**Data Structure--Final Project Proposal**

Group 2-林貞妮、林聖典、彭琮鈺、傅俊益

<I>Find what? How to beat google?

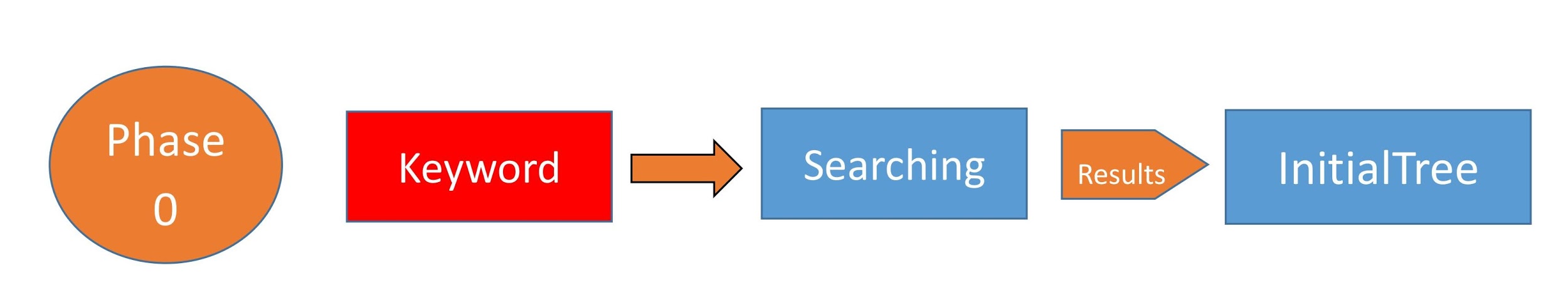
Discount restaurant. The more discount information is better.

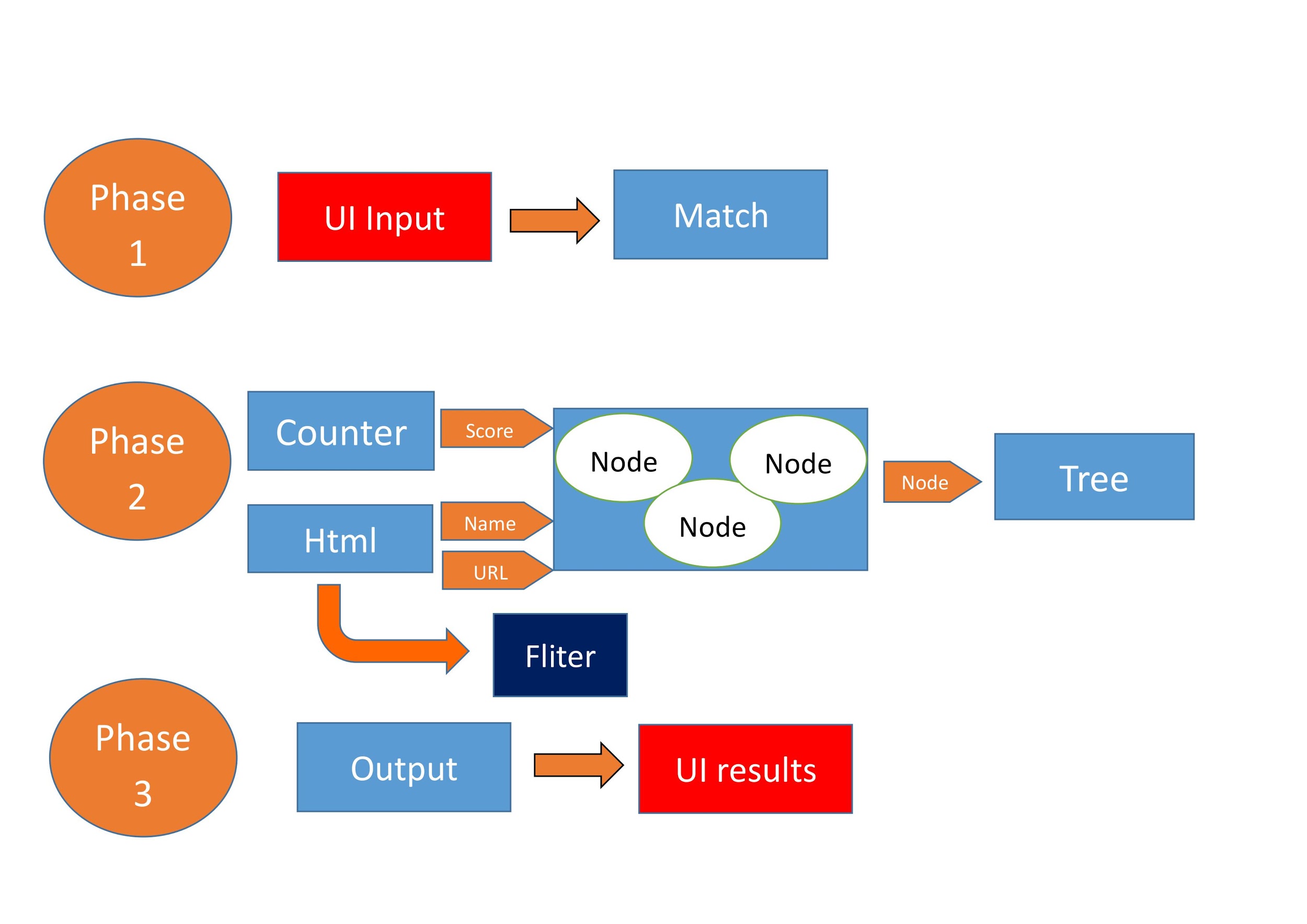
<II>Search tricks(Keyword)

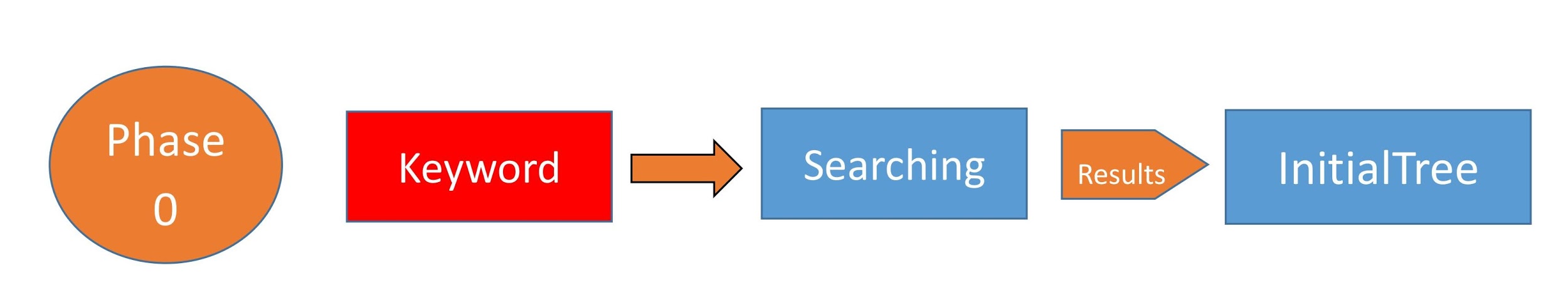
服務、環境、整潔、份量...

* 折扣：優惠、X人同行、好康、生日、起、學生證、折價、餐廳、折抵、打折、壽星、學生、特約

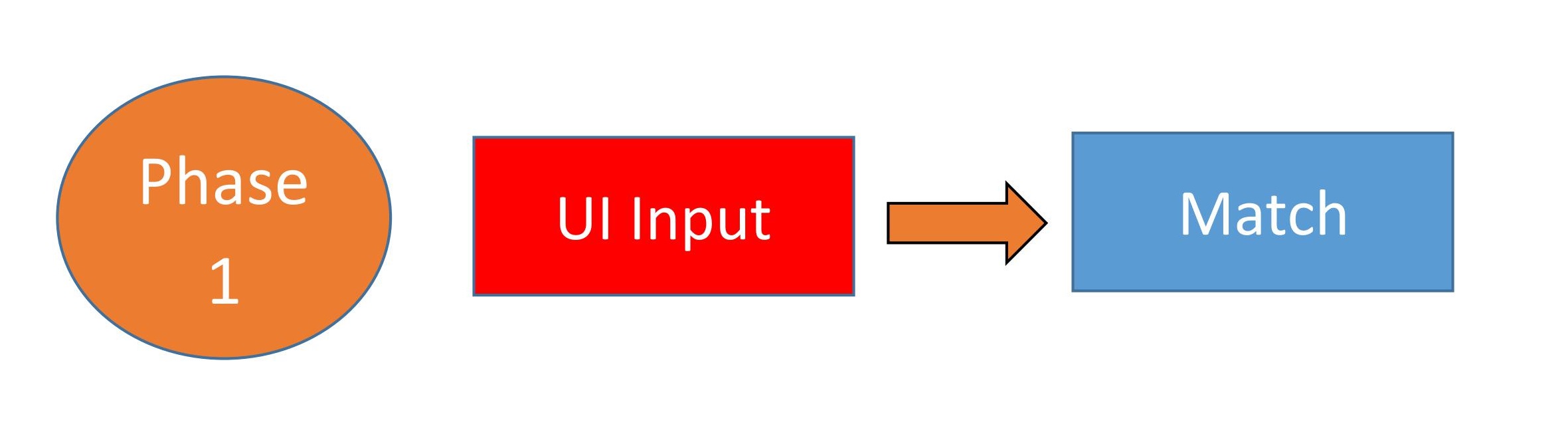
<III>System design



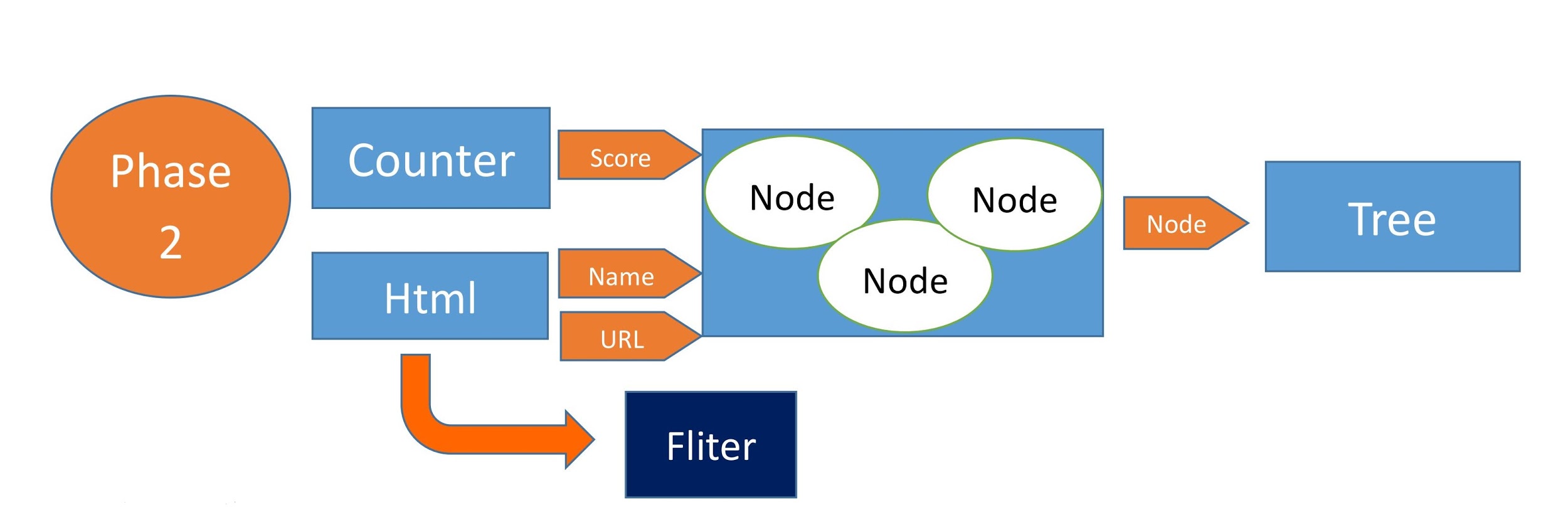




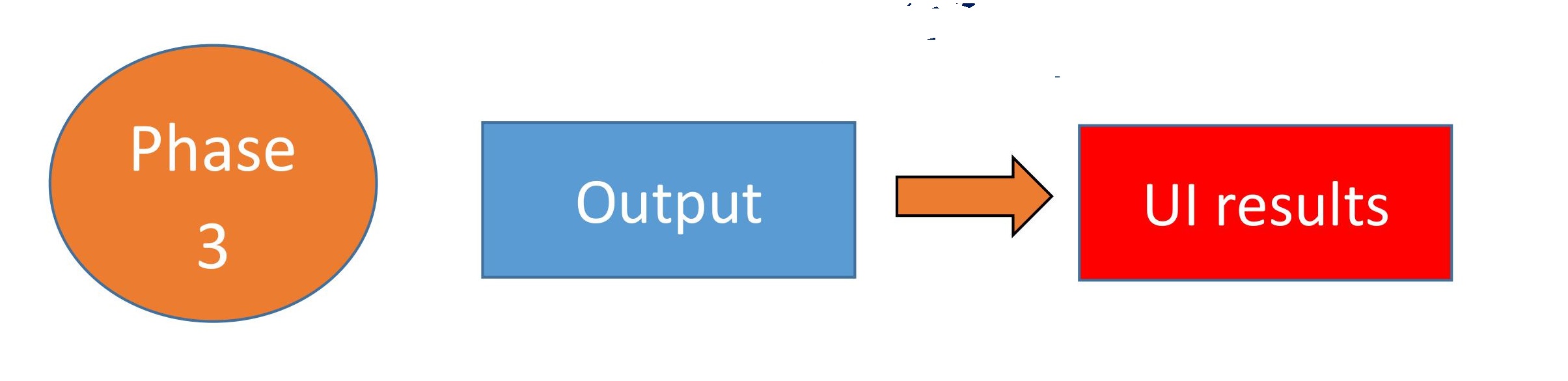
Phase 0: First, we set the keyword search engine to do the first search, the search results made into nodes and organized into InitialTree



Phase 1: create a user interface (UI), the user input keywords sent to InitialTree to compare whether there is a user input keyword



Phase 2: Tree1 matching search results made into Nodes, Node contains name, count, score, url and other attributes, where the result of the url by the HTMLTag method to catch the original code <a href=.....> Web site return derived. The result of the score is returned by the Keyword Counter operation. Finally, many nodes combined into a Tree, and use tree to help each node ranking.



Phase 3: pass the result of rank back to the page, and then output the top 20 scores of the search results

<IV>Method interface

Keyword:

public Keyword(String name, double weight)

public String toString()

WordCounter :

public WordCounter(String urlStr)

private String fetchContent() throws IOException

public int countKeyword(String keyword) throws IOException

WebPage:

public WebPage(String URL, String name)

public double setScore(ArrayList<Keyword> keywords) throws IOException

WebNode:

public WebNode(WebPage webpage)

public double setNodeScore(ArrayList<Keyword> keywords) throws IOException

public void addChild(WebNode child)

WebTree:

public WebTree(WebPage rootPage)

public double setPostOrderScore(ArrayList<Keyword> keywords) throws IOException

private double setPostOrderScore(WebNode startNode, ArrayList<Keyword> keywords)throws IOException

public void eularPrintTree()

private void eularPrintTree(WebNode startNode)

GoogleQuery:

public GoogleQuery(String searchKeyword)

public ArrayList<Keyword> disjointKeyword(String connectKeyword)

public HashMap<String, String> query() throws IOException

public ArrayList<Rerank> quickSort(ArrayList<Rerank> original)

Rerank:

public Rerank(String siteTitle, String url,double finalScore

public String toString()

Main:

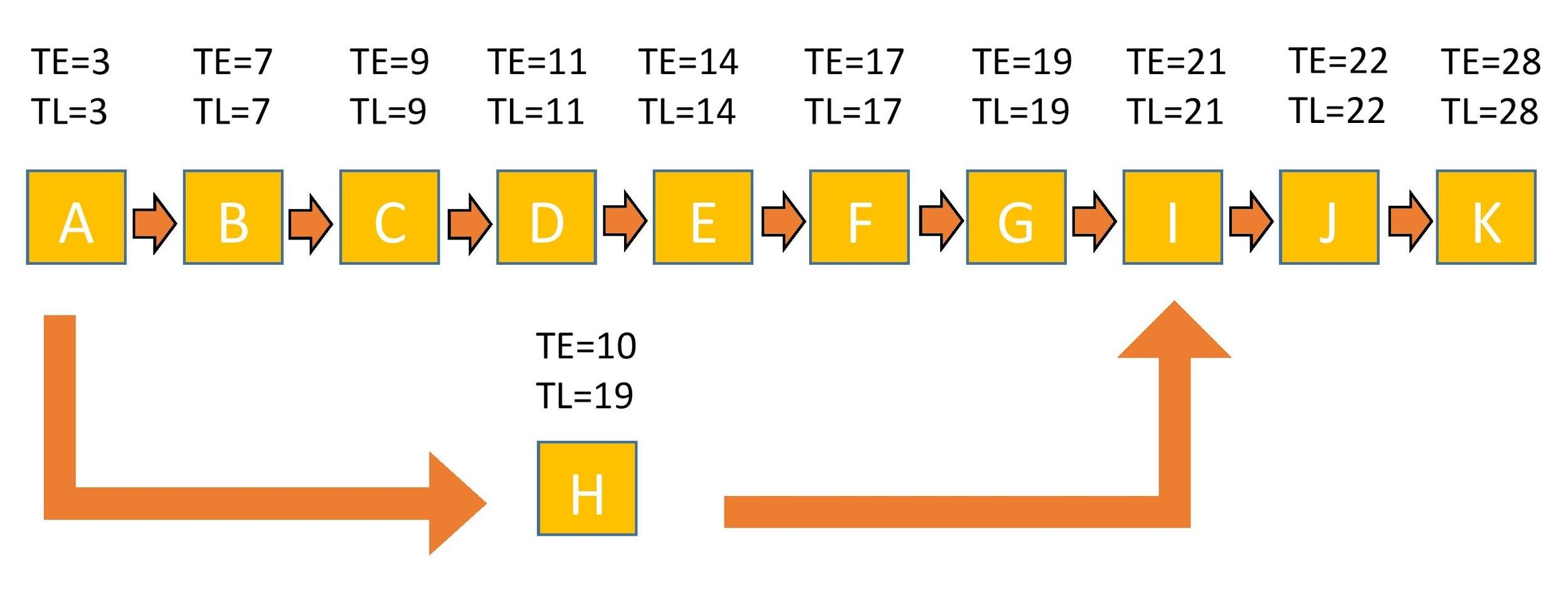
public static void main(String[] args) throws IOException

<V>Schedule(stage)

PERT

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Activity | Optimistic Time | Realistic Time | Pessimistic Time | Expected Time |
| (A)Requirements Collection | 1 | 2 | 3 | 3 |
| (B)Anlysis | 2 | 4 | 6 | 4 |
| (C)Keyword Counting | 1 | 2 | 3 | 2 |
| (D)Page Ranking | 1 | 2 | 3 | 2 |
| (E)Site Ranking | 2 | 3 | 4 | 3 |
| (F)Refine the rank of Google | 2 | 4 | 6 | 4 |
| (G)Semantics Analysis | 1 | 2 | 3 | 2 |
| (H)UI Design | 4 | 6 | 8 | 6 |
| (I)Testing | 2 | 4 | 6 | 4 |
| (J)Publish Your Work Online | 2 | 2 | 2 | 2 |
| (K)Export Your Work to App | 3 | 6 | 9 | 6 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Activity | TE | TL | Slack Time | On Critical Path |
| A | 3 | 3 | 0 | ｖ |
| B | 7 | 7 | 0 | ｖ |
| C | 9 | 9 | 0 | ｖ |
| D | 11 | 11 | 0 | ｖ |
| E | 14 | 14 | 0 | ｖ |
| F | 17 | 17 | 0 | ｖ |
| G | 19 | 19 | 0 | ｖ |
| H | 10 | 19 | 9 |  |
| I | 21 | 21 | 0 | ｖ |
| J | 22 | 22 | 0 | ｖ |
| K | 28 | 28 | 0 | ｖ |



<VI>Challenge

1. How to webrize our search engine?
2. How to make our results more precise?
3. How to make program run faster?



Group 2

Final presentation

Member: Henry Jason Jimmy Jennie

105306046,105306016,105306087,105306002



SeeFood

To find what you can not only eat delicious but also can be cheap

Contents

1. What does seefood do?(our idea)
2. Procedures in our project
3. Combine our homework
4. Set our default keywords
5. Call google search
6. Catch subsites
7. Create trees
8. Calculate scores of supersite and all of its subsites
9. Re-rank our supersites
10. Print out results
11. Webrize our search engine
12. What difficulties do we face?
13. Demo our search engine—seefood!
14. What does seefood do?(our idea)

* Seefood can find restaurants which have discounts for students.
* Surely, our goal is beating google with our better results.

1. Procedures in our project
2. Combine our homework
3. Set our default keywords
4. Call google search
5. Catch subsites
6. Create trees
7. Calculate scores of supersite and all of its subsites
8. Re-rank our supersites
9. Print out results
10. Webrize our search engine

A. Combine our homework

* We create a new class to help print out our results.

--Rerank

* We reuse those classes in our homework.

--Keyword, WordCounter, WebPage, WebNode, WebTree,

GoogleQuery

B. Set our default keywords

* We define some default keywords to make our results more precisely and meet our needs.

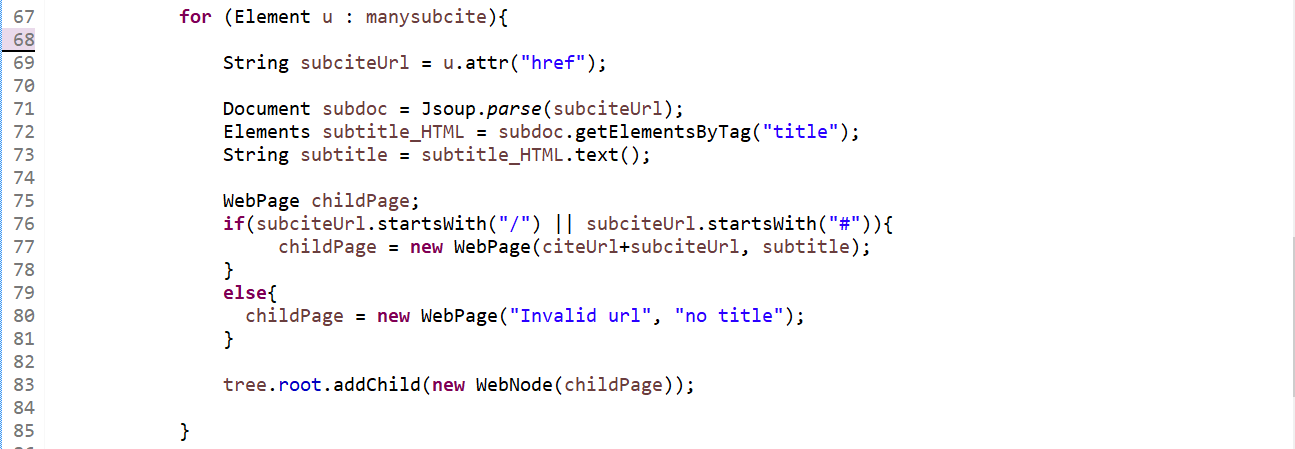
(以搜尋結果的網頁，去比對以下關鍵字是否有出現在那些網頁內，若有，則乘上權重10)

E.g. 折價、餐廳、優惠、折抵、折扣、打折、生日、壽星

C. Call google search

* Call google to search keywords(user input)
* Special: we use google built-in searching method to make our results more precisely
* We add “intitle:” in our call google search
* 結果雖然不到100個，但那是因為我們抓的是intitle，而且不會印出分數為0的網頁，所以結果較少，但更精準。

D. Catch subsites

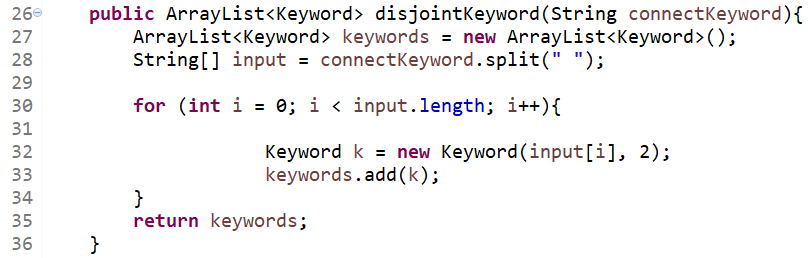
* We use a for loop to catch subsites in the for loop which we catch supersites.
* 

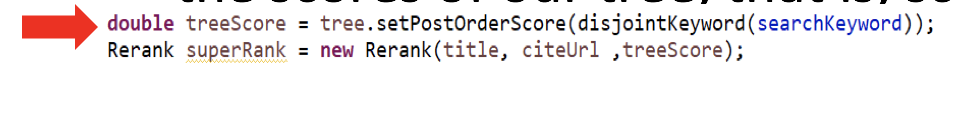
E. Create trees(一個母網頁一個)

* We use a for loop to catch subsites in the for loop which we catch supersites.
* Detail codes at right side.

F. Calculate scores of supersite and all of its subsites

• Before we calculate scores, we create a new method, disjointKeyword (String connectKeyword) to separate keywords if users input many keywords in a long string.



 • Then, we call setPostOrderScore() method in webtree class to calculate the scores of our tree, that is, scores of supersite and all of its subsites.

G. Re-rank our supersites

• We use QuickSort to re-rank our results(supersites).

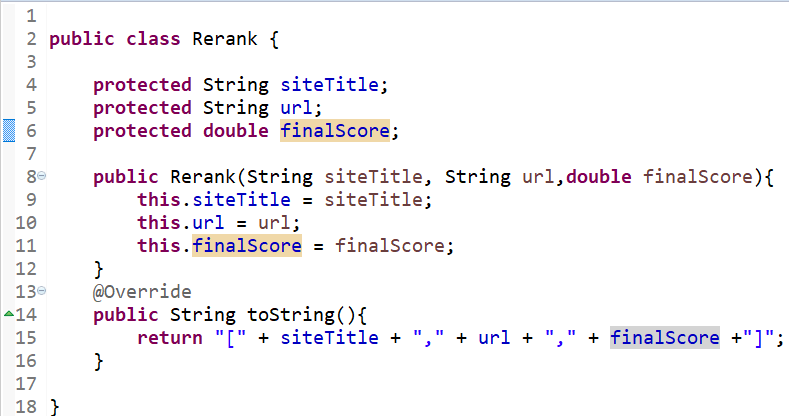


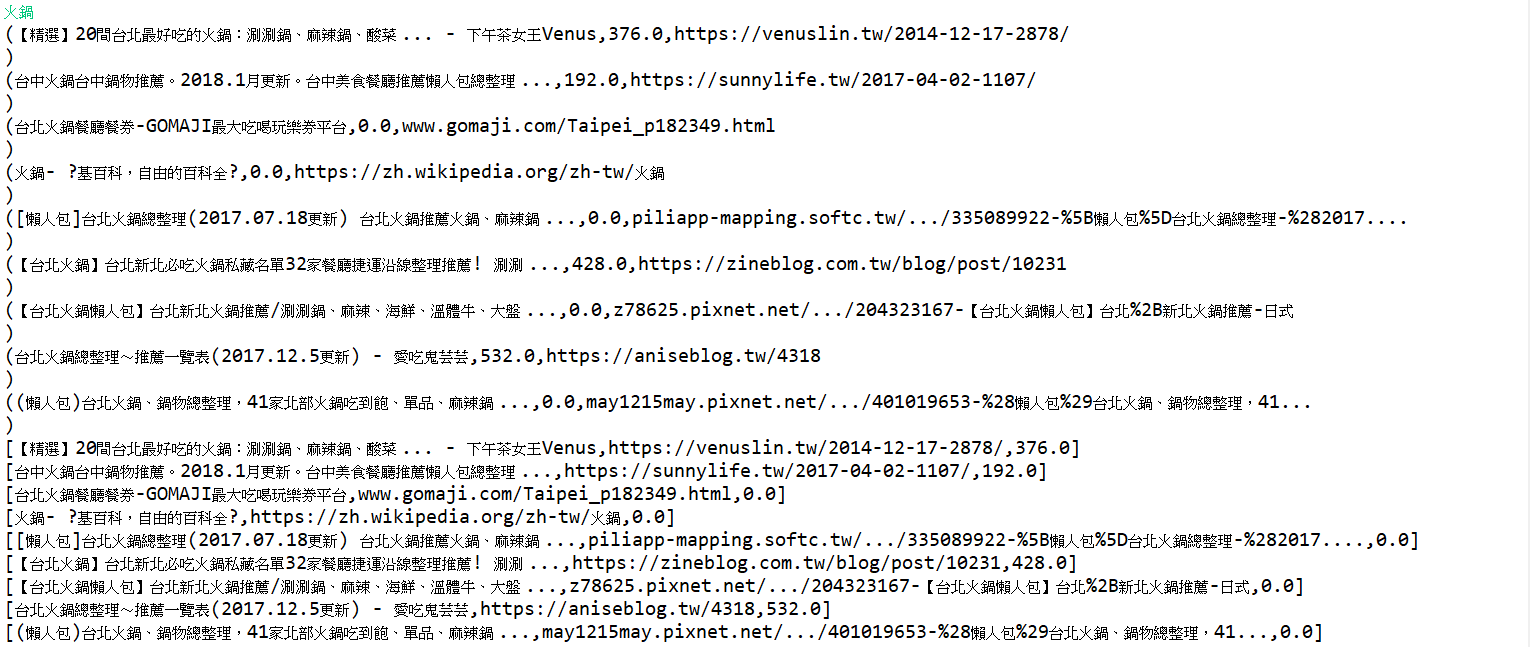
H. Print out results

• We create an object, called Rerank, to help us store three attributes of tree.

• siteTitle, url, finalScore

• Use an arraylist to store Reranks, and print it out.





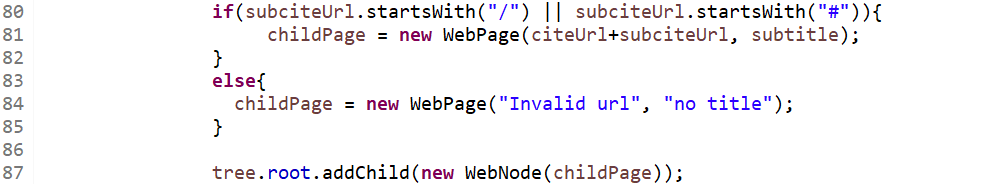
I. Webrize our search engine

•Use Apache-Tomcat to webrize our search engine.

1. What difficulties do we face?

•**MalformedURLException**

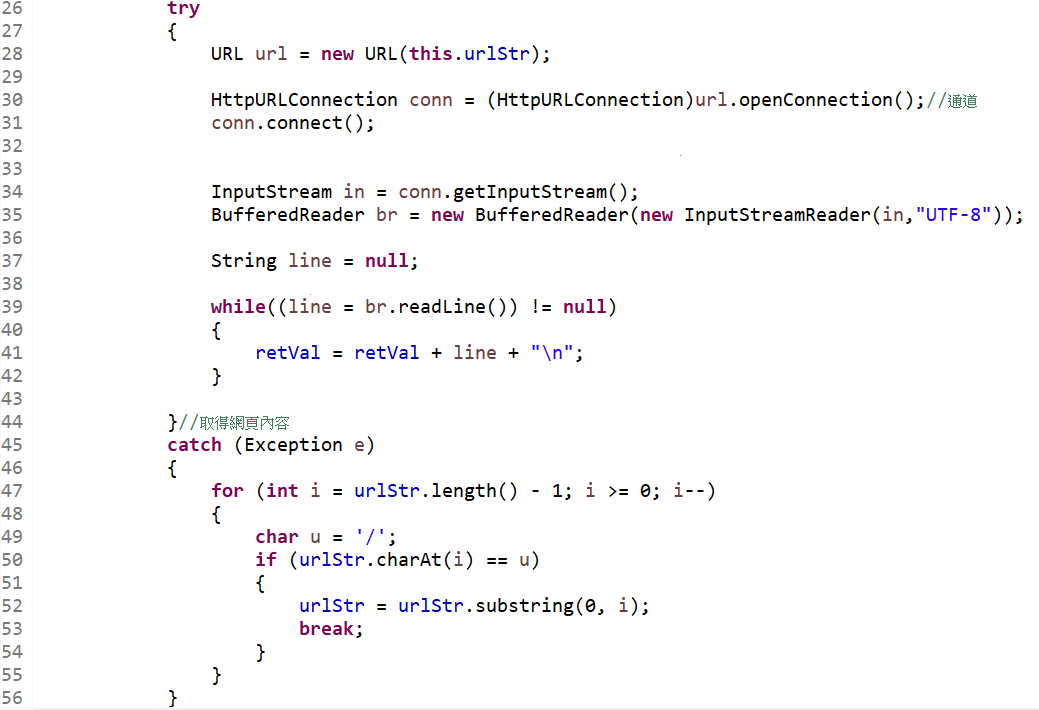
difficulty: subsites’ url has to start with “http://”, while the value in attribute “href” is not this case.

solution: we find that the value in attribute “href” should be attached after supersite’s url.

因為之前噴錯時，都是URL url = **new URL(this.urlStr);**這行出錯，推測應該是子網頁的網址出現問題，讓我們的程式跑不動，我們認為可能是網址後面是中文，讀進去後會是一堆亂碼，所以我們用**catch**內的**for loop**做，從子網頁的網址最後面往前推，，去抓**“/”**，找到較正確的子網頁網址去跑

**•NullPointerException**

difficulty: the url in attribute “href” isn’t complete, so java can’t load it.

 solution: we use a try-catch to solve this problem.

發現子網頁網址必須以http:// 開頭，所以我們用citeUrl+subciteUrl，也就是把母網頁的網址加上子網頁的部分，合成最後完整的網址

 4.Demo our search engine