1.Web Crawler

2. 我将此留给您提出一些简单的标准和算法来实现排名，例如单词出现在文档中的次数或您拥有的任何其他想法。但请解释一下你的方法。

3.我如何阅读网页？

回答：

如果您使用的是java，请使用jsoup：https：//jsoup.org/（链接到外部站点。）链接到外部站点..

如果您使用的是Python，请使用美味的汤：https：//www.crummy.com/software/BeautifulSoup/bs4/doc/

4. The core information stored by a search engine is a dictionary, called an inverted

index or inverted file, storing key-value pairs (w,L), where w is a word and L

is a collection of references to pages containing word w.

5. An array storing the occurrence lists of the terms (in no particular order)

• A compressed trie for the set of index terms, where each external node stores

the index of the occurrence list of the associated term.

Q: trie tree node 存什么 （只存index 指向 occurrence） 除掉stop word

Q: To facilitate the intersection computation, each occurrence list should be implemented with a sequence sorted by address or with a dictionary, which allows for a simple intersection algorithm similar to sorted sequence merging (Section 8.1).

Q:rank h1 里出现了几次？还是包括h2-6？