

Homework2-data mining

2016年4月22日 星期五

上午8:40

1. Data is formed by 1249596 links in total from wikipedia, all dealt with the form

From_id: to_1, to_2, ...

Each of the link has at least 1 link to another page, but still you need to use the skills that prevent dead ends. And the data sometimes is quite sparse, so please use the proper structure to store and calculate.

2. Try a small number of data first, and then try the larger dataset. The test data contains only 8996 link-to-link data.

3. Write the report when you finished. And if you cannot run on the larger data.

4. Get the top 100 ranked link id in the report as the result.