Chapter 4

New Model of the Meta-Search Engine

4.1 INTRODUCTION

This chapter introduces a new model of meta-search engine. The new model of meta-search engine consists of two types of users:

i. Administrator
ii. End-user

The administrator module enables the administrator to control the meta-search engine databases to manage stop words and search engine’s URLs. It allows the administrator to update the merged database containing search results. It also allows the administrator to create a merge table to store search results from individual search engine databases whenever required.

The end-user module provides various options for searching to the user like Aggregate search, LIKE search, Selection search, and Direct search. It also allows the user to search any text as well as image and video links.

The proposed model of meta-search engine contains several modules as shown in Figure 4.1.

Figure 4.1: Overall model
4.2 THE MODEL

As a part of research work, a new model for meta-search engine has been designed and developed. The newly designed model has two phases: Phase – I and Phase – II. Figure 4.2 shows the working of the model.

Figure 4.2: The model of newly developed meta-search engine
Sequence of activities performed by the model:

i. The model initially takes input from user with or without stop words. Performs stop word elimination, if required.

ii. The model first search for availability of search terms in database. If it exists, displays resultant links on screen as per new ranking formula.

iii. If search term does not exist in database then prepares URL (Uniform Resource Locator) query string for input search text.

iv. Then model sends request to multiple search engines, retrieves search results, performs titles and URLs separation process and stores them in arrays.

v. Stores retrieved results from each search engine in different databases.

vi. Merges all individual search results available in separate search engine wise databases by eliminating duplications.

vii. Stores merged results in merge database.

viii. Find ranks using ranking formula.

ix. Displays resultant links on screen as per new ranking formula.

4.3 FEATURES OF A NEW MODEL OF THE META-SEARCH ENGINE

Existing meta-search engines do not have their own databases for indexing purpose. They directly send request to individual fixed number of search engines for user search text and retrieves aggregate single list of result by eliminating duplicates. They are using their own ranking formulas to display retrieved links in some specific order. This model introduces databases for indexing purpose. If relevant search keywords already available in database, then it retrieves results from it in an efficient way. Moreover, existing meta-search engine uses optimization techniques for business and uses strategies like Pay Per Click (PPC). Hence, ranking order is less reliable.
Features of the new model of meta-search engine:

i. Allows information seeker to enter search text character by character and select relevant search text string based on matches with keywords available in keyword database.

ii. Stop words elimination process eliminates stop words if required, in user input string (i.e. search text).

iii. It has its own database for indexing purpose. This helps to improve response time of meta-search engine.

iv. Developed new ranking formula for getting aggregate search results, offers competitive advantage to links for their position. This increases reliability in resultant links.

v. This new model can work on selected number of search engines. i.e. the facility of selection search is provided in newly developed model.

vi. In this new model LIKE search facility is also provided, where user can rate links of search results.

vii. Direct search facility is also provided.

viii. Administrator can manage a number of search engines and their URLs dynamically. It also enables administrator to update meta-search engine database periodically.

ix. Apart from text based web search it supports searching of image and video content also.

4.4 NEED OF STOP WORD ELIMINATION PROCESS

Existing meta-search engine like dogpile, mamma metacrawler, excite, etc. retrieves links based on user input.

If user input is with stop words, e.g. “What is Software Engineering?” on meta search engine like dogpile and mamma then output will be links based on whole search text as shown in Figure 4.4 and Figure 4.8 respectively. Here, there is a focus on user input keywords as well as stop words exist with user input.
Figure 4.3 shows input with stop words on dogpile meta-search engine.

Figure 4.3: Input with stop words on dogpile meta-search engine

Figure 4.4: Search results on dogpile meta-search engine with stop words

Designing Model for Meta-Search Engine
But if user input is without stop words like “Software Engineering” on same meta-search engines then resultant links based on focused keywords as shown in Figure 4.6 and Figure 4.10. Here, there is a focus on user input keywords only. Figure 4.5 shows the screen of dogpile in which input is with stop words.

The Figure 4.4 and Figure 4.6 shows that there is a significant difference (more than 90%) in resultant links by of dogpile meta-search engine. The Figure 4.8 and Figure 4.10 shows that there is a significant difference (more than 90%) in resultant links by mamma meta-search engine also.

Hence, Stop words should be filtered from an index for efficiency because people rarely search on just those words. [9]

**Figure 4.5: Input without stop words on dogpile meta-search engine**
Chapter – 4 : New Model of the Meta-Search Engine

Figure 4.6: Search results on dogpile meta-search engine without stop words

Figure 4.7: Input with stop words on mamma meta-search engine
Chapter – 4 : New Model of the Meta-Search Engine

Figure 4.8: Search results on mamma meta-search engine with stop words

Figure 4.9: Input without stop words on mamma meta-search engine
Figure 4.10: Search results on mamma meta-search engine without stop words

This feature is implemented in this new model of meta-search engine.

In this new model of meta-search engine the input, with stop words and without stop words, the list of retrieved links are collected and presented in Figure 4.12 and Figure 4.14.

If user input is with stop words, e.g. “What is Software Engineering?” on developed meta-search engine as shown in Figure 4.11 then output will be list of links based on focused keywords as shown in Figure 4.12.

It shows that it eliminates stop words like “what”, “is”, etc. from user input before sending request for search to web.
Chapter – 4 : New Model of the Meta-Search Engine

**Figure 4.11:** Input with stop words on a new model of meta-search engine

**Figure 4.12:** Search results on a new model of meta-search engine with stop words
Chapter – 4 : New Model of the Meta-Search Engine

Moreover, if user input is, “Software Engineering” on newly developed meta-search engine as shown in Figure 4.13 i.e. input without stop words then resultant links will be as shown in Figure 4.14.

![Image of Meta-Search Engine](image)

**Figure 4.13: Input without stop words on a new model of meta-search engine**

The resultant set of links retrieved by new meta-search engine with stop words (Figure 4.12) and without stop words (Figure 4.14) are nearly same (accuracy around 90%). This shows that there is a need of stop word elimination process.
Figure 4.14: Search results on a new model of meta-search engine without stop words

4.5 INCORPORATION OF DATABASES WITH NEW MODEL OF THE META-SEARCH ENGINE

Existing meta-search engines do not have their own databases as discussed earlier. So, there is no concept of indexing process in existing meta-search engines.

New model of meta-search engine with database, based on user input search text first finds keywords in the database whether search text containing keywords are available in it or not. If found then retrieves web information from the database by sending search query.
Figure 4.14 shows new model of meta-search engine with its own database as below.

![Diagram of Meta-search engine with database](image)

Figure 4.15: Meta-search engine with database

So, meta-search engine with database increases efficiency and will not have chance of timeout problem. Necessary thing here is, updation of database over a period of time to increase reliability of search results.

**4.6 IMPACT OF NEWLY INTRODUCED RANKING ALGORITHM AND OPTIMIZING TECHNIQUES**

In meta-search engine search results should appear on particular page with some rank. In designing meta-search engine, combining all search results is one task and giving some rank to each search result is another task. Here, focus is on how to produce a better combined ranking to search results. It is likely that search engines and meta-search engines are using some techniques for optimizing results by ranking mechanisms, which may be for marketing purpose. A new ranking formula is developed which offers competitive advantage to better positioned links.
Benefits of new ranking strategy:

i. Offers to maintain competitive rank position of a link on page if tracked on some of all selected search engines.
ii. This method increases reliability of search results.
iii. Particular link may retain its popularity as well.

In meta-search engines some search results should appear on first page with some rank, some should appear on second page with some rank, etc. Meta-search engine combining all search results is one task and another one is giving some rank to each search result. Here, focus is on how to produce a better combined ranking for better results. It is likely that search engines already use some techniques within their ranking mechanisms because a ranking algorithm needs to consider multiple factors.

In meta-search engine ranking can be done by focusing on rank positions on individual search results position on a particular page of search engine. For example, for a meta search engine if there is a use of three individual search engines then let have assumption that first page of each component has 5 search results (A, B, C, D, E) in which some of them may be common. For individual page results assign some integer value to the search result on a particular position.

If search result A is on first position of a page then assign rank as 5, if it is on second position then assign rank as 4 likewise if it is on last position then assign rank as 1. There may be a chance that search result A on different search engines on different position of a page. For example, On search engine 1 it is on 1st position, on search engine 2 it is on 2nd position, and on search engine 3 it is on last position. Then, as per this example rank of A on meta-search engine will be sum of result position on all three individual search engine pages. i.e. (5+4+1=10). Similar way if search result B on one search engine on 1st position, on second search engine 2nd position and on third one it is on 1st position then its rank will be (5+4+5=14).
Another alternate method wants to suggest using count function. For example, if search result A is common of all three individual search engines of meta-search engine then the count will be 3. If it is common in any two of them then count will be 2. If it is in one of them then count will be 1. Here, there is no focus on position on a page of individual component search engine. Then render results in descending order of count by avoiding duplicate entries.

There can be role of position of search result on a page of individual component search engine with this count function related method. For example, there are five search results (A, B, C, D and E) on a page of 3 individual component search engines. Table 4.1 shows their rank as per position.

### Rank as per position

<table>
<thead>
<tr>
<th>Rank as per position</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search Engine 1</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Search Engine 2</td>
<td>4</td>
<td>-</td>
<td>3</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>Search Engine 3</td>
<td>5</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>-</td>
</tr>
</tbody>
</table>

**Table 4.1: Rank as per position**

Following table 4.2 shows search result count for this 5 search results.

### Search Result Count

<table>
<thead>
<tr>
<th>Search Result</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>3</td>
</tr>
<tr>
<td>B</td>
<td>2</td>
</tr>
<tr>
<td>C</td>
<td>3</td>
</tr>
<tr>
<td>D</td>
<td>2</td>
</tr>
<tr>
<td>E</td>
<td>2</td>
</tr>
</tbody>
</table>

**Table 4.2: Search result count**
Chapter – 4 : New Model of the Meta-Search Engine

Formula suggested here for ranking in a new model of meta-search engine is combination of above both methods

\[ \text{Rank}(x) = \frac{\sum_{i=1}^{n} P_i(x)}{\text{count}(x)} \] for all \( i \)

Where, value of \( n \) = total number of search engines
\( P_i(x) \) = Position of search result link on \( i^{th} \) search engine
\( \text{Count}(x) \) = Availability of search result link on \( n \) number of search engines

**Search Resultant Rank based on newly introduced ranking formula**

<table>
<thead>
<tr>
<th>Search Result</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>4.66</td>
</tr>
<tr>
<td>B</td>
<td>3.00</td>
</tr>
<tr>
<td>C</td>
<td>3.33</td>
</tr>
<tr>
<td>D</td>
<td>2.50</td>
</tr>
<tr>
<td>E</td>
<td>1.00</td>
</tr>
</tbody>
</table>

*Table 4.3: Ranking based on new rank assigned*

As per above table meta-search engine will render result in descending order of rank and will get their position as below.

<table>
<thead>
<tr>
<th>Position #</th>
<th>Search Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>C</td>
</tr>
<tr>
<td>3</td>
<td>B</td>
</tr>
<tr>
<td>4</td>
<td>D</td>
</tr>
<tr>
<td>5</td>
<td>E</td>
</tr>
</tbody>
</table>

*Table 4.4: Position based on new rank*
4.7 SUMMARY

Approaches for information retrieval from the web using a new model of meta-search engine have been presented. This chapter gives modular design of a new model, presents work flow of a new model and discusses various features of a new model. Also it presents need of stop word elimination process by sending input query to existing meta-search engine with or without stop words. The result shows that there is significant difference (more than 90%) in resultant links. The same feature is implemented in new model of meta-search engine and result shows that it retrieves resultant links almost same (> 90 % accuracy) with and without stop words. It also discusses about a new ranking formula for listing search results on screen by the proposed model.