow a large number of users. Among the users being followed, a portion of them will follow back for politeness or other reasons. For these people, number of their followers can't reflect their importance, so the number of followings should also be considered to measure importance of a user. Also interests of people can be reflected by their post. So user searching part can also retrieve users whose post containing tokens of query.

The project has a simple user interface with search toolbar and query results. The SERP shows main information of query results with highlight and link to original post and user. Highlight in the content helps user check if the retrieved content meet their information need quickly. And links to Weibo provide convenience for users. Example of user interface and search results are attached below.

## Conclusion

The project aim at facilitate user search and post search of Weibo. User can use user search to expand their personal network, and use post search to find posts. In future work, there are four types of improvements can be made. Firstly, I will use fresh dataset with more elements to support more search types. Secondly, I will update phrases dictionary. In the internet era, new phrases are constantly being created. It is essential to maintain an

up-to-date phrases dictionary for the analyzer. Thirdly, the search engine should support advanced search to meet user needs. Fourthly, indexing query logs, the project can support query suggestions. Fifthly, relevance feedback should be gathered from users to support query expansion. Finally, an evaluation of the search engine is needed to help modify the ranking function.

## Attachment


Figure1. home page of search engine

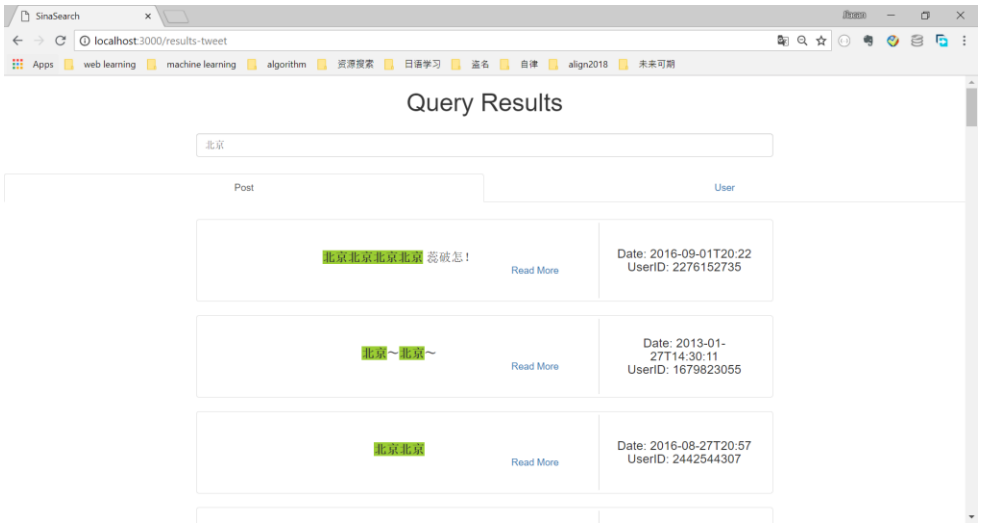

Figure 2. post query results of “北京”（Beijing）

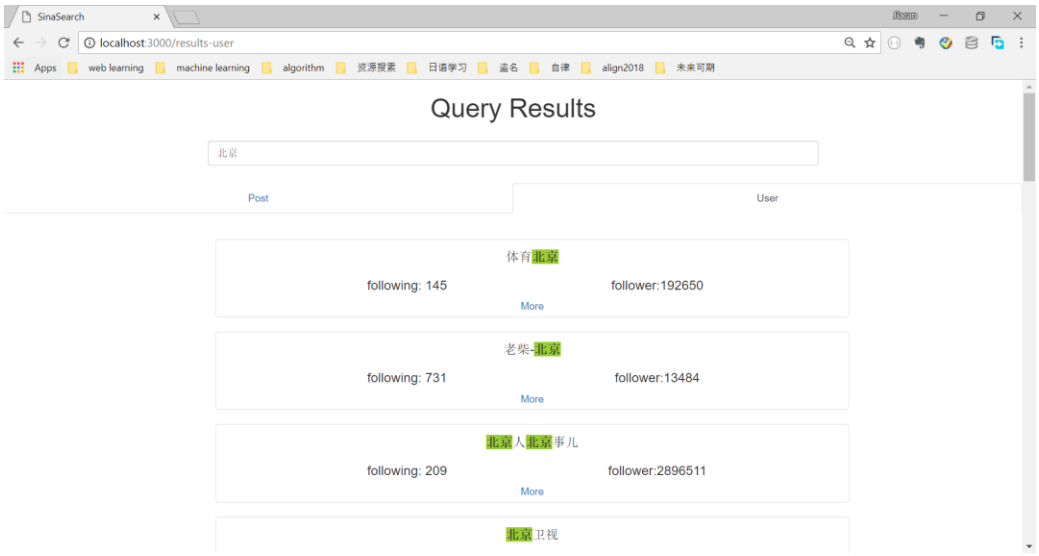


Figure 3. user query results of “北京”（Beijing）part one

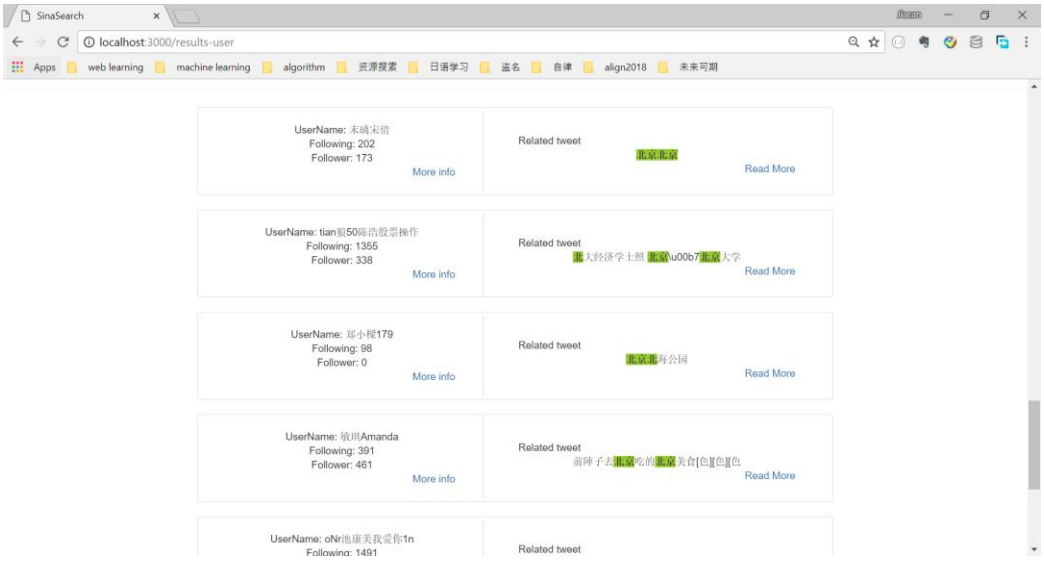


Figure 4. user query results of “北京”（Beijing）part two