Websites:
http://www.jofamericanscience.org
http://www.sciencepub.net

Emails:
editor@sciencepub.net
sciencepub@gmail.com

Journal of American Science

MARSLAND PRESS
Multidisciplinary Academic Journal Publisher

# THE INTERACTIVE INTENT ESTIMATION MODEL

[1]Dr. Rajeev Yadav and [2]Vivek Kumar

[1]Professor, Department of Computer Science, SunRise University, Alwar, Rajasthan (India)
[2]Research Scholar, Department of Computer Science, SunRise University, Alwar, Rajasthan (India)
Email-Viveks865@gmail.com

***Abstract:*** The proposed system, adapts the significance of the new context-based approach lies in the improved relevance of search results even for users not skilled in Web search. We achieve this by applying natural language processing techniques to the captured context in order to guide the subsequent search for user-selected text. Existing approaches, either analyze the entire document the user is working on, or ask the user to supply a category restriction along with search keywords. As opposed to these, our proposed method automatically analyzes the context in the immediate vicinity of the focus text. This allows analyzing just the right amount of background information, without running over the more distant (and less related) topics in the source document. The method also allows collecting contextual information without conducting an explicit dialog with the user.
[Yadav, R. and Kumar, V. **THE INTERACTIVE INTENT ESTIMATION MODEL.** *J Am Sci* 2022;18(6):30-41]. ISSN 1545-1003 (print); ISSN 2375-7264 (online). http://www.jofamericanscience.org. 5. doi:10.7537/marsjas180622.05.

**Keywords:** Interactive, Intent, Model

## Introduction:

Given the constantly increasing information overflow of the digital age, the importance of information retrieval has become critical. Web search is one of the most challenging problems of the Internet today, striving to provide users with search results most relevant to their information needs. Internet search engines have evolved through several generations since their inception in 1994, progressing from simple keyword matching to techniques such as link analysis and relevance feedback (achieved through refinement questions or accumulated. A large number of recently proposed search enhancement tools have utilized the notion of context, making it one of the most abused terms in the field, referring to a diverse range of ideas from domain-specific search engines to personalization [49].

This thesis presents a novel intent estimation strategy that interprets context in its most natural setting, namely, a body of words surrounding a user-selected phrase. The anticipation is that the growing number of searches that originate while users are reading documents on their computers, and requires further information about a particular word or phrase. Hence, the basic premise underlying our approach is that searches should be processed in the context of the information surrounding them, allowing more accurate search results that better reflect the user' s actual intensions. The implicit co-relations within the three search contexts are adapted for the intent estimation and navigation of the futuristic search.

The focus of the work carried out in this chapter is to advocate the feasibility of the co-relations and effect on several performance indicators, via following research questions (RQs): (RQ-7): What constitutes the co-relations among three search contexts in inherent exploitation-explorations? Which context component is most significant? (RQ-8): How can co-relations be adapted to design the overall system and extensive support for the user search-interactions? (RQ-9): Finally, how to design an adaptive visualization that, assist the user search task and illustrate the intent evolutions.

### 1.1 The role of Pro-search Context in Exploratory Information-seeking

A ' context' can be defined as a description of aspects of a situation and an internal representation in the cognitive state of knowledge. In ideal information system, context technology mechanism captures the concepts and relation and co-relation among in different user contexts, which is easy to reuse across searches/domains. Context information can be used to facilitate the communication in human-computer interaction [49]. The use of context is becoming important in interactive computing. Recently, there has been much discussion about the meaning and definition of context and context -awareness. However, this kind of information (context) is still not utilized much and the concept of context is not yet well understood or defined. Additionally, there exists no commonly accepted system that supports the acquisition, manipulation and exploitation of context including information units and data [49].

When discussing the information retrieval process, often the focus is on the individual activities such as formulating queries, searching document collections and presenting returned documents. However, there are situations where we need to go beyond analyzing these individual activities in isolation, and consider the groups of these activities. Traditionally, nearly 60% of users had conducted more than one information retrieval (IR) search for the same information problem. In their research, they refer to the process of repeatedly searching over time in relation to a specific but possibly evolving information problem as the successive search phenomenon.

Contextual information plays a more important role in the study of successive searches than that of isolated searches since the contexts behind a series of successive searches are probably closely related to each other, if not the same. However, finding contextual information is a complex, even for successive searches. Previous studies have demonstrated that less information is available about the users and their information needs, not to mention the fact that searches are shorter and search statements contain fewer terms than their counter parts in traditional IR searches. An individual retrieval task may be informative sometimes, but a collection of search activities provides much more information about the topic and the context if they are organized according to their time order and related search topic. It is likely that consecutive activities related to one topic can share the same context. It is, therefore, reasonable to say that the information about search topics is an important component of the context behind the users' searches or retrieval need.

For example, a search for the word "Jaguar" should return car-related information if performed from a document on the motoring industry, and should return animal-related information if performed from an Internet website about endangered wildlife. Guiding user's search by the context surrounding the text eliminates possible semantic ambiguity and vagueness.

Keyword-based search engines are in widespread used today as a popular means for Web-based information retrieval [51, 110]. Although such systems seem deceptively simple, a considerable amount of skill is required in order to satisfy non-trivial information needs. The thesis presents a new conceptual paradigm for performing search in context that largely automates the search process, providing even non-professional users with highly relevant results. This paradigm is implemented in practice in the proposed system 'IIM', where search is initiated from a text query marked by the user in a document she views, and is guided by the text surrounding the marked query in that document ("the context"). The context-driven information retrieval process involves semantic keyword extraction and clustering to automatically generate new, augmented queries. The latter are submitted to a host of general and domain-specific search engines. Search results are then semantically re-ranked, using context. The experimental results testify that using context.

In the thesis, an interactive intent-model based exploratory search system is designed and referred to as 'IIM'. Here, a client application running on the user's computer captures the context around the text highlighted by the user. The server-based algorithms analyze the context, via selecting the most important document and eventually, keywords/terms. The 'IIM' system assists the user to modify the intent to which context guides any given search, by modifying the amount of context considered. The context can be reliably classified to a predefined set of search states. A dedicated re-ranking module ultimately reorders the results received from all of the engines, according to semantic proximity between their summaries and the original context. Systems as an information specialist acting on behalf of the user, which automatically performs the search steps, from query expansion, to search engine selection, to re-ranking the results.

**1.2 Intent Estimation for Exploratory Information-seeking**

The 'Intent' is a topical dimension of user search and characterizes 'why' the user is searching, and 'how' his search evolves during search progression [76]. Characteristically, intent defined as 'immediate reason, purpose, or goal' that motivates a user to initiate or conduct a search [88] and co-exist in three aspects, i.e. Pre-search, In-search and Pro-search of the user search context. A significant fraction of user searches is influenced by the user's primary search aim i.e. 'Pre-search' context and others due to intermediate query or result in understanding

'In-search' context. An ideal information system would be able to predict the estimation of future intent (also known as 'Pro-search' context) based on the captured 'Pre-search' and 'In-search' context. The prediction of search context of futuristic search 'Pro-search' requires identification of co-relations between all three aspects of user search, therefore understanding 'why' users start searches and 'how' to predict search intent are multifaceted tasks [33, 134]. The primary focus of Information system (IS) systems has been to optimize the user-centric information retrieval and supplement the interaction related support. Conventional retrieval strategies are primarily based on term statistics, e.g. term-frequencies, inverse-document frequencies, document lengths, for the retrieval and subsequent ranking of the query results [23,30, 101]. Intuitively, the proximities instances of query terms within matched results or document can also be utilized, and proximity score could be amalgamated with traditional document-term based score in retrieval framework.

### 1.3 Proposed Retrieval Framework and Intent Estimation Model

The proposed strategy assigns more weights on exploration aspects, during initial iterations of data retrieval, and adapts to the best-efforts matching, rather exact matching. The focal point of retrieval shifts towards exploitation of related data objects, in later stages, eventually to extract highly related objects. Each intermediate user search interactions are utilized to solve the exploration/exploitation tradeoffs, and incorporated into search intent estimate. Figure 1.1 shows the evolution in the exploitation-exploration circle for a user query with retrieved objects.

In initial algorithmic iterations, the focuses is on exploration rather exploitation, as a user is uncertain of real search needs thus retrieval on the best-efforts basis is prompt. Later, the exploration circle is spanned in larger than exploitation circle, to indicate the user's state of knowledge on the current search, and this growth indicates the enhanced state of knowledge of the users' search needs, and eventually end up to a narrowed exploration circle, which implies the lesser space of uncertainty. Though, data objects under exploration circle are potential to navigate the search towards new search interest.

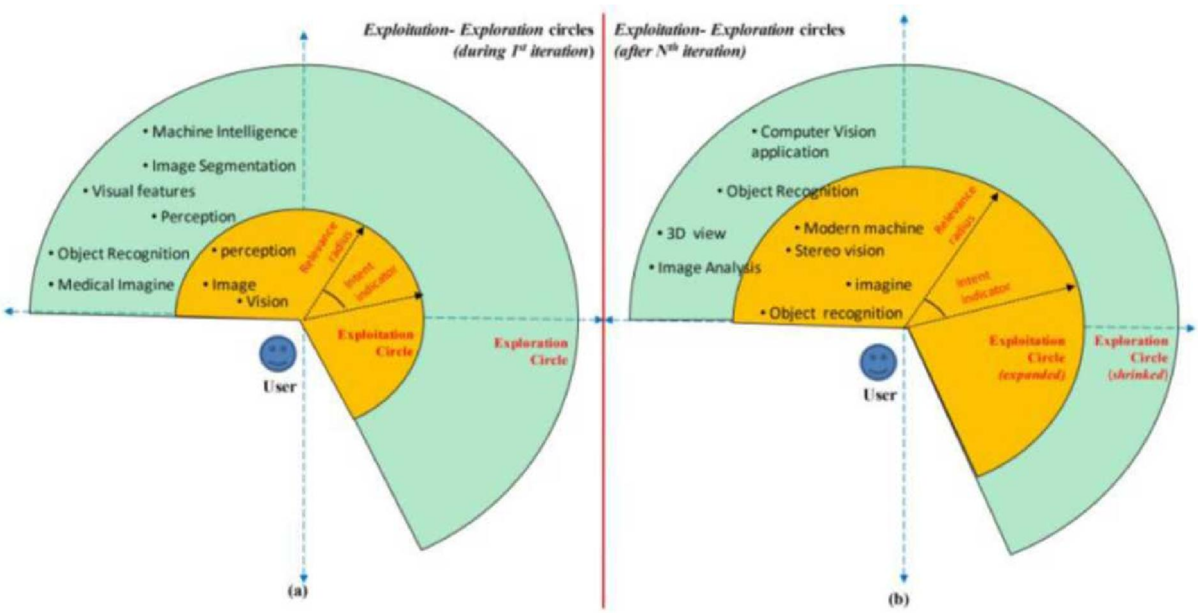


**Figure 1.1.Exploitation-Exploration evolution (a) during initial iterations, Exploitation-circle (inner circle colored in yellow) is smaller to depicts User's knowledge-state on results, and large Exploration-circle (outer circle colored in blue) to display potential results for futuristic iterations (b) in later phase of search-sessions, Exploitation-circle expands to suggest growth in knowledge state, and reduced Exploration-circle indicates reduced uncertainty over search at hand.**

For example, a user search for 'computer vision' related scientific documents, system visualize the extracted keywords from the relevant documents, and imposed relevance feedback on two terms 'perceptions' and 'object recognition'. Figure 1.1(b) displays evolve search results based on estimated search intents.

The actual user interface of the proposed system is shown in Figure 4.5, here, in addition, a classic query box and search keyword view, the interface visualizes the intent model. In the inner area, keywords close to the centre present current search intent and the outer area consists of keywords that are recommended as potential future intents. The framework thus offers the user to identify potentially interesting intents to navigate the search. The position of an individual keyword in the layout is defined by the current search intent estimate (on Relevance radius, Intent angle). The radius of a keyword signifies its relevance in the current estimate, thus the closer a keyword is to the centre, the more relevant. The angles of keywords indicate similarity 'In-search' context: two keywords with similar angles indicate the similar intents.

The components of proposed framework are shown in Figure 1.2. The user search begins by issuing simple data query (the keyword query are preferred) over the designed user-interface (UI). The user query is pre-processed/expanded for the extraction of initial data objects, the retrieval of initial results are based on algorithmic relevance (based on best-effort matching).

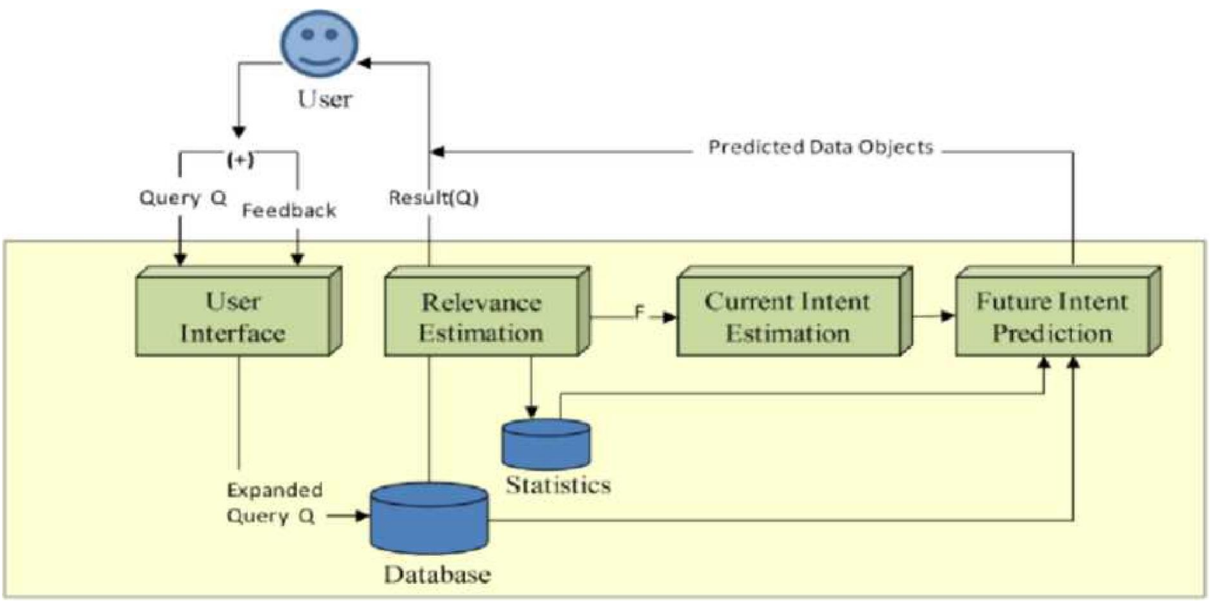


**Figure 1.2 Components and interactions of Proposed Intent Estimation Model.**

The 'relevance estimation' module of the proposed system prepares the relevance or matching between query term and pre-processed information objects (O1,O2....On). The relevance estimation at this stage is inherently based on traditional document-term (DT) statistics and query term (QT) statistics, the combination of both score is employed to retrieve the data objects for initial search interest, referred as 'Pre-search' context. The data results (Result (Q)) extracted based on 'Pre-search' context are visualized over the interface, and an active user reviews the results and co-relate with initial results. Generally, the user revises his search interest post the revision of the initial result, and an ideal exploratory search system offers opportunity for the applying the evolution in the search-thought (interest), often referred as 'In-search' context.

The user relevance-feedback (URF) is imposed in the system via labeling the result objects with positive or negative intent. The award or penalty is incorporated into the system as URF and supplied to 'Current Intent Estimation' module. Similarly, pseudo relevance feedback (PRF) identifies top-k result objects from each window in top-n and retains them in next/future iterations. This module estimates the relevance estimate for the 'In-search' context, and also evaluates the implicit co-relations with 'Pre-search' context. The co-relations are employed to predict the futuristic data objects.

The final component is responsible for the estimation of future search intent, referred as 'Pro-search' context, via a simple strategy of co-relations between the initial relevance estimate 'Pre-search' and revised intent estimate 'In-search' context. The components simply construct the overall estimate as new updated intent estimate and project the revised estimate and identify the relevant data objects. Inherently, each user-interaction significantly affects the revision in the currently estimated intents, as QTP and URF both get updated due to explicit interactions for the query reformulation and relevance-feedback respectively. The changes in the 'Pro-search' significantly affect the quantity and quality of predicted or recommended results.

In this work, interactive intent estimation model is proposed, to proactively estimate the future information needs. Three different search contexts, 'Pre-search', 'In-search' and 'Pro-Search' are adapted into intent estimation, and characterize three-different aspect of user-preferences on search task [92]. Pre-search context characterizes the initial search interest of the user and captured via algorithmic relevance measures, e.g. document-term statistics, query-term proximity statistics. Further, URF on result objects based current cognitive perception of search conceptualized as In-search context. At last, the captured top-k or most relevant retrieved results often referred as PRF are utilized to proactive search estimation, thus referred as Pro-search, to illustrate potential result objects for future searches. Each search context characterizes different user search preferences during a user information-seeking.

### 1.3.1 Workflow of Proposed Exploratory Search System

The role user search-interactions in exploratory information-seeking are significant, as each interaction steers the initiated search towards the area of interest iterative. The inherent data exploration is driven by the relevance manifestation, thus in the proposed system; a context-aware intent estimation framework is adapted to capture search evolution precisely. The component 'Ranking Model' encapsulates the various activities such as, initial relevance estimation to estimate update, feedback modeling, re-ranking of the result, and eventually the retrieval of data objects during the search task. The workflow of the proposed exploratory search system (ESS) is shown in Figure 1.3.

A user submits data request 'Query', via the input section, a typical user query contains generic keyword or terms. The proposed retrieval framework prepares a 'Pre-search' estimate (DT and QT statistics) and concurrently manages the entered user query in 'History Log'. The extracted result objects are presented to the user in three different views: a detailed view on 'Result Section' and a summarized view on 'keyword view over radar view. The user can impose his relevance feedback on a result objects over result section (via a simple click on +/-) and to rewrite current query by drag relevant keywords from radar view. The intent radar view is adapted to visualize the clustered representation of extracted keywords/terms from the retrieved documents based on 'Pre-search' context. The relevance feedbacks (URF and PRF) are supplied to 'Ranking model' of the system, to update the current estimate and further extracts the result objects for revised search intent 'In-search' context of the user search.

Rewrite current query by drag relevant keywords from radar view. The intent radar view is adapted to visualize the clustered representation of extracted keywords/terms from the retrieved documents based on 'Pre-search' context. The relevance feedbacks (URF and PRF) are supplied to 'Ranking model' of the system, to update the current estimate and further extracts the result objects for revised search intent 'In-search' context of the user search.

The proposed system offers additional result objects potentially relevant to the user-initiated search and incorporated intermediate feedbacks, via the concurrent correlation between the 'Pre-search' estimation and 'In-search' context. In this, both relevance feedbacks (URF and PRF) plays a pivotal role, as PRF promotes top-k result from top-n into the 'Exploration Intent Circle' and URF predict results object for 'Exploitation Intent circle' (as demonstrated in Figure 1.1 for user search on 'Computer Vision'). The user search and interactions are conducted over the generic keywords, either extracted from the document corpus or provided by the user during the search.

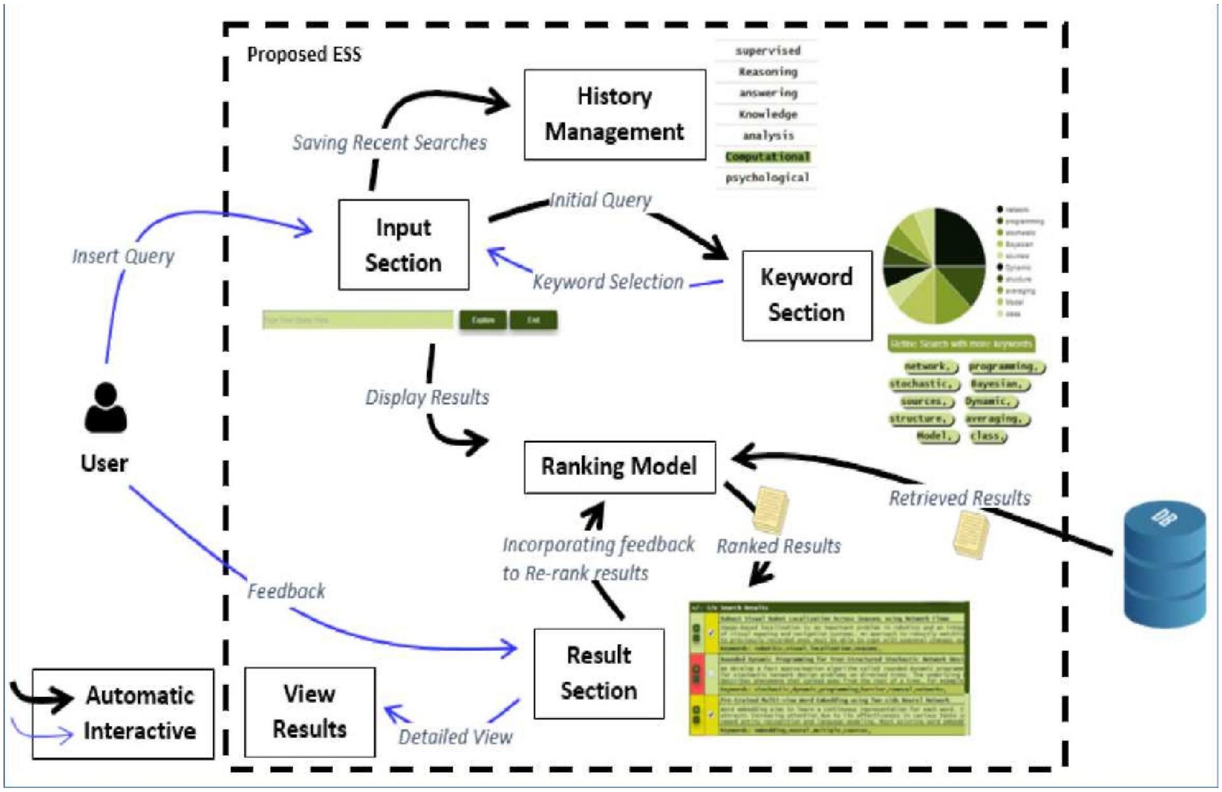


**Figure 1.3 Workflow of the proposed ESS**

### 1.3.2 Estimate Initial Intent Estimate

The proximity heuristic, however, has been largely under-explored in the literature [114, 123, 130]; mainly due the lack of certainty on how to model proximity and incorporated into an existing retrieval model [135, 140]. Though, The terms proximity is captured indirectly some retrieval models based on larger indexing units and employed in retrieval, [ 96, 103], but these models can only exploit proximity to a limited extent, since they do not measure the proximity of terms. Intuitively, the proximity score is estimated as query terms proximities within matched document/results. In the thesis, a strategy for 'Pre-search' context, via designed relevance measure is devised. The initial data samples for the user search are extracted via 'Pre-search' relevance score and subsequently passed to user review.

The key objective of the proposed work is to interactive iteratively model the PRF and interactively capture URF into intent estimation framework. Next, the notions relating to the revision in search intent are discussed. Both relevance-feedback components are adapted with appropriate definition and semantics and keeping an exploratory information-seeking into the focus. The revision of initial user search is entirely driven by the feedback incorporation into the estimate during the user search.

### 1.3.3 Revising Intent Estimates

The revision of search intents during an information search task is an essential phenomenon. The changing search goals and retrieval of new relevant result objects are the key reasons for this re-occurring scenario, during the information-seeking task. The adapted approach for the relevance modeling in proposed work is discussed in section

4.3, as part of modeling the 'In-search' context. The objective is to apply the revised/updated search interest during the search and thus modeling each component of relevance is important. In this thesis, two main drivers of relevance feedback are adapted: User relevance feedback (URF) and Pseudo relevance feedback (PRF).

The user relevance feedback (URF) is the key drivers to apply updated search interest explicitly in an ongoing search task. The semantic explanation of URF within as search task is shown in Figure 1.4, for a user search task on document collection and judgment on retrieved objects. The proposed system aide the user for applying the feedback on both intents on the currently displayed result objects, relevant (positive) and no-relevant (negative). These relevance feedbacks improve the relevance of result extraction in futuristic searches, as navigation or implicit data exploration becomes more user-centric and eventually leads to region-of-interest. Each interaction for feedback updates the current intent estimate (based on 'Pre-search' context); therefore, it is significant to model each explicit relevance feedback.

A user relevance feedback begins with the labeling of the retrieved results (Result documents terms) for a user query (Qiterms) with corresponding evaluated relevance score, over a user interface. The fetched results are generally visualized in ordered fashion or arrangement. In each iteration, a user can correlate result objects with the context of the current query with result context, and based on which label, as positive and negative intent.

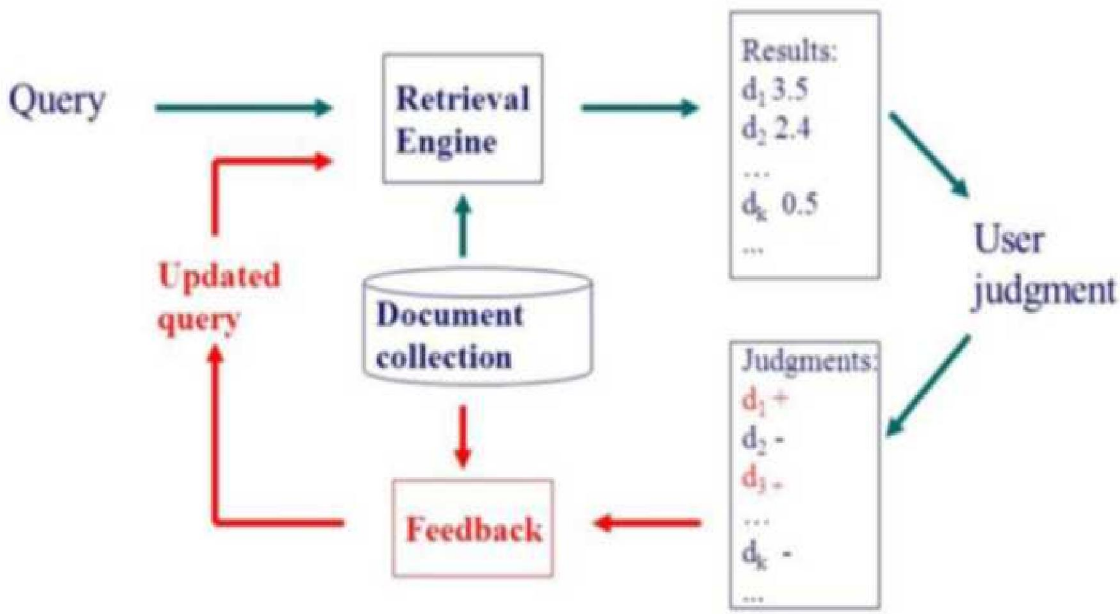


**Figure 1.4 Illustration of User-relevance feedback (URF) to revise intent estimate.**

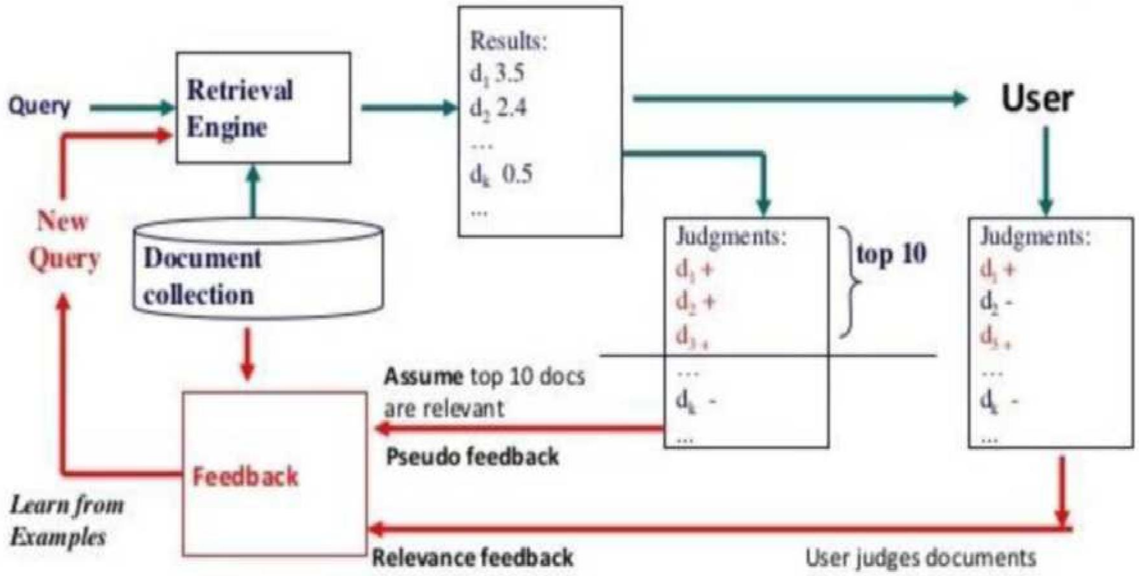


**Figure 1.5 Illustration of Pseudo-relevance feedback (PRF) to revise intent estimate**

Similarly, pseudo relevance feedback (PRF) is another driver for simulating the relevance notions into the retrieval framework of an initiated search [104]. The idea behind relevance feedback is to take the results that are initially returned from a given query, to gather user feedback, and to use information about whether or not those results are relevant to perform a new query. Figure 1.5 illustrates the contextual implication of implicit pseudo relevance feedback for a user information-seeking. The traditional PRF is commonly used to boost the performance of traditional information retrieval (IR) models by using top-ranked documents to identify and weight new query terms, thereby reducing the effect of query-document vocabulary mismatches.

In the thesis, an end-to-end PRF framework is modeled that can be used with existing IR models by embedding different URF models as building blocks of 'In-search' context to support an exploratory search task. The adapted definition of pseudo relevance feedback (PRF) incorporates use of relevance feedback information into proposed IR models. The simulated PRF framework uses feedback documents to better estimate relevance scores by considering individual feedback documents as different interpretations of the user's information need. The proposed system estimates the amount of retrieved result goes through pseudo relevance sub¬routine, among each top-n window, e.g., top-5 in top-10, top-4 in each of top-20 or top-30, etc.

### 1.3.4. Progressive Intent Estimation Model (IIM)

The evolution in search behaviour of information system is one fundamental characteristic, mainly to deal with the progression in the user search intent growth. The nature of an ideal information system is desired to progressively navigate the current exploratory searches in data regions of interest and efficiently assist in futuristic user searches. The ongoing search interactions are modelled into the systems nature and utilized in futuristic interactions. Here, exploratory search system adapts relevance measures to each stored information documents and revise these factors for each interaction.

The information retrieval research community has recently recognized the benefits of keyword search, starting to introduce keyword search capabilities into informational retrieval frameworks/models. The advantage for the keyword-based search to be supported by the underlying DBMS is quite clear; though integration task remains to be an open challenge. Most of the existing methods aides' keyword search via generating all possible results composed of relevant results such that these results can be arranged on individual ranks. This strategy is ineffective for top-keyword queries. The frequent keyword-queries results into huge answers. A progressive strategy for information retrieval that returns highly ranked results first is viable. This strategy significantly reduces the query response time via elimination of enumerating possible results and offers early completion of the user search.

In this thesis, an interactive exploratory search system is designed, in which three user search contexts are employed to the effectively capture the user search-preferences during the different phase of search and eventually incorporate into the retrieval framework. Initially, we designed the system with limited learning of search-interactions; the proposed 'IIM' system is solely based on the estimated intents (initial and revised) at various phases of the user search. The proposed system supports user to search via the 'Pre-search' estimation, and certain of documents, on which are poses the query-terms.

The progressive modeling of the proposed system begins with estimation of total documents to be go through the system for awards/penalties. In current system, we kept the estimate based on the average no. of matched documents (Avg-Match), for awards and penalties, in this total number of terms within each document will be estimated and accurately identifies as the number of document for awards and penalties. The overall strategy for the progressive modeling of user context via 'Pre-search' and 'In-search' context relevance score is described in Algorithm.

Algorithm 1.1 represents a progressive system that keep on growing itself with the increase number of the user-interactions. In the initial stage of the algorithm relevance factor (RF) corresponding to each document is defined as sum of DTscore and QTscore estimated based on the query term measures respectively. An attribute corresponding to intent log is included in each document and it keeps the information of, how many and which query terms (in user query) belongs to the document. These are two basic parameter intent log and keyword and relevance factor, where keyword vector keeps the information of query terms presence in the document.

For each relevant document, user feedback and pseudo relevance are estimated and inserted as a document attribute. In the top-k document, system estimate averages matched query proximity and document from top-2 document corresponding to average QT value, are kept with consistent relevance factor as current. Now for each document from document corresponds to avg QT to the last document of the list, new relevance factor as defined using awards and penalty mechanism. The updated relevance factor uses the nature of the document and matched QT along with URF and PRF in order to decide the award and penalty. The document of the list re-arranged according to the revised relevance factor and same strategies repeated for the next interaction of user search. It keeps on until find documents on once intended search interest.

An intent-log is maintained to keep the records of each document and corresponding keyword/terms viewed by the user during the user relevance feedback and subsequently PRF. The intent-log maintains the detail in keyword array and corresponding relevance factors (RF). Here, the relevance factor (RF) of a keyword/document-term will be controlled by the QT status, and values of QT as either 0 or 1. The strategy for updating the RF score will be integrated for each document stored in the repository.

Further, based on the matching condition with current user query terms, a document (D) is selected. The progressive strategy of search seeks an award for these resulting Ordered DocList(Di,D2, , Dn); therefore estimate the changes in current RF of document (D) as follows:

$$RF = RF + \frac{(^{(1}\sim^{RF})*1^{matched}\ Q^{T}\ \text{feedback}\ (URF_{sc()re} + PRF_{sc()re})))}{|QT|} \qquad (5\ (\ .\ )$$

and, similarly penalizes as :

$$_{RF} = _{RF}\ \_\ \frac{((\text{feedback}\ (URF_{S}core + PRF_{S}core) > 1QT\ 1)}{(QT + |\ \text{matched}\ QT\ |)} \qquad (\ .\ )$$

In the end, the entire document list (Doclist(D1, D2,..., Dr)) is re-ranked based on the updated relevance estimate and shown to the user. The system offers the opportunity for the continuation for the current search and terminates the search-interactions and ends the search process if intended search goals meet.

***Input:*** *OrderedDocList ($D_h D_2$, $D_n$) and user Query* ($Q_{t1}$^ _ $_{tl}$) ***Output:***
*Relevance Ordered Exploitable Doclist($D_1$, $D_2$,..., $D_r$)*
***Initialize***
***for*** each document in *OrderedDocList($D_1$,$D_2$, Dn)*
***Estimate*** TF-IDF and QTP for user query ($Q_{t1t2\_tk}$); //as per *Algorithm 3.1 & 3.2* //
//initialize R.F. for each document as *Pre-search relevance score//*
Relevance-Factor (R.F.)=(DTscore+QTscore);
***Re-arrange*** the *Doclist($D_1$,$D_2$,^,$D_r$)* based on current *RF*
***end***
***for*** *each document in OrderedDocList($D_1$,$D_2$, $D_n$),* //The intent-log maintain the list of keywords to TOrtritjiites, m tefiim^ current R.F. & R.F. value)//
***Initialize****Intent-log(keyword[],* R.F.) = ***{***
***for*** (QT=1; QT<=t; QT++)
{ ***if*** *(QT*= = matched is $D_i$) ***then*** *key* [Q7]=1; ***else*** *key* [QT]=0; }
***end }***
***end***
***for*** each document in *Ordered DocList(D1,D2,., Dn),*
***Apply*** *URF-PRF* mechanism and incorporate the user-preference //as *Algorithm 4.1 & 4.2//*
***end***
***for*** each document in *OrderedDocList($D_h D_2$,..., Dn)* *H* The documents for award/ penalizes identified, based on average *matched QT//*
Avg-Match= (2k=$_i$ (matched QT))/k ;
***end***
***j=0;***
***While*** (j>=0)
*{**if** (matched QT*> average) ***then*** j=j+1; ***else*** break; *}* ***for***
each document in *DocList($D_1$,$D_2$,., $D_n$),*
***Update*** (Intent-log(keyword[], R.F.)={ //also mark the contributing Doc-Terms//
***for*** (QT=1;QT<=t;QT++)
{ ***if*** *(QT*= = matched is $D_i$) ***then*** *key* [QT]=1; ***else*** *key* [QT]=0; }
***end***
***if*** (i<=j) //to award the *ma tching documents* //
RF = RF + ((1 _ RF) * |matched QT| * feedback ($URF_{score}$ + $PRF_{score}$ )))/|QT|;
***else*** // to penalizes the *non-matching documents* //
RF = RF _ ((feedback($URF_{score}$ + $PRF_{score}$ )) * |QT|)/ (QT + |matched QT|);
***end***
***Re-arrange*** the *Doclist($D_h D_2$,..., D )* based on revise relevance estimate }
***While*** (Do you want to move to next search= Yes);

### 1.4 Evaluating Counterpart Systems: YmalDB, AIDE, IntentRadar, and uRank

The validation of exploratory assistant to the user in information-search with uncertain or unclear information goals is a key component of the designed work in the thesis. The literature study outlines that, in recent years, diverse tools/systems are designed to aid the user, in analogous and contextual searches via extensive embedded strategy for automatic exploration and extensive adaptive visualization over the interface. The designed system 'IIM' offers comprehensive support to the user data exploration during the search, and thus evaluated w.r.t. potential exploratory systems. Four contextually equivalent systems are identified from the existing literature based on the relative significance among the alternatives, for the assessment of the proposed system for the user search.

The YmalDB[39] and AIDE[36] are similar in some sense, as both support inherent data exploration via novel strategy for automatic exploration, here automatic exploration primarily driven by the iterative and interactive user relevance feedback. The user interactions are supported via simple user-interface and efficient extraction of data samples. Despite these aids, both tools unable to keep the highly relevant data objects in the futuristic searches, and rarely considered the query terms proximities. IntentRadar[112] and uRank[37], are designed with an objective to visualize the search intents in interfaces along with query results, the intent visualization offers a perspective of overall search progression and the user knowledge-state. All the above systems primarily rely on traditional implicit and explicit relevance measures.

The proposed system 'IIM', adapted query terms proximities and notions of relevance-feedback to emphasis the significance of the diverse context of user search into retrieval framework and eventually steered the implicit data exploration. An ideal exploratory search system accommodates implicit and explicit relevance measures for the proactive search intent estimation. Hence, the proposed framework steers both key drivers of exploratory information-seeking, i.e. focused-search based on DT and QT measures and exploratory-browsing based on URF and PRF, eventually to support the user-centric information-seeking. In YmalDB and uRank employed PRF based intent estimation on automatic clustered data objects, AIDE apply URF based object classification to model data retrieval strategy. The IntentRadar adapts URF to model the revised intent and PRF, though primarily rely on the 'Pre-search' based intent.

A brief comparison among the counterparts and the proposed system is presented in Table 1.1, along with information-seeking parameters. The summary of visual features, e.g., Visual Accomplishment and Visual Implications, to assist the user information-seeking on the proposed system and other four ESS is given in Annexure-D. The assessment of the designed system 'IIM', equivalent ES systems are employed system and presented in the next chapter (in chapter 5), on the predefined potential search trails.

**Table 1.1 Comparison schema of proposed system ('IIM') and adapted equivalent exploration systems (over the information-seeking parameters)**

| ***Tools Search Parameters*** ^\ | *YmalDB* | *AIDE* | *IntentRadar* | *uRank* | *IIM (proposed system)* |
|---|---|---|---|---|---|
| *Prior knowledge of Information-Needs* | Full | Partially | Partially | Full | Partially |
| *Result data Visualization* | Tabular View | Sample View | Intent-radar View | Term-cloud View | Intent-circle View |
| *Exploration supported via* (action for search revision) | Drag-Drop keyword | Drag-Drop keyword | Drag-Drop keyword | Drag-Drop keyword | Click-hold keyword |
| *Relevance-feedback supported, via* | Only PRF | Only URF | URF+PRF | Only URF | URF then PRF |
| *Progression of Exploration* | Multistep and Random | Multistep and Linear | Multistep and Random | Multistep and Linear | Multistep and Random |
| *Suitable for User type* | Programmer & Expert | Programmer & Expert | Programmer & Expert | Naive & Programmer | Naive, Programmer, & Expert |

**References:**


[1] Acharya, S., Gibbons, P. B., & Poosala, V. (1999, September). Aqua: A fast decision support systems using approximate query answers. In Proceedings of the 25th International Conference on Very Large Data Bases (pp. 754-757).

[2] Abouzied, A., Hellerstein, J. M., & Silberschatz, A. (2012). Playful query specification with DataPlay. Proc of the VLDB Endowment, 5(12), 1938-1941. Andrews, K., Kienreich,

[3] W., Sabol, V., Becker, J., Droschl, G., Kappe, F., & Tochtermann, K. (2002). The infosky visual explorer: exploiting hierarchical structure and document similarities. Information Visualization, 7(3-4), 166-181.

[4] Agarwal, S., Iyer, A. P., Panda, A., Madden, S., Mozafari, B., & Stoica, I. (2012). Blink and it's

done: interactive queries on very large data. Proceedings of the VLDB Endowment, 5(12), 1902-1905.
[5] Ahn, J. W., Wongsuphasawat, K., & Brusilovsky, P. (2011). Analyzing user behavior patterns in adaptive exploratory search systems with lifeflow.
[6] Albarrak, A. A. (2018, July). RELATE: Preference-aware Correlation-based Query Refinement. In Proceedings of the 4th ACM International Conference of Computing for Engineering and Sciences (p. 1). ACM.
[7] Andolina, S., Klouche, K., Peltonen, J., Hoque, M., Ruotsalo, T., Cabral, D. & Jacucci, G. (2015, March). Intentstreams: smart parallel search streams for branching exploratory search. In Proceedings of the 20th International Conference on Intelligent User Interfaces (pp. 300-305). ACM.
[8] Ariguzo, G. C., Mallach, E. G., & White, D. S. (2006). The first decade of e-commerce. International Journal of Business Information Systems, 7(3), 239-255.
[9] Athukorala, K., Medlar, A., Oulasvirta, A., Jacucci, G., & Glowacka, D. (2016, March). Beyond relevance: adapting exploration/exploitation in information retrieval. In Proceedings of the 27st International Conference on Intelligent User Interfaces (pp. 359-369). ACM.
[9]. Baeza-Yates, R., & Ribeiro, B. D. A. N. (2011). Modern information retrieval. New York: ACM Press; Harlow, England: Addison-Wesley,.
[10]. Baigelenov, A., & Parsons, P. (2018, April). Interactivity Factors in Visualization-Based Exploratory Search. In Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems (p. LBW009). ACM.
[11]. Bailey, S. M., Wei, J. A., Wang, C., Parra, D., & Brusilovsky, P. (2018). CNVis: A Web-Based Visual Analytics Tool for Exploring Conference Navigator Data. Electronic Imaging, 2018(1), 1-11.
[12]. Balcetis, E., & Dunning, D. (2006). See what you want to see: motivational influences on visual perception. Journal of personality and social psychology, 91(4), 612.
[13]. Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. Online review, 13(5), 407-424. [14]. Barry, C. L. (1994). User-defined relevance criteria: An exploratory study. Journal of the American Society for Information Science, 45(3), 149-159.
[15]. Beigbeder, M., & Mercier, A. (2005, March). An information retrieval model using the fuzzy proximity degree of term occurrences. In Proceedings of the 2005 ACM symposium on Applied computing (pp. 1018-1022). ACM.
[16]. Belkin, N. J., & Croft, W. B. (1992). Information filtering and information retrieval: two sides of the same coin. In Communications of the ACM.
[17]. Belkin, N. J. (1993). Interaction with texts: Information retrieval as information seeking behavior. Information retrieval, 93, 55-66.
[18]. Belkin, N. J. (1990). The cognitive viewpoint in information science. Journal of information science, 16(1), 11-15..
[19]. Bennett, P. N., Radlinski, F., White, R. W., & Yilmaz, E. (2011, July). Inferring and using location metadata to personalize web search. In Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval (pp. 135-144). ACM.

6/17/2022