

## CREATIVE QUERIES FOR EXPLORATORY SEARCH

### ABSTRACT:

This article discusses our development of a new interactive search engine interface which generates more explorative and creative results, exceptional from typical accurate ad hoc search results. We will discuss in detail about the implementation of efficient algorithms on input queries to generate creative queries. This paper also discusses about different proposed interfaces and their evaluation techniques which resulted in the current interactive interface.

### MOTIVATION

Usage of traditional Search Engines like Google, Yahoo & Bing, will always get you the most accurate and point to point results. These Search engines try to know what users want by utilizing the factors like cookies and geographical location. For example when you search for a keyword “Manam” which is a tollywood movie title from a system which is either acquainted with this language or location, search engine will give you the list of results of which all of them on the first page would be related to the movie itself. But when the same keyword when used on a different system(from different geographical location and source of language) which is no way related to this domain, search results would have two different fields, in which one is related to the movie and the other to a Thai restaurant in Munchi.

Traditional search engine suits really best for all those users who knew what they want and probably who doesn't want to know the twisted solutions of the same problem. But for the users who would like to view results in a broader prospect, traditional search engine will narrow down its scope. Hence a search engine which is explorative in getting results and which can give a user different aspects of problem solving techniques rather than getting accustomed to the accurate search methodology is needed.

### INTRODUCTION

Exploratory Search[1] is for those users who are unfamiliar with the domain of their goal and even unsure about the ways to achieve their goals. Explorative Search's goal is to get a broader results view of the search query, which is achieved by manipulating the original search term using various algorithms and then regenerating new query terms. The results which are based on these new query terms are unexpected but relevant to the content of original search term. Hence a user is exposed to wider concept of the search term rather than getting narrowed to the accurate results.

### RELATED WORK

In regard with this concept, Raczinski et. al [2] & Hendler et. al[3] suggested that, in contrast to the traditional search engine's semantic web technologies, in exploratory search engine, relationships between items are exploited in a total random fashion so as to generate new interesting and unusual relations. Syzygy, Clinamen and Anomaly are the different techniques used to collect different random but connected items on web according to [1] & [2]. The new collected items are then fed to traditional search engine to generate explorative search results.

## ALGORITHMS AND APPROACH

### Syzygy:

This methodology uses WordNet dictionary. It is a large lexical database of English. Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. Syzygy of a term is formed by intersecting the union of hyponyms, holonyms and hypernyms of the term with the original vocabulary of the term.

For a search term  $t$

$$\text{syzygy}(t) = \{ h : h \in \text{union}(t) \wedge \exists h \in V \}$$
$$\text{union}(t) = \text{hypo}(t) \cup \text{hyper}(t) \cup \text{holo}(t)$$
$$\text{hypo}(t) = \{ h : h \in \text{hyponyms}(s) \}$$
$$\text{hyper}(t) = \{ h : h \in \text{hypernyms}(s) \}$$
$$\text{holo}(t) = \{ h : h \in \text{holonyms}(s) \}$$
$$\text{syno}(t) = \{ s : s \in \text{synonyms}(t) \}$$

for  $s \in \text{syno}(t)$

For e.g. Let  $t = \{ \text{live} \}$ ;

$$\text{syno}(\text{live}) = \{ \text{populate, inhabit, } \mathbf{be}, \mathbf{domicile} \};$$
$$\text{hypo}(\text{populate, inhabit, be, domicile}) = \{ \text{cliff, dwelling, house ...} \};$$
$$\text{hyper}(\text{populate, inhabit, be, domicile}) = \{ \text{be, fill up, exist, residence, abode,....} \}$$
$$\text{hypo}(\text{populate, inhabit, be, domicile}) = \{ \text{domicile, reside, camp, tent, nest} \}$$
$$\text{union}(\text{hypo, holo, hyper}) = \{ \text{populate, } \mathbf{be}, \text{ fill up, house, exist, } \mathbf{domicile}, \text{ camp,....} \}$$
$$\rightarrow \text{Syzygy}(\text{live}) = \{ \mathbf{be}, \mathbf{domicile} \}$$

### Anomaly:

Anomaly function simply makes use of WordNet's antonyms. At first, all the synonyms for query term are found and then antonyms for synonyms are generated. Words common to direct antonyms and synonym's antonyms are chosen to be anomaly output.

For a search term  $t$

$$\text{antinomy}(t) = \{ h : h \in \text{anto}(t) \wedge \exists h \in V \}$$
$$\text{anto}(t) = \{ h : h \in \text{antonyms}(s) \}$$
$$\text{syno}(t) = \{ s : s \in \text{synonyms}(t) \}$$

for  $s \in \text{syno}(t)$

For e.g.  $\text{antonym}(\text{live}) = \{ \mathbf{dead}, \text{recorded} \}$

$$\text{synonym}(\text{live}) = \{ \text{alive, animate, breathing} \}$$
$$\text{antonym}(\text{alive}) = \{ \mathbf{dead} \}$$
$$\text{antonym}(\text{animate}) = \{ \mathbf{dead}, \text{inactive, inhibit..} \}$$
$$\text{antonym}(\text{breathing}) = \{ \text{breathless, } \mathbf{dead} \}$$
$$\text{anomaly}(\text{live}) = \{ \mathbf{dead} \}$$

### Clinamen:

The Clinamen function uses the Damerau-Levenshtein algorithm which measures the distance between two strings.

For a search term  $t$

$$\text{Clinamen}(t) = \{ v : 0 < \text{dameraulevenshtein}(t, v) \leq 2 \}, \text{ for } v \in V$$

For e.g. Clinamen of LIVE = LOVE, LIES, SIZE, RIVER

## E\_ALGORITHMS

Basic algorithms of Syzygy and Anomaly with slight modifications are implemented in this interface. Clinamen is completely ignored due to its enormous distinctiveness. The essential prerequisite, unexpected but relevant search results of Explorative Search Engine interface is lost by adopting Clinamen algorithm and hence is ignored.

### E\_Syzygy

E\_Syzygy algorithm functionality is quite similar to that of Syzygy except for the resultant terms, instead of union of hypernyms, holonyms and hyponyms, intersection is considered. This modification is made due to the fact that, it is ought to take only one word from the resultant terms and it is because of the final phase of implementation(See reference Interactive Interface), where a User is provided feasibility to add any other term into creative query, if perceived necessary. Resultant set of terms would be very large and not effective if union is considered. On the other hand, taking Intersection into account, it has two advantages.

1. Lesser number of resultant terms.
2. The term selected will belong to all the three relations(Hypernym, Holonym, Hyponym) maintaining compatibility.

A term that is randomly selected from the resulting terms which when combined with the original term yield E\_Syzygy query.

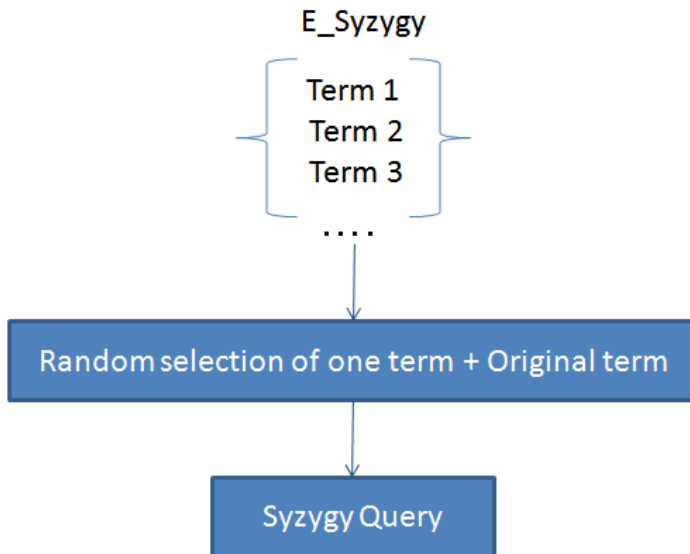


Figure 1: Explorative Syzygy Query Generation

For a search term  $t$

$$\text{syzygy}(t) = \{ h : h \in \text{rand}(\text{intersection}(t)), t \}$$

$$\text{intersection}(t) = \text{hypo}(t) \wedge \text{hyper}(t) \wedge \text{holo}(t)$$

$$\text{hypo}(t) = \{ h : h \in \text{hyponyms}(s) \}$$

$$\text{hyper}(t) = \{ h : h \in \text{hypernyms}(s) \}$$

$$\text{holo}(t) = \{ h : h \in \text{holonyms}(s) \}$$

$$\text{syno}(t) = \{ s : s \in \text{synonyms}(t) \}$$

for  $s \in \text{syno}(t)$

Rand function randomly selects one term from intersection set.

For e.g. Let  $t = \text{live}$

$\text{syno}(\text{live}) = \{ \text{populate, inhabit, be, domicile} \}$

$\text{hypo}(\text{populate, inhabit, be, domicile}) = \{ \text{cliff, dwelling, house ...} \}$

$\text{hyper}(\text{populate, inhabit, be, domicile}) = \{ \text{be, fill up, exist, residence, abode, ....} \}$

$\text{hypo}(\text{populate, inhabit, be, domicile}) = \{ \text{domicile, reside, camp, tent, nest} \}$

$\text{intersection}(\text{hypo, holo, hyper}) = \{ \text{be, go through, experience, see} \}$

$\text{E\_Syzygy} = \{ \text{live} \}$

$\text{E\_Syzygy Query} = \{ \text{be, live} \}$

## E\_Anomaly

Similar to the Anomaly algorithm, antonyms of synonyms are generated. Unlike the original Anomaly function, from the union of antonyms of synonyms, a random term is directly selected and combined with the original term to yield E\_Anomaly query. The embodied change is imperative due to following reasons.

1. To acquire more diverse results.
2. To maintain low processing over head.

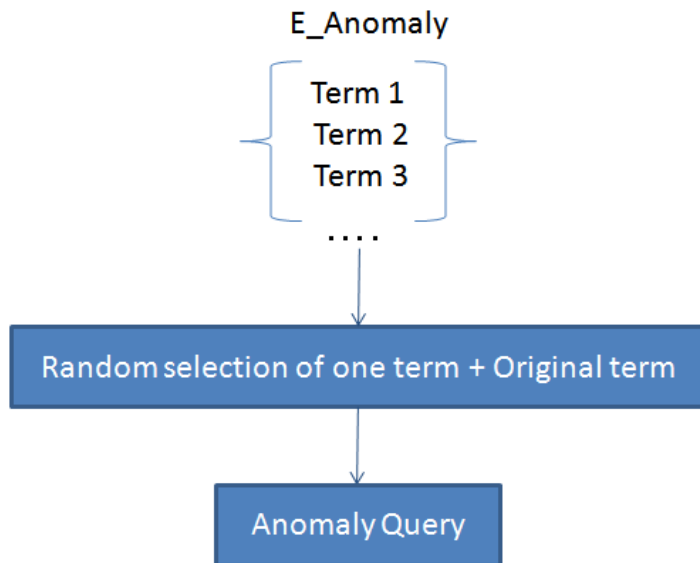


Figure 2: Explorative Anomaly Query Generation

For a search term  $t$

$\text{anomaly}(t) = \{ h : h \in \text{rand}(\text{anto}(t)), t \}$

$\text{anto}(t) = \{ h : h \in \text{antonyms}(s) \}$

$\text{syno}(t) = \{ s : s \in \text{synonyms}(t) \}$

for  $s \in \text{syno}(t)$

Rand function randomly selects one term from intersection set.

For e.g.  $t = \text{live}$

$\text{synonym}(\text{live}) = \{ \text{alive, animate, breathing} \}$

$\text{antonym}(\text{alive, animate, breathing}) = \{ \text{dead, inactive, inhibit, recorded, breathless, ....} \}$

$\text{E\_Anomaly}(\text{live}) = \{ \text{recorded} \}$

$\text{E\_Anomaly Query}(\text{live}) = \{ \text{recorded, live} \}$

## ARCHITECTURE

The diagram below depicts the basic architecture of Exploratory Search Engine.

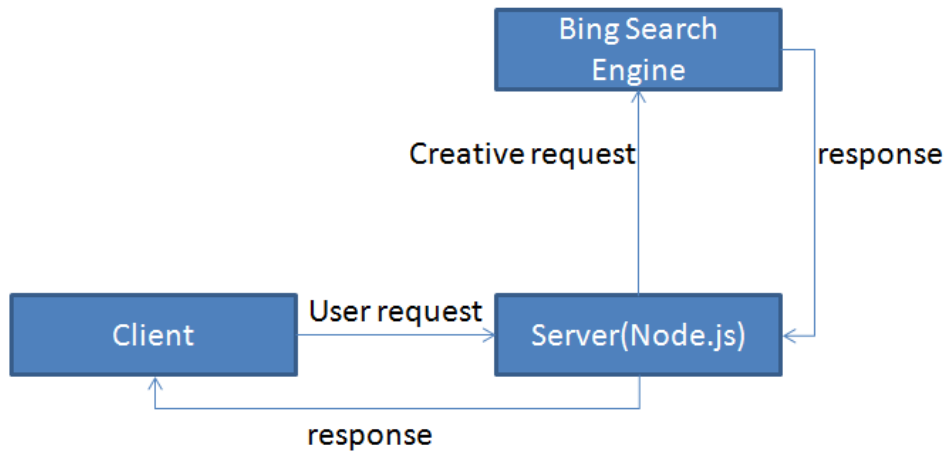


Figure 3: Basic architecture of Explorative Search Engine

The basic architecture consists of three modules

1. Client:

This module includes front end application through which a user can access the interface. A User request is directed from this module to the Server module.

2. Server:

This module handles the User request accordingly and generates a response based on the implemented algorithms. Server side scripting is done in Node.js. In-detailed functionality of Server is discussed later.

3. Bing API:

This module allows to embed or consume bing search results into Explorative interface using JSON. Source types of web and spelling suggestions are returned to server which are then consigned to client module.

## Functional Diagram of Server :

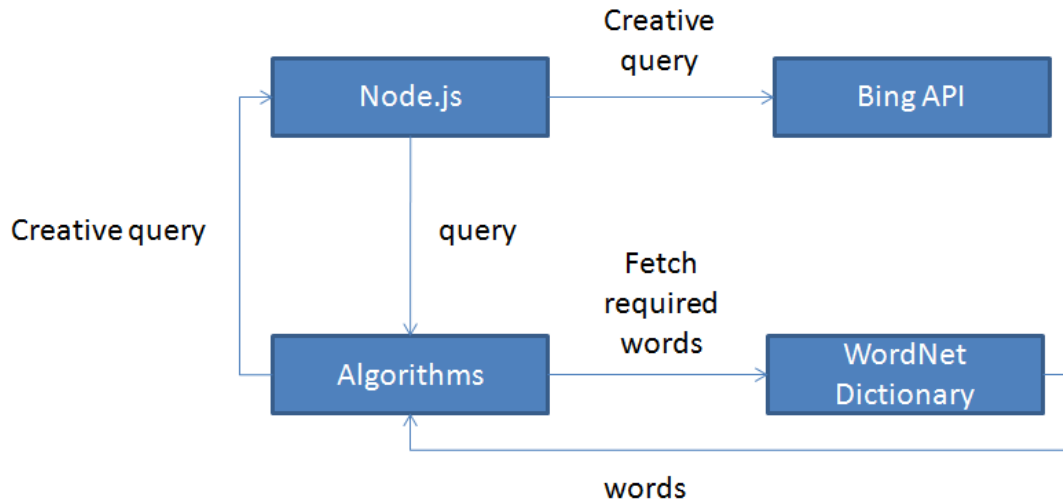


Figure 4: Functional Diagram of Server

Node.js is highly scalable (not implemented on multi threading functionality). Since that every system has a limit of creating threads, in the current case, parallelism is done in multiple levels for single user requests. For multiple users, it handles requests more efficiently. When the user request is received, Node.js transforms the initial query into creative query according to the implemented algorithms. These algorithms fetches required word relations from the WordNet dictionary using WordNet-magic(API). The evolved creative query is then used to fetch results from Bing API which are directed as the response to the client.

## IMPLEMENTATION

The diagram below depicts the main functionality of the Server.

**File Reader:** It is responsible for returning HTML, CSS, Java Script and IMG files.

**Spelling Suggestion:** It handles the request for Spelling Suggestions from Bing API for Client.

**User Request:** It handles User requests for Search queries.

**More Results:** It handles the request to explore more Explorative results.

**Response:** This module is responsible for generating response to the Client.

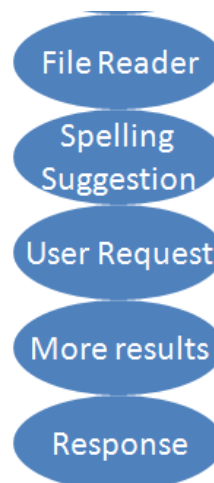


Figure 5: Main modules of Server

### User Request Handler:

All the User requests are processed by initiating Explorative and Conventional algorithms. The collective results forms the response structure.

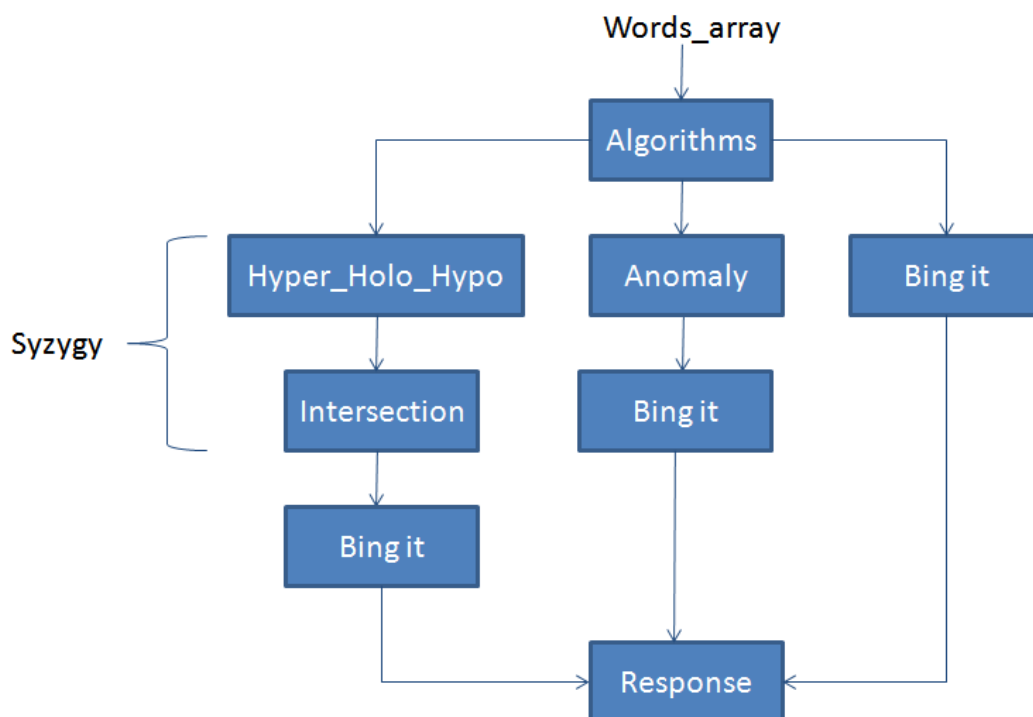


Figure 6: User Request Handler

Since Node.js works in an asynchronous tone, multiple levels of parallelism could be deployed as shown in the figure below.

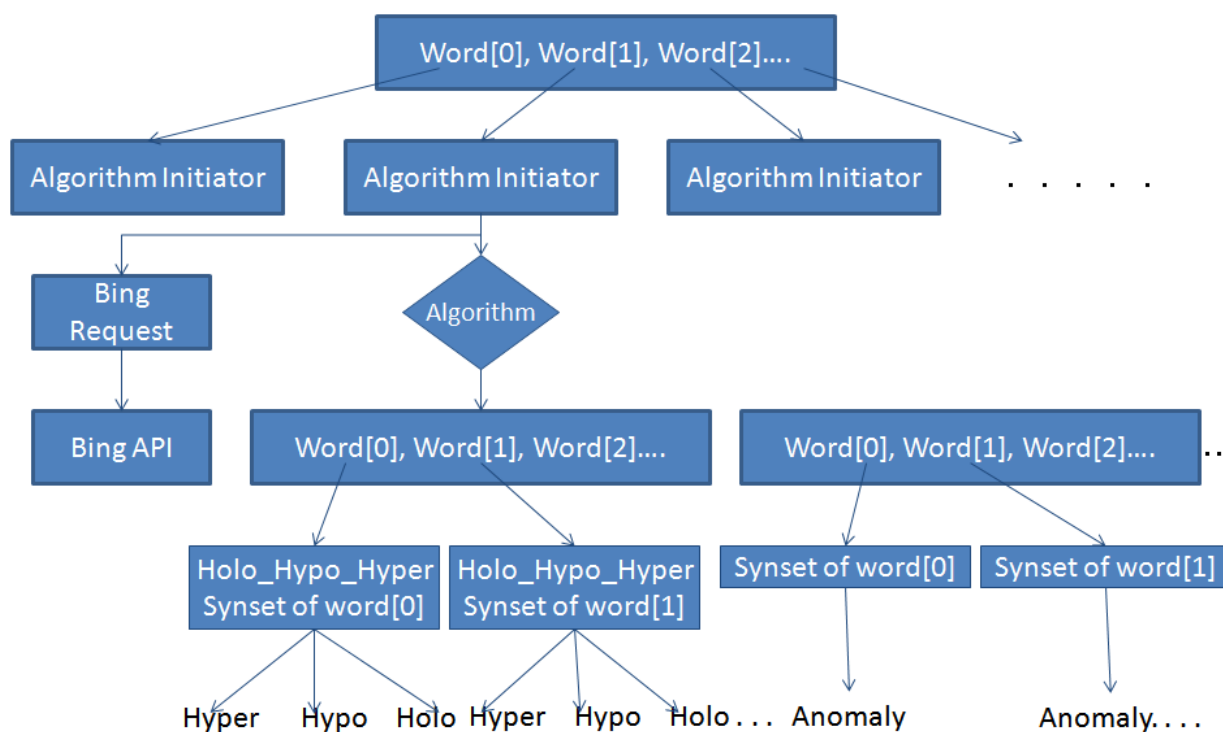


Figure 7: Multiple levels of parallelism

In a multiple words query, as soon as the process commences, individual algorithm initiators works on individual words. Algorithm initiator, in turn provokes Syzygy, Anomaly and Conventional algorithms. Sequentially, these algorithms calls their respective functions of Hypernyms, Holonyms, Hyponyms, Synonyms and Antonyms. And thus the process advances in harmony with each other deriving multilevel of parallelism

### Response Generator:

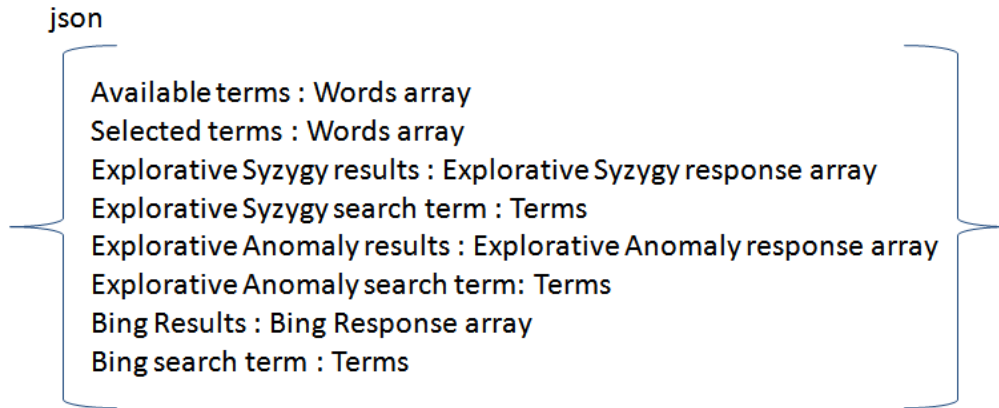


Figure 8: Server Response

The final response generated by the server in the json format is as shown in the figure above.

## USER INTERFACE DEVELOPMENT

This chapter discusses about proposed interfaces to display the results of creative queries and a survey report used to evaluate them.

Following are the three different front end interfaces contemplated to present the results. Out of which an interface is elected to be implemented based on the evaluation results.

### Query Display Interface:

It is one of the interface in which a user is exposed to different web pages overlapping on each other with visible query terms from which the results are generated. It is the user choice to select the suitable page based on the visible query terms.

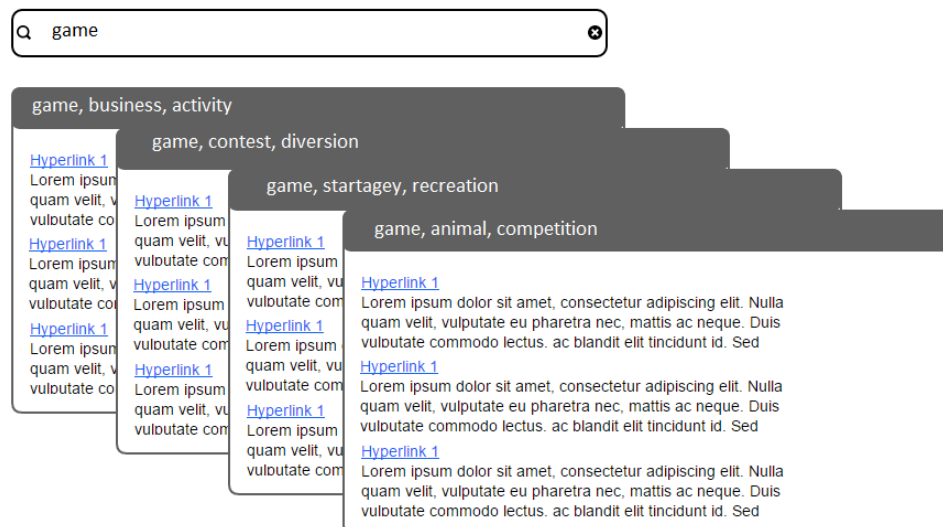


Figure 9: Query Display Interface



## Grid Display Interface

In this interface, images of website logos related to different creative query term sets are displayed in the different grids as shown below. User has the feasibility to choose one amongst them. Results are displayed considering only those websites linked to the grid.

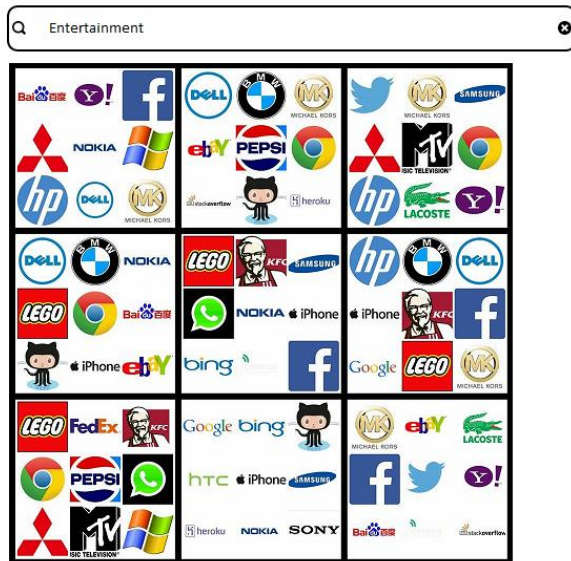


Figure 10: Grid Display Interface

## Interactive Interface

In this interface user has given complete feasibility to choose query terms generated from original query. It requires user interaction to select creative query terms from the available query terms list. User can both select and deselect terms from selected terms list.

Available Terms

activity
competition
contest
diversion
recreation
animal

Selected Terms

game
business
strategy

Submit

Search results

[Hyperlink](#)

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nulla quam velit, vulputate eu pharetra nec, mattis ac neque. Duis vulputate commodo lectus, ac blandit elit tincidunt id. Sed rhoncus, tortor sed eleifend tristique, tortor mauris molestie elit, et lacinia ipsum quam nec dui. Quisque nec mauris sit amet elit iaculis pretium sit amet quis magna. Aenean velit

Figure 11: Interactive Interface

## Preliminary Evaluation

Considering the above three interfaces, a preliminary survey has been conducted to assess each of them. The survey evaluate factors like user's conception towards creative queries, interface's complexity, supportiveness towards the goal of Explorative search engine and effectiveness.

- According to the survey, it is analyzed that almost 62.2% of candidates are interested in Exploratory search and can adopt to such interface if implemented.
- Out of all, 70% of the participants are students who use search engines mainly for academic/research purposes. Most of the students first preference is Interactive Interface and then comes Query Interface.

## AGREEMENT MODEL

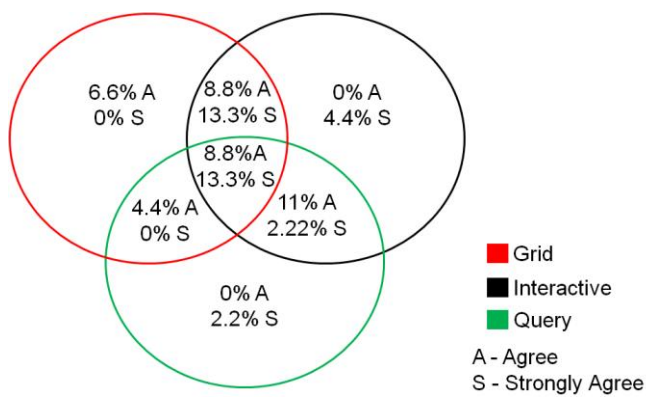


Figure 12: Agreement Model

The above model shows only the agreement relations between the three interfaces. In this case disagreement status is not considered.

According to the statistics mentioned in the diagram, percentage of strong agreement is more for Interactive Interface as compared with others. Though the individual support percentage of Interactive Interface is lower than that of Grid Interface, an overall high voting rate of 61.82% is achieved by the Interactive Interface.

## AGREE - DISAGREE MODEL

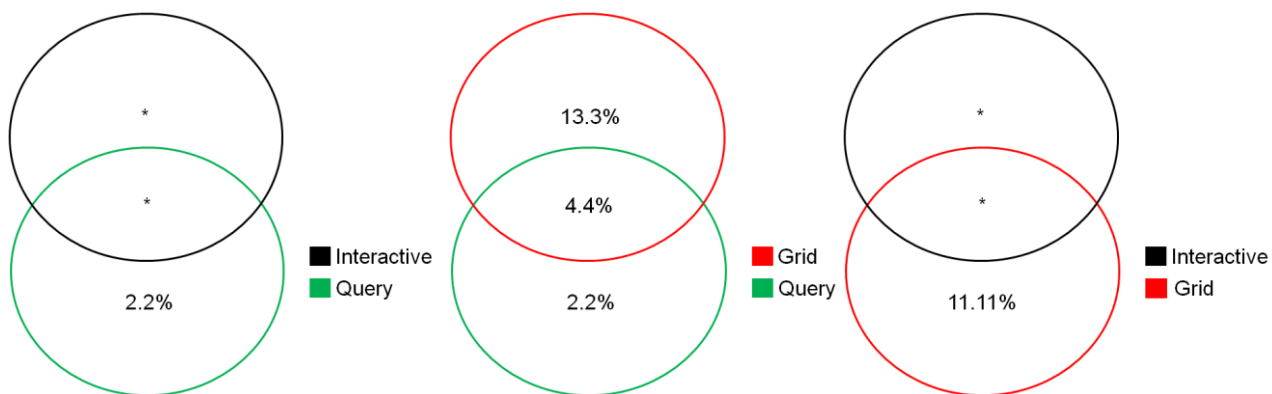


Figure 13: Disagreement Models

The Agree - Disagree model considers both agreement and disagreement statuses. These figures deal with percentage of people who are supporting one Interface and at the same time not supporting the other two. "\*" indicates zero disagreement.

From the above figures, it is discernible that, 2.2 % of Grid interface supporters are not supporting Query Interface and on the other hand, 11.11% of Query Interface supporters are not supporting Grid Interface. The most crucial conclusion that can be drawn is, there exist no entrant who disapproves Interactive Interface but 19.9% of Interactive Interface supporters disagree with the other two interfaces.

Considering all these recommendations, Interactive Interface has been tabbed for advance progression.

## Interactive interface

Interactive Interface with similar features as described in the preliminary survey is developed with slight modifications as shown in the figure below.

Live

Exploratory Search

Available Terms

Be

Experience

Go through

Recorded

Explorative Syzygy results for *experience live*

[Lorem Ipsum is simply dummy text of the printing and typesetting industry](#)  
Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book.  
<http://help@ipsum.com>

[Lorem Ipsum is simply dummy text of the printing and typesetting industry](#)  
Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book.  
<http://help@ipsum.com>

[Lorem Ipsum is simply dummy text of the printing and typesetting industry](#)  
Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book.  
<http://help@ipsum.com>

Explorative Anomaly results for *dead live*

[Lorem Ipsum is simply dummy text of the printing and typesetting industry](#)  
Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book.  
<http://help@ipsum.com>

[Lorem Ipsum is simply dummy text of the printing and typesetting industry](#)  
Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book.  
<http://help@ipsum.com>

[Lorem Ipsum is simply dummy text of the printing and typesetting industry](#)  
Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book.  
<http://help@ipsum.com>

Conventional results for *live*

[Lorem Ipsum is simply dummy text of the printing and typesetting industry](#)  
Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book.  
<http://help@ipsum.com>

[Duis aute irure dolor in reprehenderit in voluptate velit esse](#)  
Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea comm  
<http://help@ipsum.com>

[Sed ut perspiciatis unde omnis iste natus error sit voluptatem accusantium](#)  
doloremque laudantium, totam rem aperiam, eaque ipsa quae ab illo inventore veritatis et quasi architecto ata  
e vitae dicta sunt explicabo. Nemo enim ipsam  
<http://help@ipsum.com>

[quia consequuntur magni dolores eos qui ratione voluptatem sequi nesciunt](#)  
uia non numquam eius modi tempora incidunt ut labore et dolore magnam aliquam quaerat voluptatem.  
Ut enim ad minima veniam, quis nostrum exercitationem ullam corporis suscipit laborio  
<http://help@ipsum.com>

Prev

Page#1

Next

Figure 14: Interactive Interface Modeling

## Features of Interactive Interface

- Results of Syzygy and Anomaly algorithms are displayed independently.
- Every page is laid out with 10 search results. ( 3 Explorative Syzygy results, 3 Explorative Anomaly results and 4 Conventional results)
- Query terms used to fetch the Explorative results are exhibited in textual form above the results.
- Spelling Suggestions pop up only when required.
- User has the feasibility to explore more results from option appended.
- The User has accessibility to include and exclude terms into the Creative Query from the list of Available terms.

Following are the modules, a User can perform on the interface.

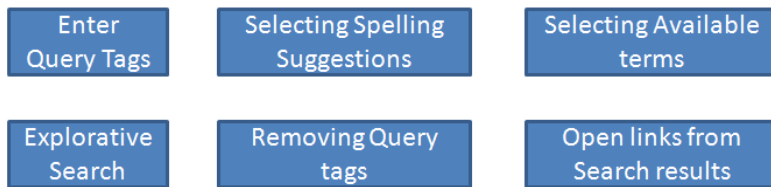


Figure 15:User Modules-I

The following flow charts provide in detailed functionality of each module.

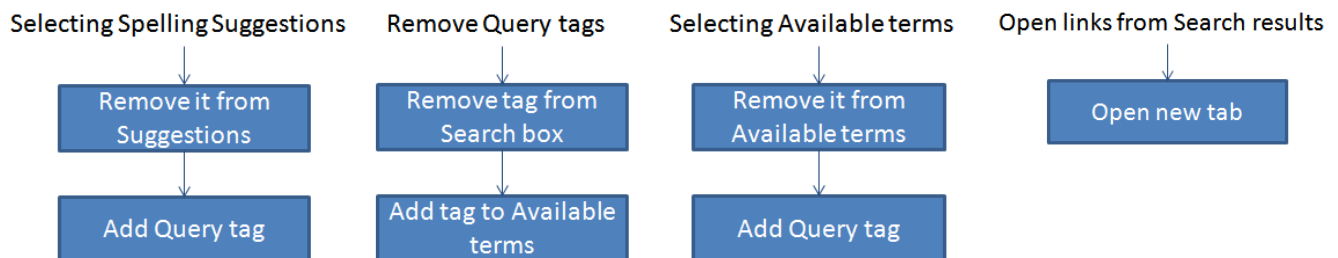
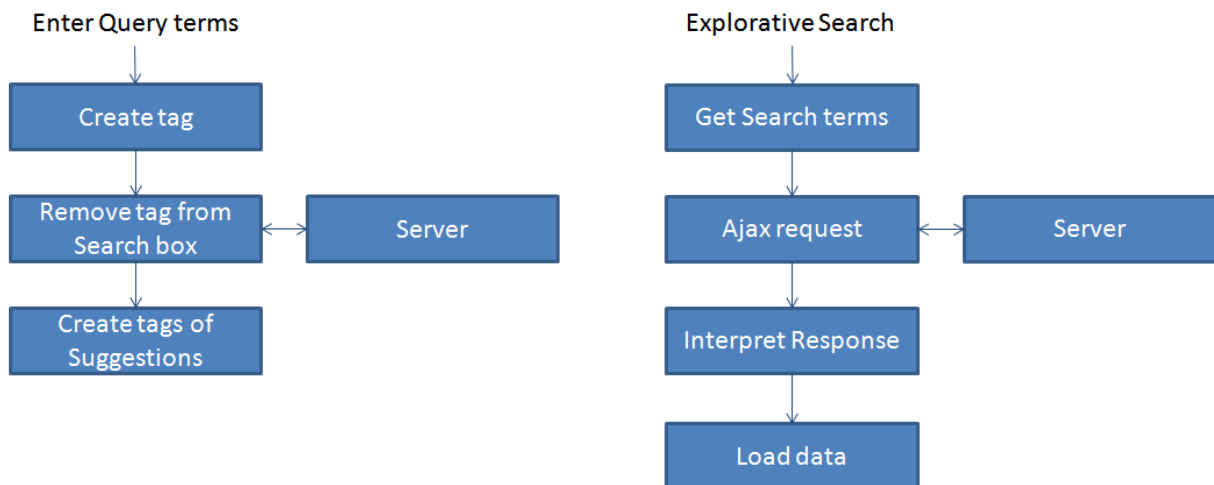


Figure 16:User Modules-II

## EVALUATION

When the interface was completely developed in all respects, a survey is carried out to bring forth the user experience.

In the survey, Users are provided with two different tasks of same complexity and nature. To preserve unbiased quantum, identities of search engines were hidden and also the order in which the results are displayed has been shuffled. The user was directed to envision as a researcher and compare Conventional with Explorative results. The conclusions realized are as follows.

The following tree graph depicts the behavior of the Users before getting familiar with the new Interface

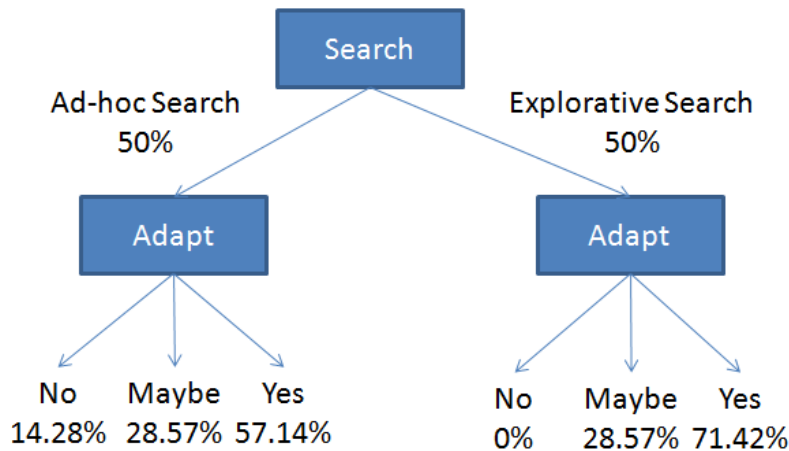


Figure 17: User Behavior-I

It represents the percentage of participants who are psychologically prepared/not-prepared to adopt the new Search Interface. 14.28 % of entries who prefer Ad-hoc Search (those searches where users of search engines exactly know what they want to find and can formulate their search queries easily), are showing resistance to adopt the new interface. These are the users who are not mentally prepared for new interface. In the later stages, we get to know the deviation in the behavior of these users, after getting acquainted with the new interface.

The following tree graph depicts the behavior of Users after getting familiar with the new Interface.

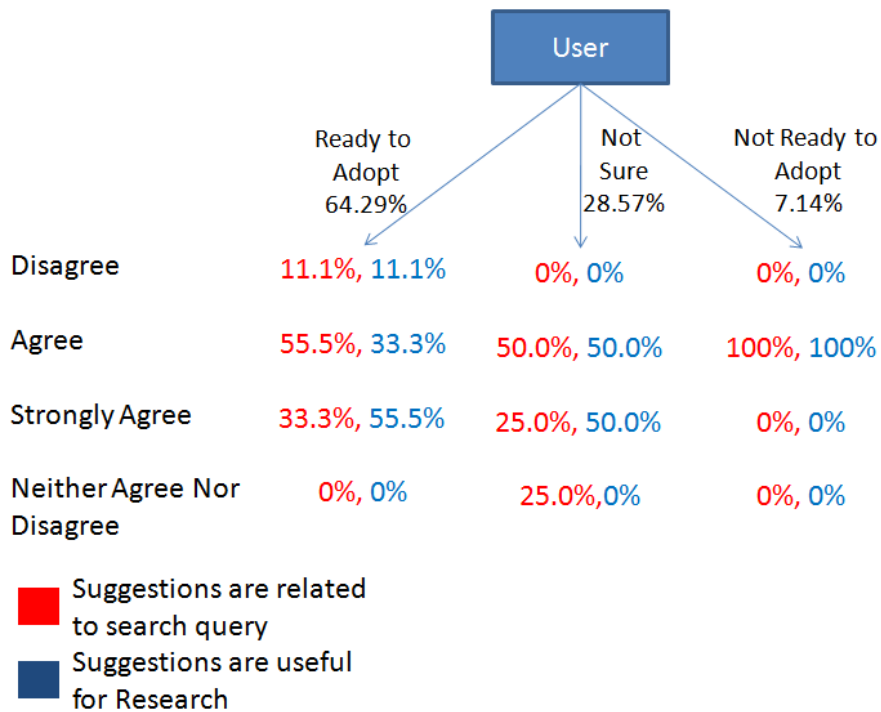


Figure 18: User Behavior-II

From the above figure, it can be noticed that, those passive users who are not ready to adopt new interface before, have given 100% compliance with the conduct of Interactive Interface which indicates that this interface was impressive enough to motivate passive Ad-hoc Search users to adopt new interface. Also there are people who are ready to adopt the Explorative interface, but has disagreed with both of the relevancy and effectiveness of suggestions. But this percentage is being dominated by the percentage of people who cited their conformity.

The following graph correlates Conventional & Explorative results.

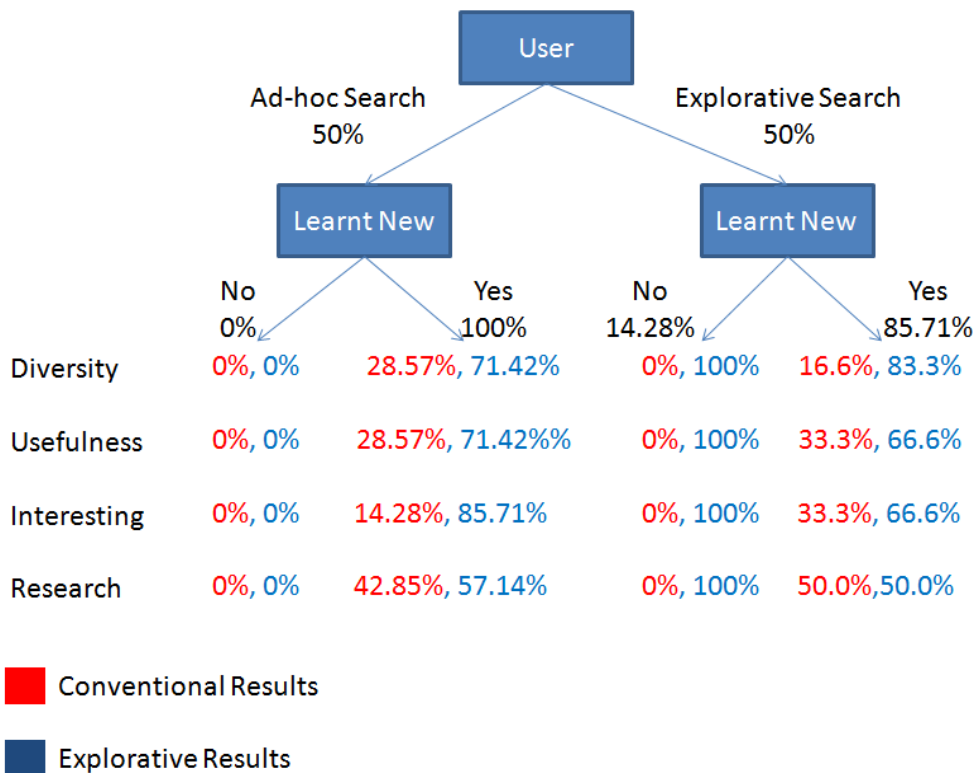


Figure 19: User Behavior-III

In accordance with the above results, it can be assured that Explorative Search is endorsed by many users in all the aspects. Although less than one fourth of the participants who voted for Explorative Search, have affirmed that they haven't learned anything new from it, all of them put up for the same results with respect to other factors (Diversity, Usefulness, ...). Moreover intriguingly, those passive users who previously stepped back from explorative search interface, has also got fascinated by its features and up voted Explorative results.

## CONCLUSION AND FUTURE WORK

In this project, we have successfully developed an Interactive Search Engine Interface which generates effective, diverse, fruitful and compelling Explorative results. But the processing time of the implemented mechanism is still very high (around 30 seconds). In the future work, this limitation could be eliminated by deploying Inverted Indexing process and/or kill with iron (Adding hardware) process. Also the WordNet dictionary used, has a very limited set of words which has to be enhanced.



## REFERENCES

[1] **Exploratory Search**

[https://en.wikipedia.org/wiki/Exploratory\\_search](https://en.wikipedia.org/wiki/Exploratory_search)

[2] **Creative Search Using Pataphysics** by Fania Raczinski, Hongji Yang, Andrew Hugill

[https://scholar.google.co.in/citations?view\\_op=view\\_citation&hl=en&user=sRce9ogAAAAJ&citation\\_for\\_view=sRce9ogAAAAJ:TFP\\_iSt0sucC](https://scholar.google.co.in/citations?view_op=view_citation&hl=en&user=sRce9ogAAAAJ&citation_for_view=sRce9ogAAAAJ:TFP_iSt0sucC)

[3] **The syzygy surfer: (Ab)using the semantic web to inspire creativity** by James Hendler and Andrew Hugill

[https://scholar.google.co.in/citations?view\\_op=view\\_citation&hl=en&user=sRce9ogAAAAJ&citation\\_for\\_view=sRce9ogAAAAJ:e5wmG9Sq2KIC](https://scholar.google.co.in/citations?view_op=view_citation&hl=en&user=sRce9ogAAAAJ&citation_for_view=sRce9ogAAAAJ:e5wmG9Sq2KIC)