# 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 Munche.

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 words. 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

syzygy( 𝑡 ) = { ℎ ∶ ℎ ∈ union( 𝑡 ) ∧ ∃ ℎ ∈ 𝑉 }

union( 𝑡 ) = hypo( 𝑡 ) ∪ hyper( 𝑡 ) ∪ holo( 𝑡 )

hypo( 𝑡 ) = { ℎ ∶ ℎ ∈ hyponyms( 𝑠 ) }

hyper( 𝑡 ) = { ℎ ∶ ℎ ∈ hypernyms( 𝑠 ) }

holo( 𝑡 ) = { ℎ ∶ ℎ ∈ holonyms( 𝑠 ) }

syno( 𝑡 ) = { 𝑠 ∶ 𝑠 ∈ synonyms( 𝑡 ) }

for 𝑠 ∈ syno( 𝑡 )

For e.g. Let t = { live } ;

syno( live ) = { populate, inhabit, **be**, **domicile** };

hypo(populate, inhabit, be, domicile) = { cliff, dwelling, house ... };

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

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

union(hypo, holo, hyper) = {populate, **be**, fill up, house, exist, **domicile**, camp,.... }

**->**Syzygy( live ) = { **be**, **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

antinomy( 𝑡 ) = { ℎ ∶ ℎ ∈ anto( 𝑡 ) ∧ ∃ ℎ ∈ 𝑉 }

anto( 𝑡 ) = { ℎ ∶ ℎ ∈ antonyms( 𝑠 ) }

syno( 𝑡 ) = { 𝑠 ∶ 𝑠 ∈ synonyms( 𝑡 ) }

for 𝑠 ∈ syno( 𝑡 )

For e.g. antonym( live ) = { **dead**, recorded }

synonym( live ) = { alive, animate, breathing}

antonym( alive ) = {**dead**}

antonym( animate ) = { **dead**, inactive, inhibit..}

antonym( breathing ) = { breathless, **dead** }

anomaly( live ) = {**dead**}

## Clinamen:

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

For a search term t

Clinamen ( t ) ={ 𝑣 ∶ 0 < dameraulevenshtein (t, 𝑣 ) ≤ 2 }, for 𝑣 ∈ 𝑉

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

## E\_ ALGORITHMS

Basic algorithms of Syzygy and Anomaly with slight modifications are implemented. Clinamen is completely ignored due to it's 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

In the original Syzygy, union of hypernyms, holonyms and hyponyms are considered as a resultant terms.

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.

For a search term t

syzygy( 𝑡 ) = { ℎ ∶ ℎ ∈ intersection( 𝑡 ) }

intersection( 𝑡 ) = hypo( 𝑡 ) ∧ hyper( 𝑡 ) ∧ holo( 𝑡 )

hypo( 𝑡 ) = { ℎ ∶ ℎ ∈ hyponyms( 𝑠 ) }

hyper( 𝑡 ) = { ℎ ∶ ℎ ∈ hypernyms( 𝑠 ) }

holo( 𝑡 ) = { ℎ ∶ ℎ ∈ holonyms( 𝑠 ) }

syno( 𝑡 ) = { 𝑠 ∶ 𝑠 ∈ synonyms( 𝑡 ) }

for 𝑠 ∈ syno( 𝑡 )

For e.g. Let t = live

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

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

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

hypo(populate, inhabit, be, domicile) = { domicile, reside, camp, tent, nest,.... }

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

### 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.

For a search term t

anomaly( 𝑡 ) = { ℎ ∶ ℎ ∈ anto( 𝑡 ) }

anto( 𝑡 ) = { ℎ ∶ ℎ ∈ antonyms( 𝑠 ) }

syno( 𝑡 ) = { 𝑠 ∶ 𝑠 ∈ synonyms( 𝑡 ) }

for 𝑠 ∈ syno( 𝑡 )

For e.g. t = live

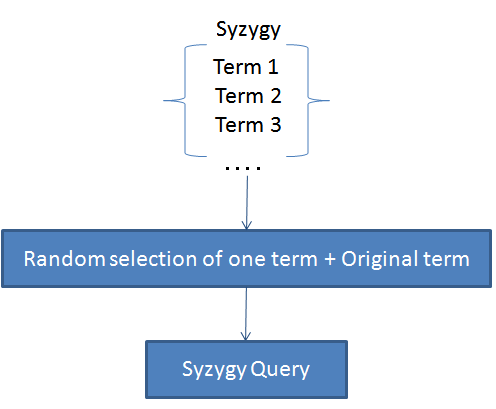
synonym( live ) = { alive, animate, breathing}

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

# EXPLORATIVE QUERY GENERATION

## Syzygy Query Generation:

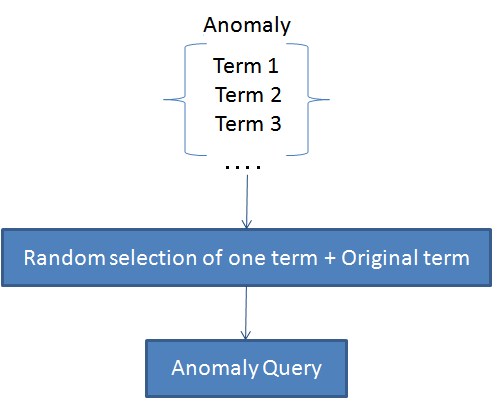
From the resultant set of Syzygy terms, only one word is selected randomly and combined with the original query term to form Explorative Syzygy Query.



*Figure 1: Syzygy Query Generation*

## Anomaly Query Generation:

From the resultant set of Anomaly terms, only one word is selected randomly and combined with the original query term to form Explorative Anomaly Query.

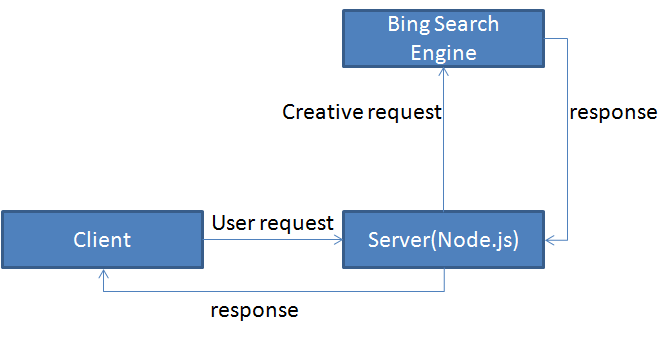


*Figure 2: Anomaly Query Generation*

# BASIC ARCHITECTURE

Client, Server and Bing Search Engine are part of basic architecture. In this project, design and development of Client and Server is done.

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.

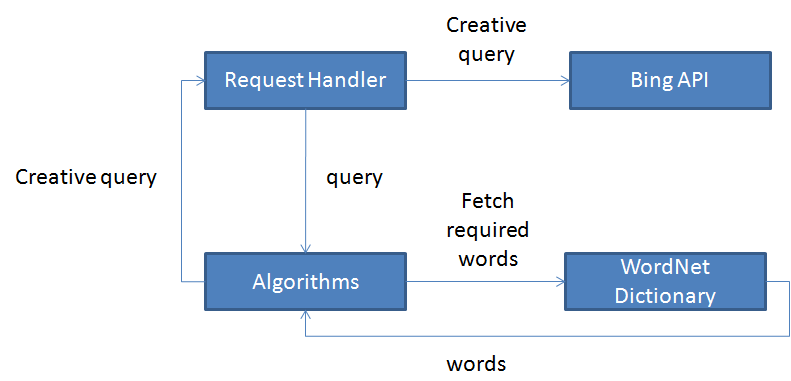
2. Server:

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.

## Internal Architecture of Server

In Server side implementation, there are four modules. They are Request Handler, Algorithms, WordNet API and Bing API, which are performing their own functionality to handle client request(see client requests). For Server side scripting Node.js is used.

*Figure 4: Functional Diagram of Server*

Node.js is highly scalable, instead of using multi-threading(limited number of threads) functionality its using asynchronous function calls so there is no limit for calling asynchronous functions.

After receiving request from user, Request handler 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).

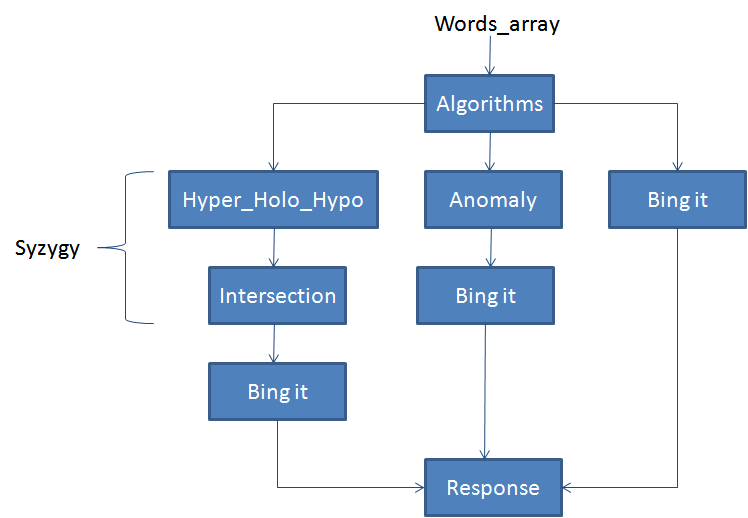
The evolved creative query is then used to fetch results from Bing API which are directed as the response to the client.

# REQUEST HANDLER

It contains the implementation regarding client request handling. The **diagram** below depicts the main functionality of the Server.

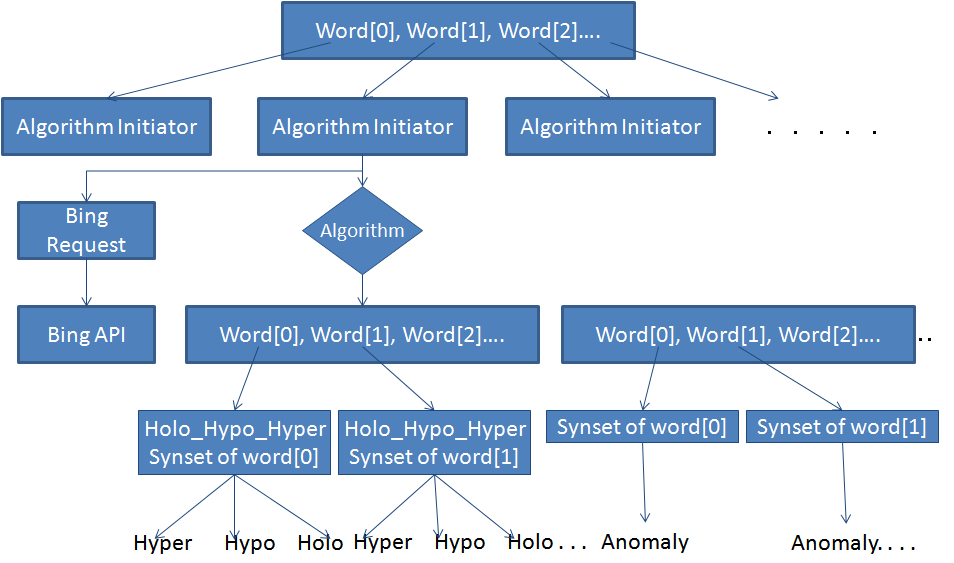
|  |  |
| --- | --- |
| **File Reader**: It is responsible for returning HTML, CSS, Java Script and image files.  **Spelling Suggestion**: It handles the request for Spelling Suggestions from Bing API for Client.  **Search Request**: It handles User requests for Search queries.  **More Results**: It handles the request to explore more Explorative results.  **Response**: This module is responsible for generating response to the Client. | 3.png  *Figure 5: Main modules of Server* |

## Search Request Handler:

All the Search 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 have been implemented as shown in the below figure 7.

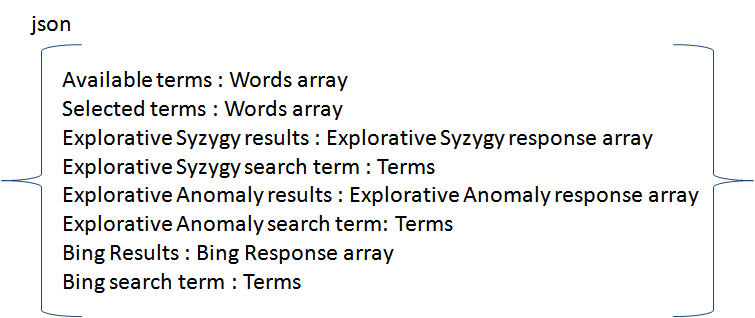


*Figure 7: Multiple levels of parallelism*

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:**

The final response generated by the server in the json format is as shown in the figure 8.



*Figure 8: Server Response*

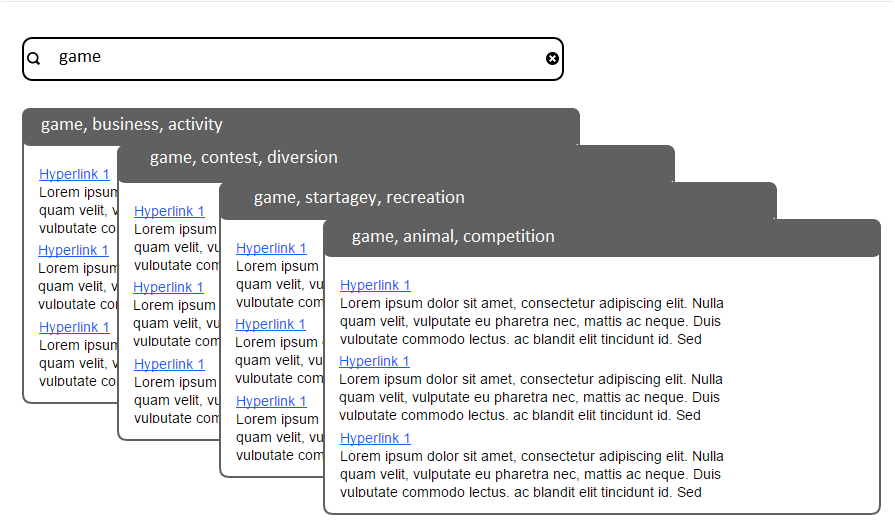
# USER INTERFACE DEVELOPMENT

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:**

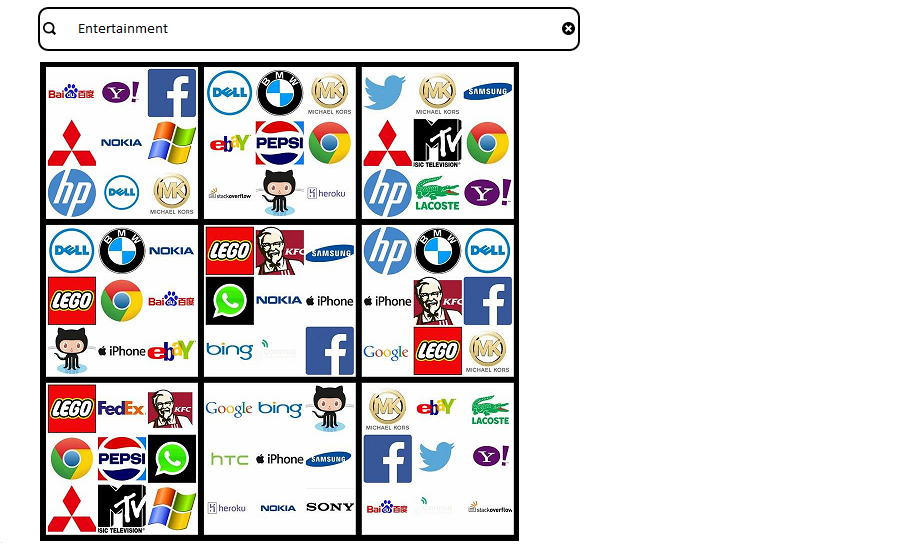
It is one of the interface in which a user is exposed to different web pages overlapping on each other with visible query terms from which the results are generated. It is the user choice to select the suitable window, based on the visible query terms displayed on the window header.



*Figure 9: Query Display Interface*

### Grid Display Interface

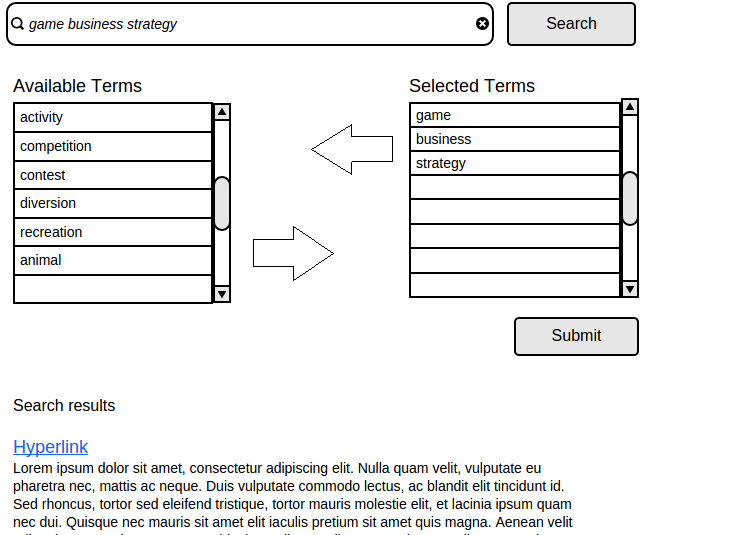
In this interface, images of website logos related to different creative query term sets are displayed in the different grids as shown below. User has the feasibility to choose one amongst them.



*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.



*Figure 11: Interactive Interface*

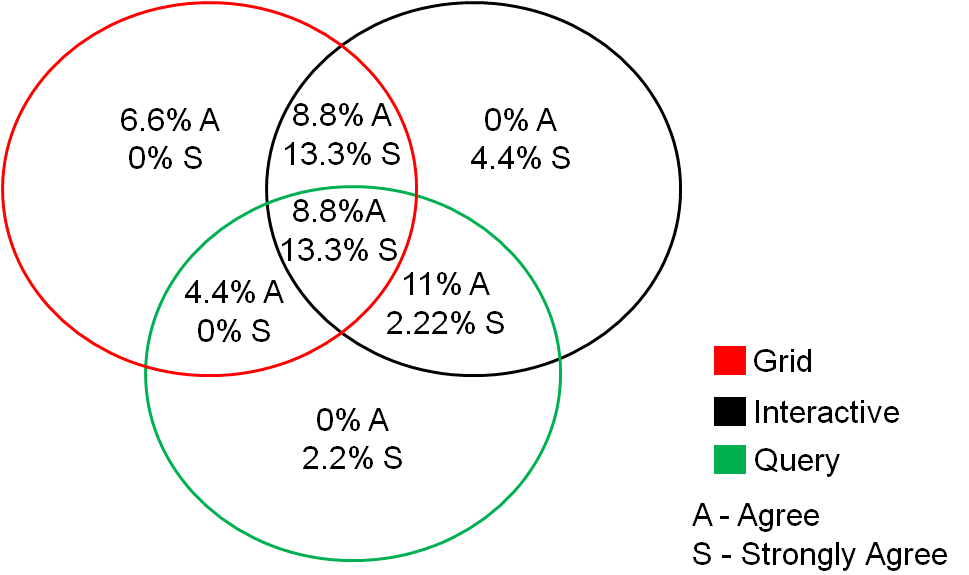
**Preliminary Evaluation**

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.

**AGREEMENT MODEL**

In Agreement model, only those users, who showed agreement for at least one interface are considered. Phrasing the response options, scale requires the respondent to first decide how supportive the interface was and then whether they "strongly agree", "agree", "neither agree nor disagree", "disagree" or "strongly disagree", with the specific interface.

****

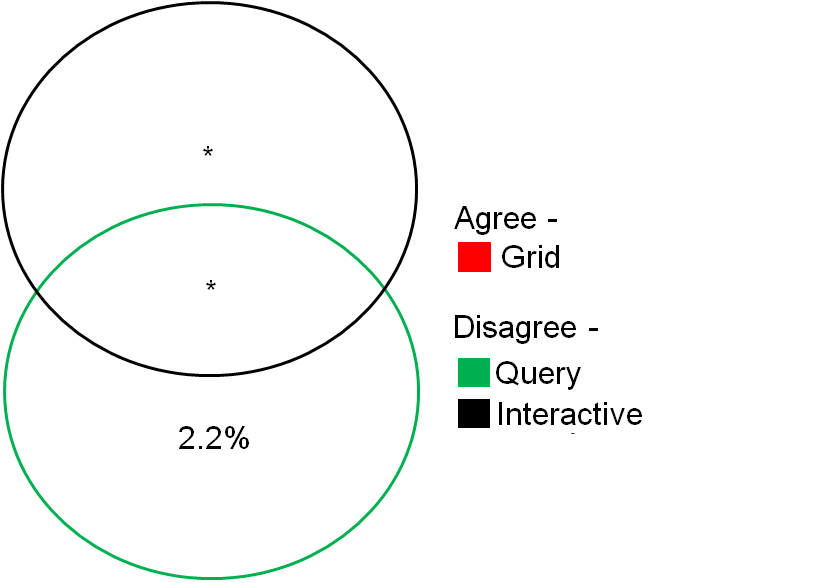
*Figure 12: Agreement Model*

In figure 12, 4.4%S in Interactive interface are those users who strongly agreed with this interface and have neither agreed nor disagreed with both Grid and Query interfaces. 11%A, in the intersecting part of Interactive and Query interfaces, are those users who have agreed with Interactive and Query interfaces but have neither agreed nor disagreed with Grid interface. 13.3%S, in the intersecting section of all the three interfaces, is the percentage of those users who have strongly agreed with all the three interfaces.

It is clear that, 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 alone.

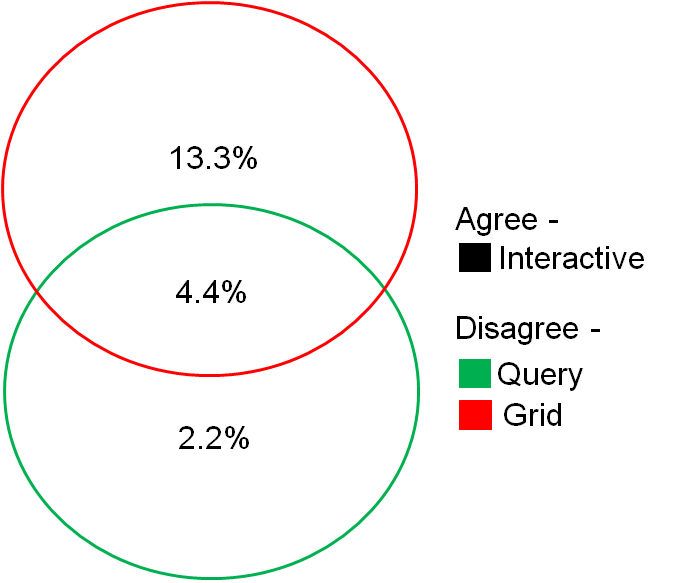
**AGREE - DISAGREE MODEL**

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



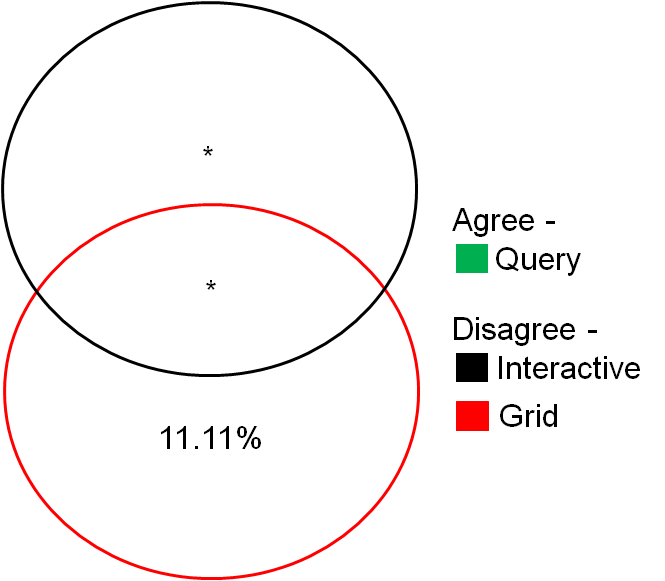
*Figure 13a: Grid Agreement Model*

Figure 13a shows that, out of all the users who agreed with Grid interface, 2.2% of them has disagreed with Query Interface and none of them disagreed with interactive interface. There are no users who disagreed both the interfaces together.



*Figure 13b: Interactive Agreement Model*

Figure 13b shows that, out of all the users who agreed with Interactive Interface, 2.2% of them has disagreed with Query Interface and 13.3% of them disagreed with Grid Interface. There are 4.4% users who disagreed both the interfaces together.



*Figure 13c:Query Agreement Models*

Figure 13c shows that, out of all the users who agreed with Query Interface, 11.11 % of them has disagreed with Grid Interface and none of them disagreed with Interactive Interface. There are no users who disagreed both the interfaces together.

From the above figures, as disagreement percentage of Interactive Interface is 0%, Grid Interface is 24.41%(11.11%+13.3%) and Query Interface is 4.4%(2.2%+2.2%). On the whole Interactive Interface is dominating the other two interfaces in this model as well.

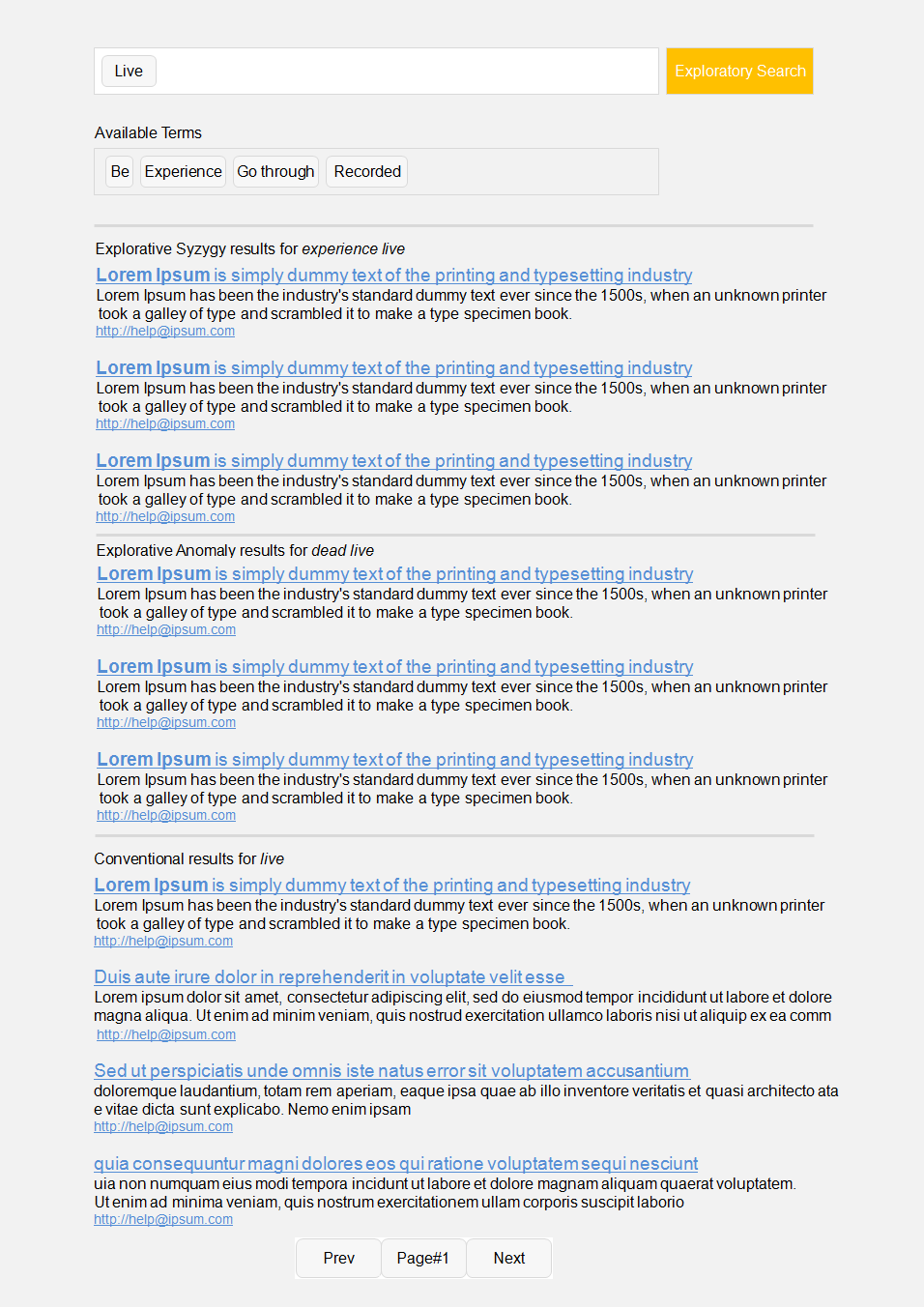
**PARTICIPANT'S COMMENTS:**

Above all, a participant comment in the survey, "In general, I think the displays of the first two interfaces(Grid and Query Interfaces) are rather overwhelming, especially the grid display. I think the third display(Interactive Interface) is a decent idea, but generally seems like too much effort for a normal search query. ", suggests that users feel more comfortable with Interactive Interface as compared with other interfaces.

Hence it can be concluded that, from both the models and participants comments, users got highly fascinated by Interactive Interface and thus it 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.



*Figure 14:Interactive Interface Modeling*

**Features of Interactive Interface**

* Explorative Syzygy , Explorative Anomaly and Conventional search results 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 displayed in textual form above the results.
* Spelling Suggestions pop up only when there exists a suggestion from Bing.
* User has the feasibility to explore more results.
* The User has accessibility to include and exclude terms into the Creative Query from the list of Available terms and/or spelling suggestions.

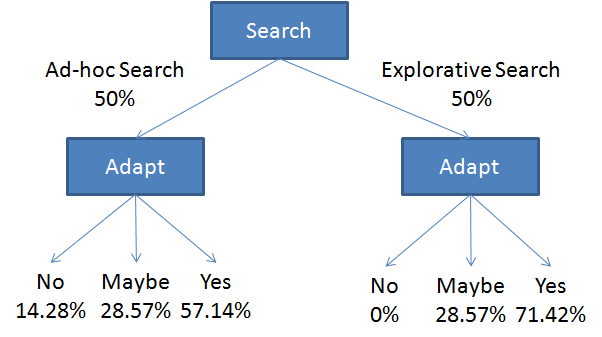
# EVALUATION

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 for the two tasks. 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.



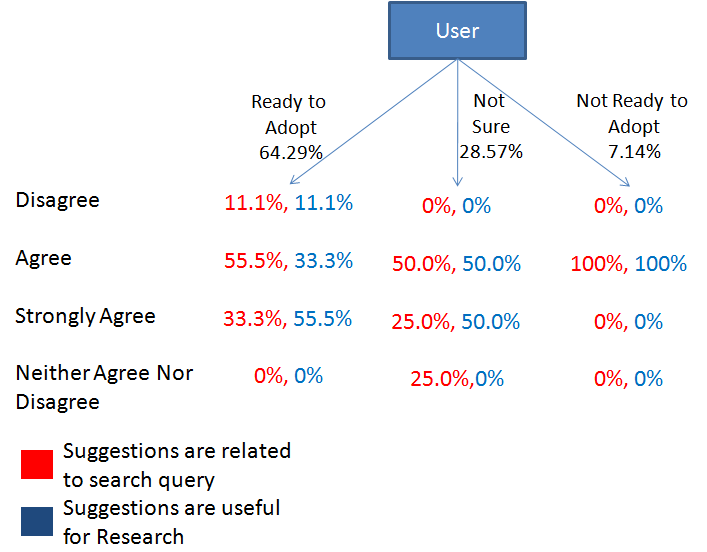
*Figure 17:User Behavior-I*

Participants are queried about the type of search they are interested in. While, 50% of them showed up on Ad-hoc Search( those searches where users of search engines exactly know what they want to find and can formulate their search queries easily), the remaining 50% showed up on Explorative Search.

Later when they are asked whether they would like to adapt new interface, 71.42% of participants who supported for Explorative Search, confirmed with yes and 28.57% were not sure but none of them refused to adapt.

In case of users who supported Ad-hoc Search, 57.14% of them confirmed with yes, 28.57% were not sure and 14.28% showed their resistance to adapt the new interface. These 14.28% of the users are those, who were not yet prepared for the 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.

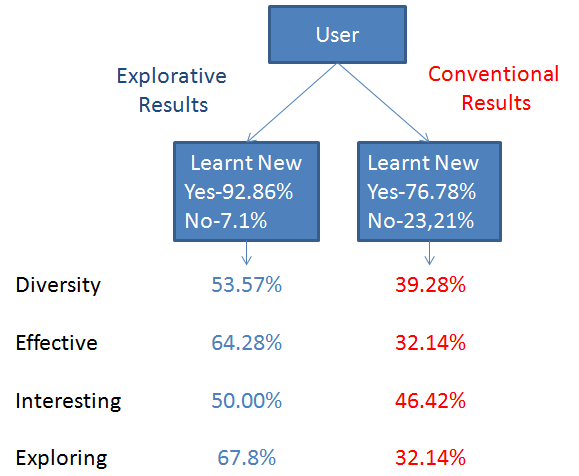


*Figure 18:User Behavior-II*

In the Interactive Interface, users have feasibility to include suggested terms into the Creative query. In the Survey, Users were enquired about whether these suggested terms are related to the original search query and useful for research purposes. From the above figure, it can be noticed that, those passive users(7.14%) who were not ready to adopt new interface before(see figure), have given 100% compliance with the conduct of Interactive Interface, which indicates that this interface was impressive enough to motivate even passive Search users to adopt new interface.

Among the users who are ready to adopt the new interface,11.1% of them has disagreed with both of the relevancy and effectiveness of suggestions. But this percentage(11.1%) is being dominated by the percentage(33.3%+55.5%) of users who cited their conformity.

The following graph correlates Conventional & Explorative results. The percentage of people, who learnt new things from Conventional results and Explorative results, are derived and these two search results are compared with respect to Diversity, Effectiveness, Interesting and Exploring factors.



*Figure 19:User Behavior-III*

From the above figure 19, clearly we can see dominance(92.86% over 76.78%) of Explorative results in terms of learning something new.

When the two search results are compared, with respect to all the 4 factors(Diversity, Effectiveness, Interesting and Exploring) mentioned in the figure 19, Explorative results showed their dominance over Conventional 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 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 later in the future.

# REFERENCES

[1] White, R. and Roth, R: " *Exploratory Search: Beyond the Query-Response Paradigm*" Morgan and Claypool, San Rafael, CA(2009).

[2] Fania Raczinski, Hongji Yang, Andrew Hugill: "*Creative Search Using Pataphysics*", Proceedings of the 9th ACM Conference on Creativity & Cognition, pp. 274-280(2013).

[3] James Hendler, Andrew Hugill: "*The syzygy surfer: (Ab)using the semantic web to inspire creativity*", Int. J. of Creative Computing, Vol.1, No.1, pp.20 - 34(2013).