Goooooooooogle

BEATGOOGLE!

Proposal

TEAM2

104205002 彭光宇 104306007 李沛宸

105306008 姜姵君 105306018 林冠志 105306088 陳昌碩

◦ **Subject**

Subject
 Medical prevention,treatment, and regimen.

Websites for ranking
 https://www.cancer.org/
 https://www.cdc.gov/
 https://www.drugs.com/
 https://www.hhs.gov/

Keywords for ranking
 Cancer
 Healthcare
 Respiratory disease
 Regimen

◦ **System Specifications**

Users
 Any person who likes to type keywords to search for
 information

Purpose of using this system
 To give proper medical related information to users.

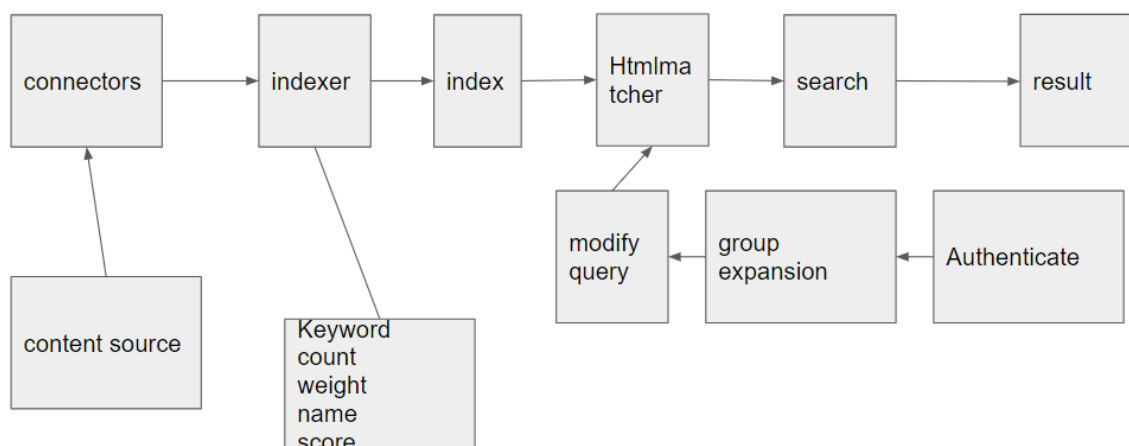**What is the function of the system,
and how to use it?**
 While users using this system to search,
 it will calculate the times of occurrence of the typed-in
 keywords in the websites,
 and then multiply it with the weights that users want it to be.
 By doing so, the system will identify relevance of the search
 result and the original keywords.
 Just enter keywords and it is ready to do search.

◦ Score Formula Suggestion

Input example: cold

For each contains	score = score + x
cold, colds	1
deal with, deals with, dealing with, cure, cures, curing, curable, relief, relieve, relieves, relieving	0.5
effective, effectively	1
Dr, Doctor (case sensitive)	2
report, study, research	1.5
caution, careful, beware	1

◦ Class Diagrams

◦ **Stage and Schedule**

Stage1
1. Given a set of keyword and URLs
2. Define score formula based on keyword appearance
3. Rank the initial set of URLs based on score
4. Return rank, score, the count on appearance of each keyword for each URL

Stage2
1. Given a set of keyword and URLs
2. Define score formula based on keyword appearance in the URLs and their sub
3. Rank the set of URLs based on score including all their sub URLs
4. Return rank, score, and a tree structure for its sub URLs along with the number of appearance of each keyword in each node

Stage3
1. Given a set of keyword (without URLs)
2. Use search engine(Google, Yahoo...) to get potential URLs
3. Apply the ranking on Stage2 to the URLs

Stage4 (12/31)
1. Derive relative keywords from the discovered Web sites
2. Iteratively do the same analysis on Stage3

Time Planning
System Structure Design (1~2 weeks)
Programming and Testing (6~7 weeks)
Adjustment and Demo Preparing (1~2 weeks)

○ **Challenge**

1. Allow the keyword to be in Chinese
2. Build a web site for our search engine