UNIFIED IMPLICIT AND EXPLICIT FEEDBACK FOR MULTI-APPLICATION USER INTEREST MODELING

A Dissertation

by

UKWATTA KANKANAMALAGE SAMPATH JAYARATHNA

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Frank Shipman
Committee Members,	Richard Furuta
	James Caverlee
	Ann McNamara
Head of Department,	Dilma Da Silva

August 2016

Major Subject: Computer Science

ABSTRACT

A user often interacts with multiple applications while working on a task. User models can be developed individually at each of the individual applications, but there is no easy way to come up with a more complete user model based on the distributed activity of the user. To address this issue, this research studies the importance of combining various implicit and explicit relevance feedback indicators in a multi-application environment. It allows different applications used for different purposes by the user to contribute user activity and its context to mutually support users with unified relevance feedback. Using the data collected by the web browser, Microsoft Word and Microsoft PowerPoint, Adobe Acrobat Writer and VKB, combinations of implicit relevance feedback with semi-explicit relevance feedback were analyzed and compared with explicit user ratings.

Our past research show that multi-application interest models based on implicit feedback theoretically out performed single application interest models based on implicit feedback. Also in practice, a multi-application interest model based on semi-explicit feedback increased user attention to high-value documents. In the current dissertation study, we have incorporated topic modeling to represent interest in user models for textual content and compared similarity measures for improved recall and precision based on the text content. We also learned the relative value of features from content consumption applications and content production applications. Our experimental results show that incorporating implicit feedback in page-level user interest estimation resulted in

significant improvements over the baseline models. Furthermore, incorporating semi-explicit content (e.g. annotated text) with the authored text is effective in identifying segment-level relevant content.

We have evaluated the effectiveness of the recommendation support from both semi-explicit model (authored/annotated text) and unified model (implicit + semi-explicit) and have found that they are successful in allowing users to locate the content easily because the relevant details are selectively highlighted and recommended documents and passages within documents based on the user's indicated interest. Our recommendations based on the semi-explicit feedback were viewed the same as those from unified feedback and recommendations based on semi-explicit feedback outperformed those from unified feedback in terms of matching post-task document assessments.

DEDICATION

In memory of my **father** Mr. Jayarathna (1948-2001)

you left fingerprints of grace on our lives. You shan't be forgotten

This is for you, **Nim** and **Pahan**; you have been a constant support and encouragement during the challenges of graduate school and life. I am truly thankful for having you in my life

This is for you, **Mom**. Thanks for always being there for me

ACKNOWLEDGEMENTS

Foremost, I would like to express my sincere gratitude to my advisor Dr. Frank Shipman for the continuous support of my Ph.D. study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis and getting a tenure track job. I could not have imagined having a better advisor and mentor for my Ph.D. study.

Besides my advisor, I would like to thank the rest of my thesis committee: Dr. Richard Furuta, Dr. James Caverlee and Dr. Ann McNamara, for their encouragement, insightful comments, and hard questions.

I thank my fellow colleagues (past and present) at Shipman Lab and Center for Study of Digital Libraries for the stimulating discussions (mostly about food), and for all the fun we have had in the last six years; Atish, Faryaneh, Luis, Gabe, Suin, Josh, Caio, Prasanth and Arun.

Most importantly, none of this would have been possible without the love and patience of my family. My immediate family to whom this dissertation is dedicated to, including my wife Nimmi and son Pahan have been a constant source of love, concern, support and strength all these years. I would like to express my heart-felt gratitude to my family and my in-laws which have aided and encouraged me throughout this endeavor.

Finally, I sincerely appreciate the financial support from Department of Computer Science & Engineering that funded parts of the research discussed in this dissertation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Page

1. INTRODUCTION

In everyday tasks, users often handle a large volume of information from the web, and routinely face with difficulty in finding useful information. As the amount of information available for consumption causes "information overload," there's a high demand for personalized approaches for information access. *Personalized information delivery* is a possible solution for this particular problem in gathering, personalizing and providing appropriate information to users. The quality of these personalization efforts are mostly depends on the information beyond what is merely expressed in a user's query. The traditional "one-size-fits-all" approach used in search systems has been replaced by the idea of this "personalizing" results for specific users based on the user preferences, context, and particular information need (Liu and Belkin 2010). These customization efforts may take the form of filtering out irrelevant information from available resources and identifying additional resources of interests to the users. Personalization also allows providers to gather, filter content, adjust and format to an individual user's needs and preferences. Considerable research has addressed the problem of personalization in the context of search and assessing the relevance of document to the user's information needs (James, Hinrich , Todd , Rob , Don et al. 2002, Micarelli, Gasparetti, Sciarrone and Gauch 2007, Pasi 2010, Bennett, White, Chu, Dumais, Bailey et al. 2012). Research in personalization is ongoing in the major fields such as *information retrieval, data mining, artificial intelligent*, among others. The effectiveness of these approaches are strongly contingent upon the quantity as well as the quality of information available about the user and her preferences.

(Barla 2011) explains how the user modeling and personalization can be separated into three distinct stages of *data collection*, *user model inference*, and *adaptation and personalization*. This process has cyclic characteristics in personalization and adaptation by continuously acquiring new relevance information about the user and by refining a user interest model to better reflect the learned inference to serve personalization efforts. There are number of issues associated with the each step of the process. The data collection should balance between user privacy and the amount of data to needed to deliver successful personalization. When a user accesses the personalization system for the first time, there is not enough information to provide efficient personalization. This *cold-start problem* poses a challenge to early application activity. These personalization activities in the early stage of the system usage can be critical to user retention (Tsiriga and Virvou 2004). More generally, the sources of such data should not pose additional burden on the user and the data collection process should be unobtrusive by nature. Finally, the inference techniques should be able to maintain a relationship to user characteristics with changes to their personal development, interest and knowledge.

Relevance feedback is an interactive activity in which the system engages the user in iteratively formulating a user model to fulfill information needs based on the user's expectations. Such a user-system interaction is not usually a single user-system interaction based solely on a user query and a resultant list of items that the system has evaluated as relevant. There has been a shift from this "blind" and closed behavior of first generation of search systems to assessment of multiple relevance dimensions motivated by the deep study of the notion of relevance (Saracevic 2007, Pasi 2014). Inferring perceived relevance of information content delivered to the user is a central task of interactive information retrieval systems (Moshfeghi, Pinto, Pollick and Jose 2013). Perceived relevance can then be used to represent user preferences (Kelly 2009), used as an input for a search tasks (Ruotsalo, Peltonen, Eugster, Głowacka, Konyushkova et al. 2013) and to measure the user's satisfaction (Fox, Karnawat, Mydland, Dumais and White 2005) with the personalization effort of the system.

Although relevance feedback has received a great deal of attention in the user modeling literature on IR and search personalization, very little work has been done to study the process of unifying these heterogeneous relevance feedback in multi-application environments. Many existing personalized information delivery require user interventions in terms of explicitly indicating interests, or interrupting users during their activity to recognize user preferences. The work presented in this dissertation addresses this rarely investigated topic: the potential of aggregating activity across multiple applications for user interest modeling. While there are theoretical or software frameworks for distributed user modeling, assessments of modeling techniques are almost always reported in terms of single applications. In this work, we present and evaluate a multi-application modeling technique that combines implicit and semi-explicit feedback across multiple everyday applications. The following section addresses problems and issues, and Section 3 provides an overview of related work. Section 4 describes the architecture, interfaces, and other capabilities. Evaluation of user models and findings are discussed in Sections 5. Finally, Section 6 addresses conclusions and future work.

2. PROBLEMS AND ISSUES

Detailed knowledge about a user's interests is beneficial in web search, advertising, and personalized recommendations as well as in content targeting. The goal of personalized recommendations is to support users by identifying documents or the parts of a document that best match user's interests during an open-ended information gathering task. Such recommendations can result in a more efficient use of the user's time, e.g. that their time is spent on the most relevant documents.

2.1 Too Many Documents, Too Little Time

Our past research shows that time is frequently a limiting factor in web search tasks: there are too many documents to assess and too much reading to do. The problem in such a search task is that even with the best web search engines, and the most effective query formulations, these tasks require people to work through long list of documents to examine potentially relevant documents or part of a document. Most users skim early documents, find portion of a document relevant to the current query, and determine additional information needs that result in further queries and more documents to process (Bae, Kim, Meintanis, Moore, Zacchi et al. 2010).

A user's query provides the most direct evidence for a particular information need when creating a user model, and most existing retrieval and personalized information delivery systems rely solely on query inputs to create these user models (Shen, Tan and Zhai 2005). However, query inputs are often short and natural language is inherently ambiguous, therefore the resulting user interest models are inevitably impoverished.

Perhaps due to the difficulty in expressing a more precise query, many queries consist of only a few keywords to model the actual information need (Jansen, Spink and Saracevic 2000). These short queries often contain only marginally informative content about user's actual intension therefore may return search results not relevant to the intended query concept (Stamou and Ntoulas 2009). In addition, query term mismatch is often compounded by synonymy and polysemy (Carpineto and Romano 2012), resulting in user confusion. In order to mitigate the inherent ambiguity of queries, web search engines are employing user models to customize search results based on the inferred interests of the user. The belief is that detailed knowledge about a user's interests, i.e. the *user interest model*, can improve support of searching and browsing activities as every user has a particular goal and a distinct combination of context and background knowledge (Sieg, Mobasher and Burke 2007).

2.2 Relevance Feedback

As an alternative approach to improve the interest modeling, explicit feedback can be used to verify with the user how relevant or useful or satisfying the given documents are for her information need. But in a real world scenario, users are usually reluctant to make the frequent ratings of documents without an immediate benefit from their efforts (Grudin 1994, Kelly and Teevan 2003, Bae, Kim, Meintanis, Moore, Zacchi et al. 2010). Also, users rate far fewer documents than they read which is basically due to the interference of providing frequent explicit feedback with their normal reading and browsing patterns (Sarwar, Konstan, Borchers, Herlocker, Miller et al. 1998). User activity beyond explicit relevance feedback can also be used to infer interests. Annotation and clipping behaviors provide more direct evidence of user interest while browsing and reading behaviors, such as dwelling/reading time, mouse clicks, mouse movements and scrolling provide more indirect evidence of user interest.

2.3 Challenges in Personalized Information Delivery

Even though personalized information delivery has the potential to provide users accurate results relevant to search intensions, personalization is particularly challenging due to two key issues. First, it requires identifying the interests of users in semi-persistent user profiles. Estimating user preferences in a real user interaction with a web search engine is a challenging problem, since the interactions tend to be more noisy than controlled settings (Agichtein, Brill and Dumais 2006). Second, given the user preferences recorded in a user profile, personalized information delivery requires a way to alter the presentation of search results to reflect those preferences. This dissertation is focused on the first of these problems. A challenge for user interest modeling is that a particular user interacts with a limited amount of information while working on any particular task. It also takes time for users to search and select information before they understand what they really want. As a result, user modeling techniques may not understand what is of value to the user until it is too late and their interest has shifted.

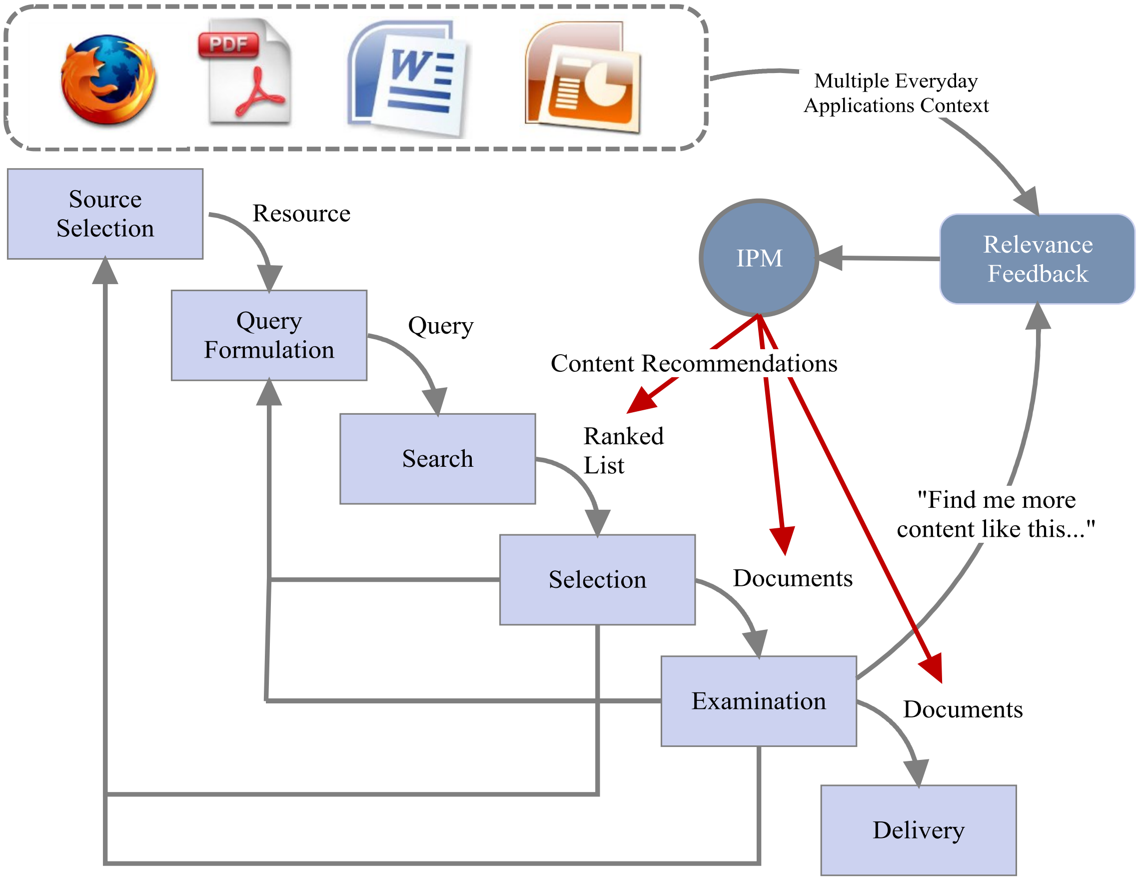

Figure 1: IR Cycle with Interactions in Everyday Applications

2.4 Multi-Application Environments

Real-world personalization is often dynamic in nature and information delivered to the user can be automatically personalized and catered to individual user's information needs (Lu, Agarwal and Dhillon 2009). Figure 1 presents the standard information retrieval process in a web search environment and interactions across multiple everyday applications. People interact with different applications, and have extra information about the content they are interacting with. These interactions results in implicit feedback (e.g., click-through data, reading time) and semi-explicit feedback (e.g., annotations) data that varies depending on their task and the type of information being explored. For example, a user may examine a list of search results in a web browser; she may use MS Word or PDF Reader to examine the contents of individual documents; she may use a note-taking tool to keep track of interesting snippets; and she may use MS Word or a presentation tool to author her own interpretation of what she has found. Therefore, a user model extracted from a single application is unlikely to be as effective as a user model based on the aggregate activity across applications (Badi, Bae, Moore, Meintanis, Zacchi et al. 2006). The particular approach being explored here looks to broaden current techniques by including a variety of direct and indirect evidence of interest across multiple applications.

3. LITERATURE REVIEW*

3.1 Relevance Feedback

Relevance feedback has a history in information retrieval systems that dates back well over thirty years and has been used for query expansion during short-term modeling of a users' immediate information need (Kelly and Teevan 2003). Relevance feedback has been one component of the notion of context applied towards interactive search where the user can explicitly interact with the system to judge the relevance of information presented to her needs (Salton and Buckley 1997). With the combination of the context and explicit indication of relevance to the information, systems can better capture user preferences and alter the presentation of information. In recent years, there has been a shift from explicit to implicit techniques motivated by the need of obtaining preferences unobtrusively interrupting or burdening users. With these implicit techniques, user-system interactions are learned or inferred to collect elements of context in these interactions (Kelly and Belkin 2002, Fox, Karnawat, Mydland, Dumais and White 2005, Speretta and Gauch 2005). Therefore to capture the user’s interests, two main techniques of relevance feedback may be employed, namely *(i) implicit*: information can be derived by studying users behavior while using services *(ii) explicit*: information can be gathered by a direct intervention of the users themselves by filling some kind of predefined forms (Ruthven and Lalmas 2003, Viviani, Bennani and Egyed-Zsigmond 2010).

---

* Jayarathna, S., Patra, A., and Shipman, F. “Unified Relevance Feedback for Multi-Application User Interest Modeling,” Proceedings of the 15th ACM/IEEE-Cs Joint Conference on Digital Libraries, pp.129-138, © 2015 Association for Computing Machinery, Inc. Reprinted by permission. http://doi.acm.org/10.1145/2756406.2756914

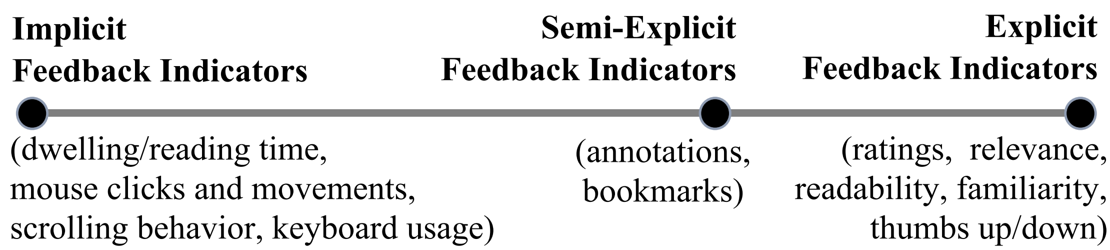

Figure 2: Types of Relevance Feedback Indicators

Figure 2 shows how user actions form a continuum from implicit to explicit feedback. There is a clear tradeoff between the quantity and quality when comparing implicit feedback with explicit feedback. Explicit feedback indicators are higher in quality but lower in quantity because it is rather burdensome to enter a rating for every item a user liked or disliked (Liu, Xiang, Zhao and Yang 2010). On the other hand, implicit feedback indicators are abundant in quantity but lower in quality because they must be interpreted by heuristic algorithms that make assumptions about the relationships between the observable low-level actions and the high level goals of users. In (Nichols 1998), authors evaluated the costs and benefits of using implicit feedback indicators over explicit feedback indicators. The results suggested that the implicit ratings can be combined with existing explicit ratings to form a hybrid system to predict user satisfaction. In (Jawaheer, Szomszor and Kostkova 2010), authors showed that implicit and explicit positive feedback complement each other with similar performances despite their different characteristics. This implies that systems can be designed to use the correlation between implicit and explicit feedback to tune the interest modeling algorithms based on implicit feedback.

3.2 Explicit Relevance Feedback

User guided modeling techniques such as explicit relevance feedback systems rely on the information provided by the user to build a user model (Marios, Efi, Panagiotis and George 2013). Explicit feedback requires users to assess the relevance of documents or to indicate their interest in certain aspects of the content. Explicit evidence can also be obtained by direct intervention, asking the user about their preferences usually by giving a questionnaire when interacting with the system and asking users to select keywords or topics pertinent to their interests (Germanakos, Tsianos, Lekkas, Mourlas and Samaras 2008). Alternatively one can ask users for feedback about the items they have browsed using binary evaluations (e.g., like/dislike), ratings (5-point scale) and text comments. The data collected may contain demographic information such as age, gender, marriage status, profession, interests and/or preferences or personal information (Gauch, Speretta, Chandramouli and Micarelli 2007). In addition some methods allow users input via checkboxes and text fields by selecting values from a range. Many commercial systems have been exploring personalization for some time based on user preferences in order to customize interfaces. For instance, iGoogle[†] (Casquero, Portillo, Ovelar, Romo and Benito 2008), My Yahoo[‡], NetVibes[§], and uStart[**] are commonly utilized for customizing user interfaces by collecting user preferences to create user profiles and services to adapt in order to increase the information accessibility. These web site contents are then

† http://www.google.com/ig
‡ https://my.yahoo.com
§ http://www.netvibes.com
** http://www.ustart.org

dynamically organized based on the collected user preferences. The user preferences in an interest model can also serve personalize services provided by other applications in order to improve the user satisfaction.

There are numerous techniques in the IR research such as Curious Browser where explicit feedback are in the form of user ratings of document relevance such as “relevance”, “readability” and “topic familiar before” ratings (Zigoris and Zhang 2006). WebMate (Chen and Sycara 1998) learn and keeps track of user interests incrementally with multiple pages provided explicitly by the user as relevance guidance. It extracts keywords from these pages and uses them for keyword refining in query formulations. Similarly, InfoFinder (Krulwich and Burkey 1997) system learns user profiles from sample documents that users submit while browsing. The system learns general profiles from the text that are likely to represent the users’ interests in document topics. Similarly, in contextual relevance feedback (Harper and Kelly 2006, Limbu, Connor, Pears and MacDonell 2006), the search results list is filtered based on user-collected document piles that are used as user profiles.

Explicit feedback has the advantages that it can be easily understood, is fairly precise and requires no further interpretation (Claypool, Le, Wased and Brown 2001). However these techniques also have some disadvantages. Generally, asking a user to complete a preliminary questionnaire or to identify keywords/topics of interest interferer with the natural interaction of the user (Hijikata 2004). Grading pages or rating items might also takes time away from the user’s main activity. Both direct and semi-direct explicit methods require users to invest effort and their willingness varies according to the application they are interacting with (Schiaffino and Amandi 2004). Because of these additional burdens on the user, and/or privacy concerns, users may not choose to participate. Also users may not accurately report their own preferences, interest or demographic data and the user's interest may change over time by making their user model increasingly inaccurate.

In some cases, users enjoy providing and sharing their feedback. This is most evident in services relevant to consumer products such as movie ratings Netflix[††], Movielens[‡‡] and sites dedicated to collecting and sharing streaming music such as Pandora[§§], Last.fm[***]. All these explicit relevance feedback collection techniques have the advantage that the form of the replies are more standardized than other relevance feedback techniques such as implicit feedback. The main drawback is that the user's interaction with the system may be disrupted due to unwillingness of the user to provide the preference information due to lack of trust or time to participate in the process (Marios, Efi, Panagiotis and George 2013). In addition, users may not accurately or fully report their preferences and systems may not have facilities to update when the preferences have changed.

3.3 Semi-Explicit Relevance Feedback

Some user actions, particularly bookmarking and clipping, can be interpreted as semi-explicit feedback in that the user's action is a clear evidence of their desire to re-

†† http://www.netflix.com
‡‡ https://movielens.org
§§ http://www.pandora.com
*** http://www.last.fm

access this content. There are also various applications and services that can enable creating tags, highlights and other types of annotations allowing the users to provide additional information sources while reading electronic documents.

While reading *printed* documents, it is a common practice to write down various types of notes, underlines, and highlights as a mean of storing our thoughts, marking interesting parts of documents and for the ease of navigation later on. Similarly, many online tools allow such behaviors with electronic documents and add to the value of the information presented (Oard and Kim 2001). *Annotations* created by user can be consider as a form of user's context while reading documents (Navrat 2012) or a body of words marked among text with the meaning of its position and content and what text it contains (Haiqin, Zheng and Qingsheng 2003). For a particular annotation, the surrounding text defines its *context*. A user can mark-up a portion of a document by highlighting a paragraph or attaching an electronic sticky note. Not all reading results in user annotations. Annotations are most likely when people read materials crucial to a particular task at hand and are infrequent when reading for fun (Shipman, Price, Marshall and Golovchinsky 2003). Annotations are used to identify which documents or portions of documents are interesting. But, if a document is large, users will frequently skim or stop reading when they feel they have met their information need. Consequently, potentially better document contents are left having never been reviewed (Badi, Bae, Moore, Meintanis, Zacchi et al. 2006). Visualizations can also draw user's attention to similar documents or document parts (Bae, Kim, Meintanis, Moore, Zacchi et al. 2010). Such visualizations include colors and icons to highlight annotated contents in a document overview (Price, Schilit and

Golovchinsky 1998). Spatial hypertext systems such as VIKI (Marshall and Shipman 1995) and VKB (Shipman, Hsieh, Maloor and Moore 2001) use similar visualization techniques to provide system-identified "interesting document contents" to support navigation. M4Note (Rudinei, Renan, Jose, Valter and Maria 2004) is designed as a way of providing annotations as metadata for indexing, retrieval, semantic processing and content enrichment. This can generate a structured document with an underlying description model that can be used in computations such as personalized tag hierarchies to support content enrichment. Recent work has also been conducted to study the user's post-click behaviors relevant to interactions with text selections or highlights (Guo and Agichtein 2012, White and Buscher 2012). In these studies, text selection is used to find the search performance for queries by clustering users based on their similar behaviors.

3.4 Implicit Relevance Feedback

Implicit feedback techniques have the advantage that the necessary data can be collected easily without burdening users. Implicit interest indicators are based on user actions and not on explicit value assessments. During a search task, readers may indicate their interest in documents by how they interact with them: by how much of the document they examine (e.g. how far into a document they scroll); and through other behaviors and events that are specific to the tools they are using. This interest may be recorded as users interact with documents and may be characterized via feature extraction. In learning from user behaviors, personalization attempts to infer user interests from logs of user activity, such as dwell time, click through, and other salient behaviors that can be easily captured. For example, the Curious Browser (Claypool, Le, Wased and Brown 2001) records various types of implicit feedback include aspects of mouse usage, keyboard usage and the time spent viewing documents.

There has been a broad range of research conducted in the area of interpreting user interactions data for implicit relevance feedback. This work can be divided into research aiming at *object-level* feedback and research on *segment-level* feedback (Buscher, Van Elst and Dengel 2009). In many situations records of user activity have been used to estimate object-level relevance, that is relevance for entire documents (Oard and Kim 2001). This is in line with object-level feedback that is needed in the classical information retrieval scenarios where systems are adapting the response set for a query (Gerard 1971). Studies of these object-level implicit feedbacks have often focused on correlation between reading time and explicit feedback based on document length and textual features (Morita and Shinoda 1994, Claypool, Le, Wased and Brown 2001, Fox, Karnawat, Mydland, Dumais and White 2005). There is strong evidence that user's spend more time on interesting articles than uninteresting ones. For example, there is a weak correlation between the document length and associated reading times because users tend to read in part and not entirety Additionally, assumptions about user work practice s complicate generalizing results but study of a more naturalistic scenario (Kelly and Belkin 2001, Kelly and Belkin 2002, Kelly and Belkin 2004) found that there was no general relationship between display time and the user's explicit ratings of document relevance. The high variation of display time with respect to the different user and different task has led to adjusting display time thresholds for implicit feedback based on task type (White and Kelly 2006). This confirmed the idea of adjusting display time thresholds according to task type leads to improved performance. (Rafter and Smyth 2001) showed for one specific task, adjusting individual measures can correlate the display time with user interest. In addition to considering display/reading time, additional studies have found that scrolling, exit type (Fox, Karnawat, Mydland, Dumais and White 2005), click-through (Joachims, Granka, Pan, Hembrooke, Radlinski et al. 2007) has been found to provide good indications of interests.

Segment-level feedback is much less explored compared to document or object-level methods. The user search behavior for estimating passage relevance for re-ranking is mostly done by studying correlation between segment-level display time and segment-level feedback from an eye tracker (Buscher, Dengel and Van Elst 2008) (Buscher, Dengel and Van Elst 2008, Buscher, Van Elst and Dengel 2009). In (Kong, Aktolga and Allan 2013) user behavior information from section relevance has been used to improve section ranking. More specifically, four types of user search behaviors, dwell time, highlighting, copying and click at various section levels were used to improve section rakings. This study also reveals characteristics of user behaviors at segment-level. Based on the resultant dataset, authors claim that about 50% of segment-level dwells are shorter than 2 seconds, suggesting users skim many sections instead of reading them to entirety. For clicks, authors reveal that users tend not to click on sections in the top part of pages because they are already being displayed resulting in a position bias for section clicks.

Mouse clicks and movements can also indicate the relevance of content to the user. (Claypool, Le, Wased and Brown 2001) report the amount of scrolling on the web page along with several other implicit measures and their relation to explicit indicators. The dwell time on the web page, the amount of scrolling and the combination of dwell time and scrolling led to the most arcuate predictions of user behaviors. In addition (Kríž 2012) has shown that time spent scrolling is a strong indicator of users potential interest in content. (Kantor, Boros, Melamed, Meñkov, Shapira et al. 2000) explore how users would follow the mouse pointer with their eyes while reading content. (Chen, Anderson and Sohn 2001). Their study indicates that, when participants use the mouse, 75% of the time the mouse and eye move to the same region of the screen. This suggests that there is a high correlation between mouse movements and eye movements. (Cooke 2006) confirms these findings by suggesting many users are active mouse users as they search for information, with 69% of the time mouse movements matching the eye movements. (Guo and Agichtein 2008) shows mouse movements as a way of inferring query intent based on the trajectories of mouse movements. The mouse movement data was superior to click-through data. The average accuracy of intent inference from click-through data was 62.95% while it was 70.2% using mouse movements. (Mueller and Lockerd 2001) suggest users tend to rest their mouse while reading and more detailed analysis of a user mouse movement can infer the user query intent. (Rodden and Fu 2007) further investigate this idea and introduce a user study to detect mouse movements in real time in order to recognize them as they occur. The study suggests that users tend to hesitate on links or text before clicking and that could potentially indicate there is more information on the page that are of interest to the user.

3.5 Topic-Level Relevance Indicators

The document content filtering based on a user interest model can be done by topics discussed in relevant documents and then by representing these document collections in a vector space (Tang and Vemuri 2005). The learning algorithms used in traditional vector space models are usually divided into supervised learning, unsupervised learning and semi-supervised learning. The process of supervised document filtering is called classification and unsupervised filtering is called clustering.

Topic models learn bag of words from a collection of documents without any supervision (Stevens, Kegelmeyer, Andrzejewski and Buttler 2012). Topic models assume generative model which can be used to model a collection of documents by topics. These generative topic models can reveal topic level relations based on the words used within a document. Three major distinct approaches for topic modeling are the Latent Dirichlet Allocation (LDA) (Blei, Ng and Jordan 2003), Latent Semantic Analysis (LSA) (Dumais 2005) and Non-negative Matrix Factorization (Lee and Seung 1999). As shown by (Xu, Liu and Gong 2003, Shahnaz, Berry, Pauca and Plemmons 2006), NMF outperforms traditional vector space approaches for document clustering such as LSA and learn concise topics with similar performance with LDA. However, NMF learns more incoherent topics compared to LDA(Stevens, Kegelmeyer, Andrzejewski and Buttler 2012).

A particular document can be encoded in an n-dimensional vector where n is the total number of terms in the corpus. Each vector defines the relative importance of corresponding terms with respective to the semantics of the given document (Salton, Wong and Yang 1975). In this *vector space model*, a collection of documents can effectively represented as a document-by-term matrix with a positive weight per corresponding term presented in the document or zero value otherwise. Given a term-by-document matrix with inherent non-negativity, the NMF (Lee and Seung 1999) can learn the underlying semantics or patterns in a text collection based on non-negative lower rank factors. The documents can be reconstructed combining these learned semantic features and set of documents with common features can be represented by a cluster.

(Harvey, Crestani and Carman 2013, Vu, Song, Willis, Tran and Li 2014) utilize LDA to determine user profile based on the latent topics from relevant documents. In this work, topic space is determined based on the relevant documents extracted from the query logs from user's web search history. Mehrotra (Mehrotra 2015) explores the possibility of modeling users search tasks by coupling topical interests with the search task behavior to learn user representations. (Majumder and Shrivastava 2013) present an approach treating online service platforms (OSP) such as search engines, news websites, ad-providers etc., as black boxes and extract their output to formulate latent topic personalization (LTP).

The topics of the relevant documents are often obtained from human generated online ontologies such as Open Directory Project (ODP)[†††] (Bennett, White, Chu, Dumais, Bailey et al. 2012, Raman, Bennett and Collins-Thompson 2013, White, Chu, Hassan, He, Song et al. 2013). In addition, click entropy (Teevan, Dumais and Liebling 2008, Teevan,

[†††] www.dmoz.org

Dumais and Horvitz 2010, Song, Nguyen, He, Imig and Rounthwaite 2011) uses the ODP distributions for analyzing search content of pages. Web pages with low entropy is considered to have higher search focus (Kim, Collins-Thompson, Bennett and Dumais 2012). These approaches are mainly limited in functionality because many documents may not contain topics covered in online ontologies. Also, human-generated topics require expensive manual effort to categorize each document.

3.6 Hybrid (Implicit and Explicit) Relevance Feedback

Hybrid relevance feedback methods attempt to exploit the benefits of implicit and explicit approaches. Hybrid methods can generate accurate user preference because implicit feedback lowers the user's workload and explicit feedback compensate for the sparseness and inadequacies of implicit feedback (Paliouras, Alexandros, Ntoutsis, Alexopoulos and Skourlas 2006). Sela (Sela, Lavie, Inbar, Oppenheim and Meyer 2015) examine users' interests in various news topics measuring the subjective satisfaction of news editions along with objective measures to infer actual interest in news items. Results suggest user interest is weakly correlated with reading duration, article length and reading order with explicit measures predicting interest in clearly defined topics.

There is clearly a tradeoff between the quantity and quality when comparing implicit feedback with explicit feedback. In (Nichols 1998), authors evaluate the costs and benefits of using implicit feedback indicators over explicit feedback indicators. The results suggest the implicit ratings can be combined with existing explicit ratings to form a hybrid system to predict user satisfaction. In (Jawaheer, Szomszor and Kostkova 2010), authors show the implicit and explicit positive feedback complement each other with similar performances. Similarly, comparison of the implicit and explicit feedback in use of the Curious Browser reveals the time spent on a page, amount of scrolling on a page and the combination of time and scrolling had a strong correlation with the explicit feedback. This implies the systems can be designed to use the correlation between implicit and explicit feedback to tune the interest modeling algorithms based on implicit feedback. The WAIR system (Zhang and Seo 2001) learns the user interest by observing user interactions and then training on the explicit feedback data. After this learning phase, the system can estimate the relevance feedback implicitly based on the learned observations. The learned information is used to create a user profile and this profile is used in generating queries for retrieval process. In (Liu, Xiang, Zhao and Yang 2010), implicit and explicit feedback indicators are unified using a matrix factorization model (called Co-rating) that can effectively cope with the heterogeneity between these two forms of feedback. Similarly, in (Wang, Rahimi, Zhou and Wang 2012), a unification model based on matrix factorization called expectation-maximization collaborative filtering (EMCF) is introduced.

3.7 Distributed User Modeling

Multi-application systems provide opportunities to gather user data from outside of the individual application itself. Aggregated user data may be useful to address the cold-start problem as well as the sparseness of user data. Connecting data from different sources and services from distributed application environments is in line with advancements in multi-core and multi-tasking architectures. While there are theoretical and software frameworks for distributed user modeling, assessments of modeling techniques are almost always reported in terms of single applications. With a better understanding of the user interests, adaptive systems can provide better personalization. Sharing and reusing the user model information between applications can bring the advantage for profile providers as well as profile consumers by enriching the user models.

Current systems that provide personalized services to users are mostly develop their own proprietary application environments in ad-hoc manner as a part of a specific application requirement (Dim and Kuflik 2012). These proprietary user models are of evidence in system developer's focus on specific characters of their users in order to provide a specific service (e.g., movie recommender system). Over the years, these user models and their application environments are moved from providing complete, monolithic solutions in user modeling servers (Kobsa 2007) to dynamic solutions in the areas of interoperability and interlinking (Leonardi, Abel, Heckmann, Herder, Hidders et al. 2010, Carmagnola, Cena and Gena 2011). User models can be developed by adapting the content consumed or produced by the user, and their specific task, background, history and information needs (Renda and Straccia 2005). These models can bring users' attention to valuable content via personalized presentations. (Berkovsky, Kuflik and Ricci 2008) presented a definition of mediation to introduce cross-system personalization using the technique to integrate and match user modeling data. Recognizing the user interest based on observed user activity is confounded by idiosyncratic work practices. As a result, systems that aggregate evidence of user interest from a wide variety of sources are more likely to build a robust user interest model.

There are two main approaches to user modeling in a component-based architecture. These vary based on the degree of centralization of the user models. Decentralized (or distributed) user modeling had its roots in agent-based architectures; here fragments of user model are kept and maintained by each independent application. Another important distinction among user modeling approaches is whether the model is represented via features or content. Feature-based user models define a set of feature-value pairs representing various aspects of the user, such as interest in a specific category or a level of knowledge in a specific area. Content-based approaches take into account the user's area of interest, as an example, the textual content of documents the user has previously indicated as relevant. These systems generate recommendations by learning user needs with the analysis of available rated content.

In a centralized approach, the integrated user model is stored in a central server and the model is then shared across several user-adaptive applications. Apart from alleviating the applications re-inventing the wheel, centralized user model give an opportunity to share the same user model between several applications. These include generic user modeling servers such as IPM (Bae, Kim, Meintanis, Moore, Zacchi et al. 2010), CUMULATE (Brusilovsky, Sosnovsky and Shcherbinina 2005, Yudelson, Brusilovsky and Zadorozhny 2007), UMS(Kobsa and Fink 2006) and PersonisAD (Assad, Carmichael, Kay and Kummerfeld 2007) as well as framework developed for mashing up profile information (Abel, Baumgartner, Brooks, Enzi, Gottlob et al. 2005, Abel, Henze, Krause and Plappert 2008, Abel, Heckmann, Herder, Hidders, Krause et al. 2009, Houben, Leonardi and Van Der Slujis 2009) to facilitate aggregated user data.

PersonisAD is a distributed framework for building ubiquitous computing applications. It defines a user model based on data gathered from different sensors and combines their preferences using resolvers to provide a tailored experience. CUMULATE is a generic modeling server developed for a distributed E-Learning architecture to help students select the most relevant self-assessment quizzes by inferring their knowledge of a predefined set of topics based on authored relationships among activities in the educational applications and topics. UMS is a user modeling server based on the LDAP protocol which allows for the representation of user interests using a predefined taxonomy for the application domain.

Attempts to bridge user models in various systems require conversion of the user models data between various applications, domains and adhering to semantic representations (Martinez-Villaseñor, Gonzalez-Mendoza and Hernandez-Gress 2012). Some of these have been done using mapping techniques of user models (Vassileva, McCalla and Greer 2003, Bennani, Chevalier, Egyed-Zsigmond, Hubert and Viviani 2012) and more recently using machine learning methods (Berkovsky, Kuflik and Ricci 2008). These user modeling systems do not easily comply with a standard format, technique or vocabulary to enable user modeling interoperability (Martinez-Villaseñor, Gonzalez-Mendoza and Hernandez-Gress 2012).

This dissertation is based on the immediate need for approaches to setup user interests and the distinctions between them to be constructed based on the content encountered rather than pre-agreed upon by the contributing applications.

4. SYSTEM ARCHITECTURE*

4.1 Interest Profile Manager

The Interest Profile Manager (IPM) is user profile server (see Figure 3) to support the personalized delivery of content across multiple applications. The IPM collects user activity across many applications and infers user interests using this implicit and semi-explicit interest information. It also shares the inferred user interests with registered applications that ask for it. The IPM can easily communicate with any application that can be modified to include the interest profile client software component enabling user interest modeling capability in existing applications.

We have used the Mozilla-Firefox web browser and Visual Knowledge Builder (VKB)(Shipman, Hsieh, Maloor and Moore 2001) applications to present search results and also to visualize recommendations. The three other applications provide additional activity data but do not include visualizations: PDFPad which is an acrobat add-on; IPCWord which is a Microsoft Word add-on; IPCPowerPoint which is a Microsoft PowerPoint add-on. Records of user activity in PDFPad, Mozilla, MS Word and MS PowerPoint are stored in the IPM and drive the visualizations that the IPM generates for each of the application registered for relevant notification request (Jayarathna, Patra and Shipman 2015). For our implementation, we utilize VKB to act as an overview application

---

* Jayarathna, S., Patra, A., and Shipman, F. "Unified Relevance Feedback for Multi-Application User Interest Modeling," Proceedings of the 15th ACM/IEEE-Cs Joint Conference on Digital Libraries, pp.129-138, © 2015 Association for Computing Machinery, Inc. Reprinted by permission. http://doi.acm.org/10.1145/2756406.2756914

for a web search (see Figure 4). An interest profile is made up of the aggregated heterogeneous interest evidence collected from these different IPM clients.

The IPM defines the XML communication interface so that application clients can interact with IPM over TCP/IP. The IPM framework includes two modules involved in estimating the user interest, the Estimation Manager and the Estimation module which is again decomposed to 3 sub-modules: Multi-Application Weighting module, Implicit Feedback Module and Semi-Explicit Feedback Module.

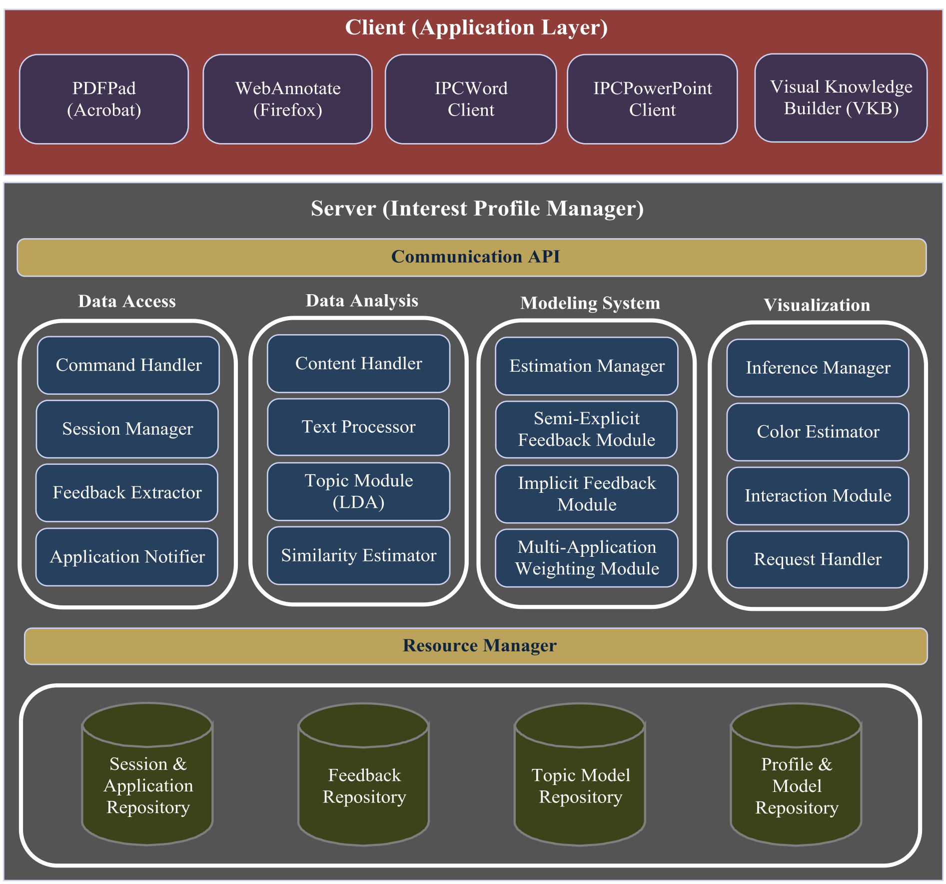

Figure 3: Interest Profile Manager Architecture

The Estimation Manger provides a generic high level interface to the other modules within the IPM and also enables multiple modules to estimate the user's interests using different algorithms. In the Multi-Application Weighting module (see section 4.5.2 for discussion on multi-application weighting), each application is assigned a weight based on the particular user's activities in the various applications. These learned weights are used to merge the estimated interests from the different applications when modeling the overall user interest. The implicit and semi-explicit relevance modules handle the implicit and semi-explicit relevance feedback indicators respectively. The combined outputs from these two sub-modules are used to estimate the final unified user interests for a search task.

The Resource Manager communicates with data repository to update the user interests according to the user activity data sent from application clients. The Data Repository also saves session data both in terms of contextual and temporal features so that the user activity can be defined as a group of search tasks related to each other in order to make inferences about evolving information needs. This is particularly important because if we are able to accurately identify changes to the users' information seeking intent, then we will be in a better position to limit the application of particular inferences about user interests (Jones and Klinkner 2008). The Data Repository also saves both types of feedback data and application data received from application clients for further processing at the estimation modules.

4.2 Interest Representation

Although each application has unique information that may be used to gauge human interest, this interest assessment needs to be sharable among the different applications to be useful in building the complete interest model of a user.

The IPM depends on an abstract XML representation for receiving interest-related information from applications and for broadcasting inferred interest to client applications. Because we realize that we cannot foresee all of the ways different applications will allow users to interact with documents, the representation is extremely general and extensible. Thus an interest profile consists of a document identifier, an application identifier, and a list of application-specific attribute/value pairs. In this way, new applications only have to inform the IPM of the attributes and how they demonstrate user interest when registering.

While some of these applications support two-way communication, this is not required (see Figure 5); an application could merely provide information to the IPM or only receive interest information from the IPM. In the current architecture, VKB, PDFPad and WebAnnotate support two-way communications while Microsoft Word and PowerPoint support one-way communication. Applications also can be categorized into (i) *Consumption Applications*, for examining/annotating existing content; and (ii) *Production Applications*, for creating/authoring content (see Figure 6).

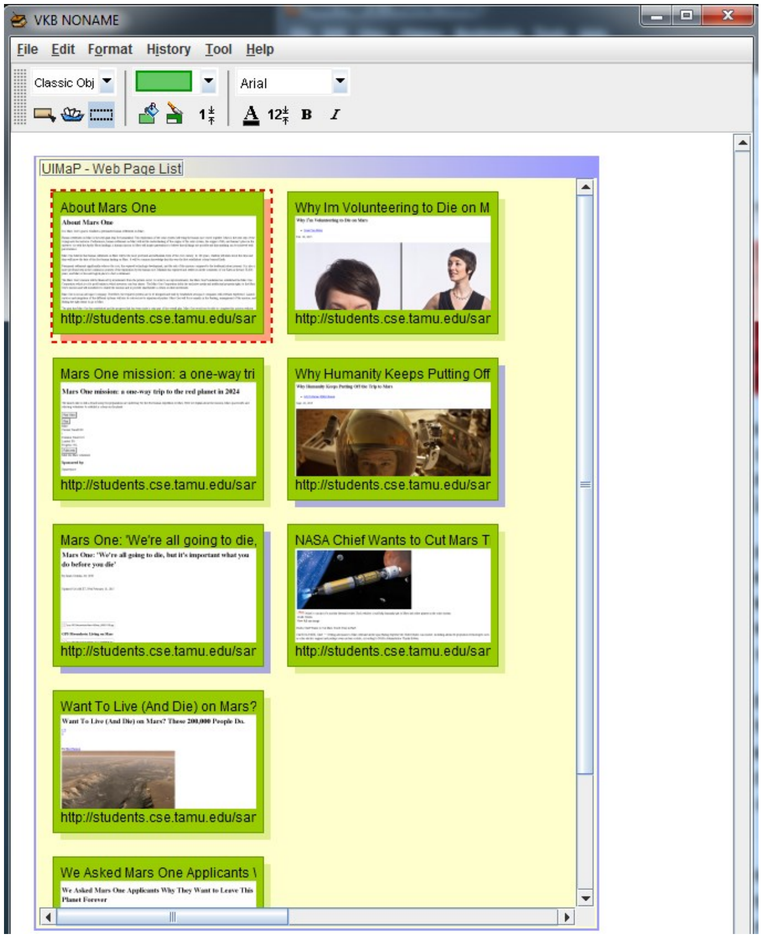

Figure 4: VKB Search List with Visualizations

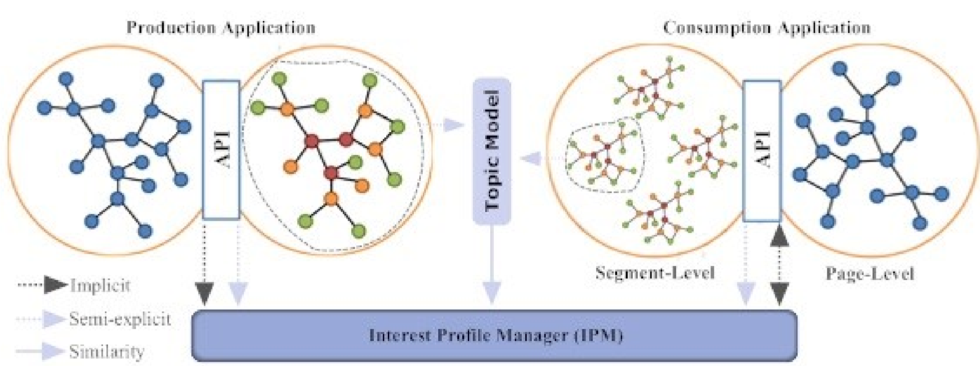

Figure 5: Everyday Applications System Architecture

4.3 Explicit Feedback

Whenever a document is opened in Microsoft Word or PowerPoint, event handlers are registered for user events. Event handlers save each interaction and their values locally and send them to the IPM. Additionally, the content of the document and document characteristics are sent to the IPM at the time of closing the document.

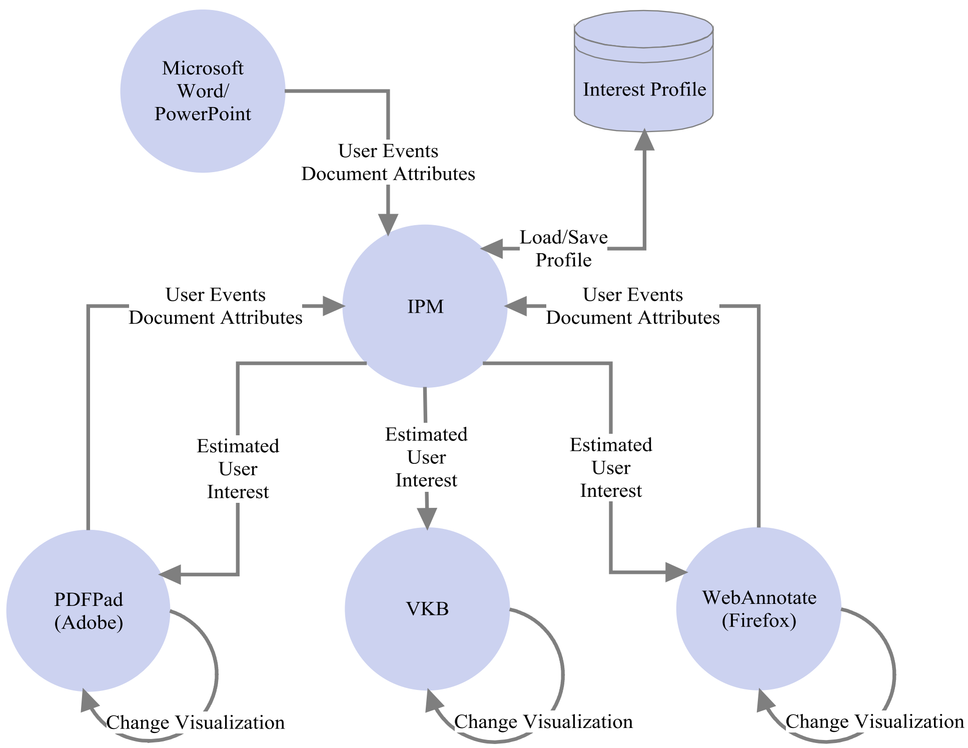

Figure 6: IPM Event Transition from Individual Applications

Similarly, WebAnnotate parses raw text to identify every paragraph when a new web page is opened. It also appends mouse and keyboard events in a buffer and saves the color and relevance score assigned to each annotation until the browser is moved to the background. All the raw information is sent to the IPM in an XML format at the next focus out event or web page close event. The buffer is reset once the focus is brought back to the web page.

During an information gathering activity, useful documents may be long and cover multiple subtopics; users may read some segments and ignore others. The browser plug-in WebAnnotate (Bae, Kim, Meintanis, Moore, Zacchi et al. 2010) enables basic annotation capabilities so that users can make persistent annotations on web pages and passages and get suggestions within these documents based on estimated user interests. The interest classes can be defined based on annotations' color, type and content in WebAnnotate. To identify segments of new or unread documents to bring to the user's attention, these classes are then compared against the segments of the document currently displayed in WebAnnotate generated by the text-tiling algorithm. When a match is identified, an underline (based on the intensity of the inferred interest value) of the appropriate color for the class is used to signal the similarity. In Figure 7 the user has opened the Wikipedia page for the Human Genome Project and highlighted text related to the history of the project. It can be seen that other paragraphs are underlined with the same color indicating that they are similar to the passage highlighted.

In the current study, WebAnnotate was extended to include three types of explicit ratings for content: "page relevance", "page familiarity", and "paragraph relevance" on a 5-point scale. After each paragraph annotation WebAnnotate allows the user to mark individual paragraphs as relevant or not to their task (see Figure 8).

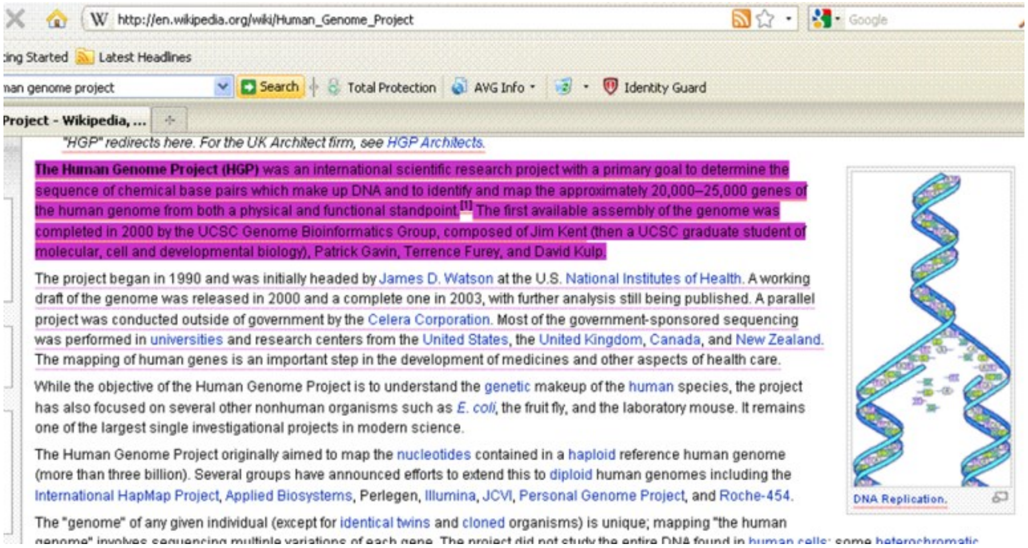

Figure 7: WebAnnotate User Highlights and System Recommendations

A user might also use Microsoft Word or PowerPoint applications to open, read or modify some documents. The user's actions while working on these applications can also be used to infer user's interests. MS Word and PowerPoint consider all the data in one document to belong to a single interest class. The default color (in the current research study 'Blue' color) of the application is used to define the interest class.

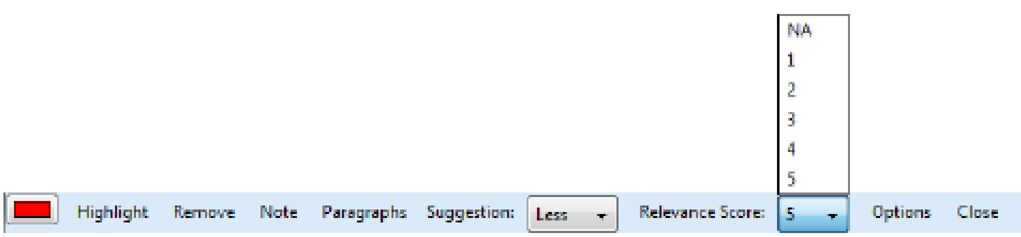

Figure 8: WebAnnotate Toolbar for Rating Paragraphs

4.4 Implicit Feedback

We utilize a set of the implicit feedback indicators during a document reading activity to characterize the interactions between the user and documents. These document reading activities include user actions during a passive reading in a consumption application (web browser or PDF reader). This consists of time spent in a document, number of mouse clicks, number of text selections, number of document accesses and characteristics of user scrolling behaviors such as number of scrolls, scrolling direction changes, time spent scrolling, scroll offset, and total number of scroll groups. Furthermore, we collect time spent on a production application (MS Word or PowerPoint), focus in/out and other formatting activities. Table 1 summarizes the user events and document attributes collected from both production and consumption applications during this research study.

The interest profile broadly contains three types of interest indicators, characteristics of the user, the document as a whole, and the textual content of the document (see Table 1). The user features are derived from implicit feedback data. All these features vary from one user to another as they heavily depend on the individual practices. Document features are high level features of the documents that are the same across users. Finally, document text features are generated from the user's annotations in consumption applications and from the user's authored content from production applications. Document text content provides evidence of more focused interest than the general document features. Such evidence is important when identifying the specific parts of documents that are expected to be relevant.

Table 1: Interest Indicators from Applications

Interest Category	Microsoft Word/PowerPoint	Browser (Firefox)
User characteristics (Implicit Feedback)	Click, double click, right click, focus in/out, total Time, edit time, idle time, away time	Click, double click, right click, focus out, total Time, reading time, away time, number of scrolls, number of scrolling direction changes
Document characteristics (Fixed Features)	Size, number of characters, images, links, last access time, number of slides, text boxes	Images, links, document relevance and familiarity score (explicit)
Textual characteristics (Semi-Explicit Feedback)	Text Authored	Text Annotated

Another type of feature important in this work is content similarity. Content similarity metrics are used to measure the overlap between the textual content of the user's previous interactions and any future text content. These similarities are computed between text considered valuable to the user (authored or annotated text) and all other paragraphs displayed in the browser and documents available in other applications. The similarity score represents the user's interest expressed through the textual content. In this work, Latent Dirichlet Allocation (LDA) is used to compute the content similarity (see section 4.5.1) using the Hellinger Distance measure and are then normalized to be between [0-1] using min-max normalization.

4.5 Models of User Interest

The IPM uses the document attributes (e.g. metadata, term vectors, user-assigned color of annotations) to determine classes of user interest. Attributes of the document as a whole and textual characteristic of document segments are selected based on evidence of interest in individual documents. To aid in the creation of descriptions of document classes, the IPM includes term vector and metadata analysis capabilities as well as text tiling capabilities to allow clients and the IPM to analyze text at the sub-document level. Currently, user-assigned annotation color is used to identify the known members of an interest class while the identification of documents and document components similar to that class is based on the other document attributes and user characteristics.

The next subsections describe the use of topic modeling for similarity assessments of textual content in the user model or of potential value to the user, the weighting of features across the different applications, and the development of semi-explicit and unified feedback models.

4.5.1 Topic Modeling of Textual Content

Before introducing our topic modeling approach for inferring user interests, we first give a brief review of the statistical model Latent Dirichlet Allocation (LDA) and its parameters used in this research study. LDA (Blei, Ng and Jordan 2003) is a hierarchical Bayesian model that assumes each document is a finite mixture of a set of topics K and each topic is an infinite mixture over a set of topic probabilities. Unlike clustering methods, LDA does not assume that each document can only be assigned to one topic. Given a document collection, we use LDA to find a set of topics discussed in the document collection. Each topic is represented as a set of words that have a higher probability than others to appear in the text unit related to the topic. Based on the probability distribution of words in each topic, we can calculate the probability that each document may contain a topic and obtain a document-topic assignment.

We set LDA parameters; a number of topics $K = 5$ to match the number of topic clusters anticipated, two smoothing parameters $\alpha = 0.01$ and $\beta = 0.01$ (McCallum 2002). As words are the only observable variables in an LDA model, conditional independence holds true for the outputs of LDA model which are document-topic and topic-words distributions Φ and Θ.

For a corpus containing D documents (see Figure 9), the parameters, the $D \times K$ matrix of document-topic probability distribution per each document and the $K \times W$ matrix of topic-words probability distribution per each topic must be learned from the data. Parameter fitting is performed using collapsed Gibbs sampling (Porteous, Newman, Ihler, Asuncion, Smyth et al. 2008) with sampling and burn-in iterations set to 1 and 5 respectively. We look at the difference in the content from two text units by first computing the LDA document-topic distributions Φ_i and Φ_j $(i, j = 1..K, i \neq j)$ and then by calculating the divergence between these two document-topic distributions. The smaller the divergence is, the stronger the associated similarity is.

We performed an evaluation to determine the feasibility of topic modeling divergence methods in our context and to select among alternative topic modeling approaches. Based on those results, we use Hellinger distance (Bishop 2007) to compare the similarity between document-topic distributions (Equation 1).

$$D_{LDA+H}(\Phi_i || \Phi_j) = \sqrt{\frac{1}{2} \sum_{i,j=1}^{K} \left(\sqrt{\Phi_i} - \sqrt{\Phi_j}\right)^2} \quad \textbf{(1)}$$

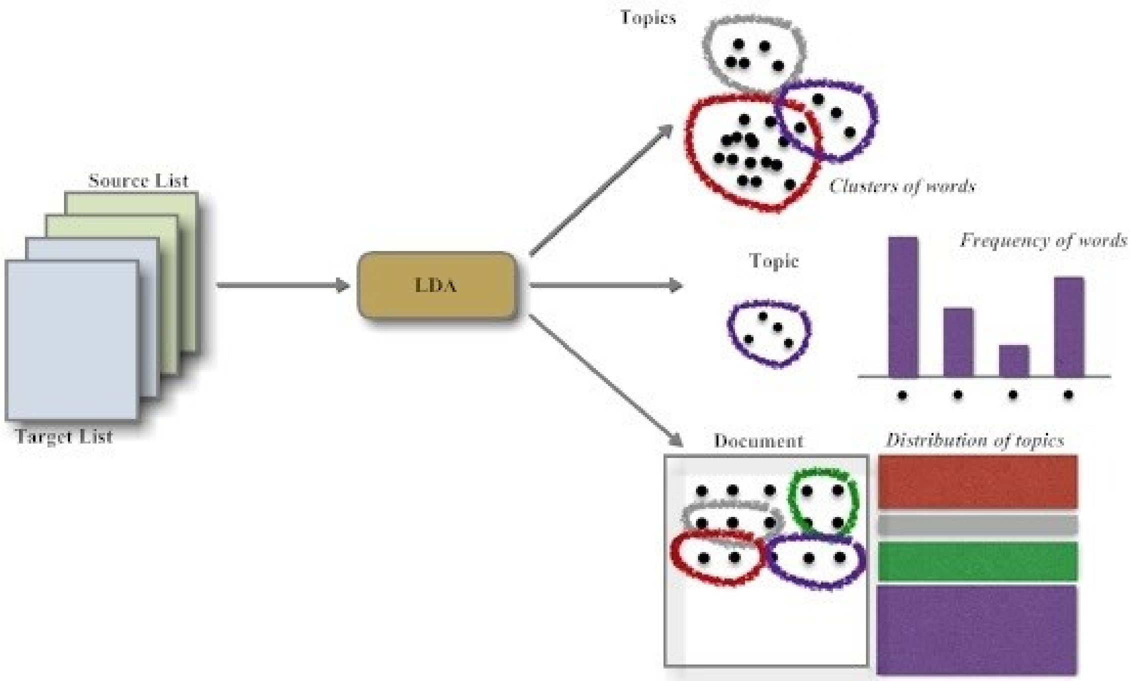

Figure 9: Semi-Explicit Topic Modeling of Text Content

4.5.2 Multi-Application Weighting

Once we have user, document, and textual characteristics as well as textual similarity measures, we need to weight the various features to predict the likelihood of interest in the target. Rather than using one set of weights for all users, we train the interest model using weighted K nearest neighbor (WKNN). This enables weights to adapt to the user-specific patterns present in the feature space. The weights for the features result in a classifier algorithm that predicts relevance the score for each paragraph on a 5-point scale. From here onwards, we denote *C* as the relevance label.

In this work, we have combined two variants of KNN, i.e., attribute-weighted and distance-weighted KNN to a build our weighted KNN classifier. By introducing a feature weight component in the distance metric (Equation 2), the quality of the feature is also considered in addition to the difference in value of the feature. Thus, more useful features are given more weight while the less useful features have less weight in the ultimate distance measurement. As a result, useful features have greater impact on the distance function compared to irrelevant features.

$$d(x, y)_w = \sqrt{\sum_{j=1}^{d} w_{cj}^2 (y^j - x^j)^2} \quad \textbf{(2)}$$

where c = class(x), $x \in F$, w_{cj} = weight of feature j belonging to class c.

Since we intend to learn the individual importance of each feature corresponding to each class, we have implemented a normalized version of the class dependent RELIEF algorithm, NCW-R (Marchiori 2013). All the feature weight vector values are initialized to zero and updated iteratively by processing each data point *x* in *X* as per Equation 3.

$$w_c = \sum_{x \in X_c} \left\{ \sum_{z \in WKNN(x,c)} -|x - z| + \sum_{\substack{z \in WKNN(x,\acute{c}) \\ \acute{c} \notin c}} |x - z| \right\} \Big/ N_c \quad \textbf{(3)}$$

4.5.3 Semi-Explicit Feedback Model

In this section, we first focus on the user interest model based on semi-explicit and implicit relevance feedback. For the semi-explicit model, we use baseline-LDA to infer content similarity and use it in the user interest estimation to determine how likely a page or a segment is of interests to a user.

Suppose at time t, the user has annotated a segment from document d_{ti} whose previous annotations (from same user) are $a_1, \ldots, a_n$. We update our baseline-LDA model by the modified Rocchio algorithm (Rocchio 1971, Shen, Tan and Zhai 2005) computing the centroid vector of all annotations created by the user for the given task and interpolating it with the previous source document vector to obtain an updated term vector (Equation 4). In this context we define the set of annotations as the combination of the relevant user annotations from the browser and the produced text from content producer applications (MS Word or PowerPoint).

$$\vec{Q}_t = \lambda \vec{Q}_{t-1} + (1 - \lambda) \frac{1}{n} \sum_{i=1}^{n} \vec{a}_i \qquad \textbf{(4)}$$

where $\vec{Q}_{t-1}$ is the previous source vector, n is the number of annotations the user created immediately following the current annotation, and λ is the parameter that controls the influence of the annotations on the inferred user model. In our experiments, λ is set to 0.5.

4.5.4 Unified Relevance Feedback Model

Previous work (Liu, Xiang, Zhao and Yang 2010, Wang, Rahimi, Zhou and Wang 2012) shows that implicit relevance feedback alone is not adequate to estimate the interest of a user during document interactions in some situations. The results suggested that the implicit ratings can be combined with existing explicit relevance data to form a hybrid system to predict user interest.

For a target document d_{ti}, we define a scalar valued interest prediction from the observations of user behavior as,

$$r_i = \mu R_E(i) + (1-\mu) R_I(i), \quad 0 \le R_E(i) \le 1, \quad 0 \le R_I(i) \le 1 \tag{5}$$

where $R_E(i)$ is the similarity score estimated from semi-explicit feedback model, $R_I(i)$ is an implicit feedback estimated from the following equation, and $\mu = 0.8$ is a heuristically tuned scaling factor representing the relative importance of the implicit feedback. We calculate $R_I(i)$ from,

$$R_I(i) = \sum_{j \in F} w_j f_j(i) \tag{6}$$

where w_j is the weight for each feature j of the implicit feedback generated from WKNN. All the features were normalized to zero mean and unit variance.

4.6 Dynamic IPM Architecture

With the lessons learned from initial static implementations of the system, in next sections, we describes updates to the IPM to support dynamic user interest modeling along with added functionality from multiple everyday applications.

4.6.1 Semi-Explicit Model

Figure 10 presents the scenario where, the dynamic system architecture is handling semi-explicit user activities (authored-text and/or annotated text). In the current application environment (See Figure 11), the WebAnnotate tool from the Firefox web browser and the PDFPad annotation tool for the Adobe Acrobat Writer support creation of user annotations. Similarly, Microsoft Word and PowerPoint support creation of authored-text (IPC stands for Inter-process communication). Each time a user creates an authored-text or annotation, this information is propagated to the IPM via IPC through XML data packets.

Each annotation from a webpage or pdf document is considered a source segment and added to the Source List in the IPM Text Processor module. We apply ***Dice's coefficient*** measures to find how similar a source segment and the segments from the current Source List. Dice's coefficient is used to measure how similar two strings are in terms of the number of common bigrams (a pair of adjacent letters in the string).

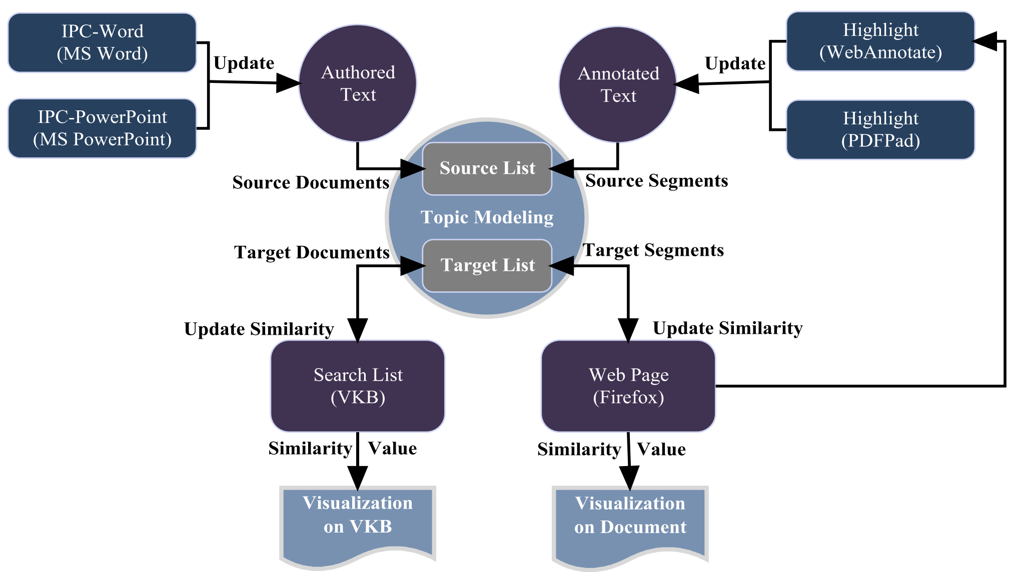

Figure 10: Semi-Explicit Relevance Feedback System Architecture

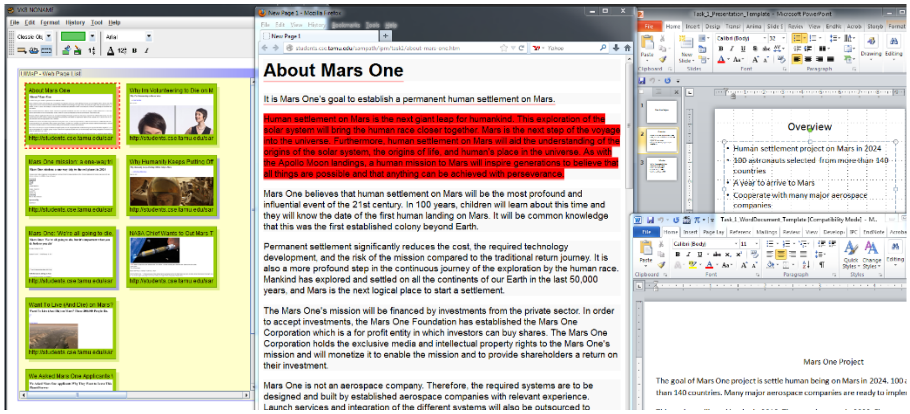

Figure 11: Semi-Explicit model Applications

If the similarity of the segments is over the similarity-threshold (value of 0.5 is empirically selected in this research study), then the two segments are merged. If the current source segment from the application is below the similarity-threshold (compared to all the segments from Source List), a new source segment inside the Source List will be created.

We apply similar calculations in order to insert source documents as well as target documents and target segments to appropriate lists. We find that calculating the similarity using Dice's coefficient is computationally inexpensive compared a more sophisticated topic modeling approach such as LDA.

4.6.2 Unified model

The Figure 12 scenario presents the situation where there is only implicit relevance feedback from user interactions available for user interest model construction. For example, say the user is currently interacting with a web page in the Mozilla Firefox browser retrieved from VKB Search List and the relevance feedback data are generated from the web page through the WebAnnotate tool. An implicit relevance feedback record (sliding-window) for the current user is retrieved every 10 seconds (or whenever user focus-out from the browser application) and sent back to IPM implicit relevance feedback module. This current sliding-window record is aggregated with the user profile and running-interaction event record for implicit rating calculation.

All the individual interaction event instances (each instance is a set of feature values) for the currently active web document are weighted by default feature weight values. Next these weighted feature values are normalized via Weka normalization method and used in equation 6 for implicit rating calculations. The previous interaction-record (page implicit rating) is now updated with the current implicit rating value. All the rated implicit interaction instances are forwarded to the weight-learning module for the feature weight learning process.

Next we calculate the rating similarity for the rest of the VKB search list. We define a ***Proxy similarity*** for each VKB search list web document; which is the Dice's Coefficient similarity between currently active web documents in the Firefox browser. We propagate learned similarity for the current active web document by calculating the VKB similarity using following equation,

$$R_{VKB}(j) = R_{Proxy}(j) \times R_I(i) \tag{7}$$

where, $R_{VKB}(j)$ is the similarity value for document j in VKB search list, and $R_{Proxy}(j)$ is the proxy similarity between document j and i. $R_I(i)$ is the implicit similarity value of the currently active web page i.

When both semi-explicit and implicit relevance feedback is available, we update the previous equation 5 to support the proxy similarity of the VKB search list web pages and to calculate the new similarity value,

$$R_{VKB}(j) = \mu R_E(i) + (1-\mu)\left[R_{Proxy}(j) \times R_I(i)\right], \quad 0 \le R_E(i) \le 1, \quad 0 \le R_I(i) \le 1 \tag{8}$$

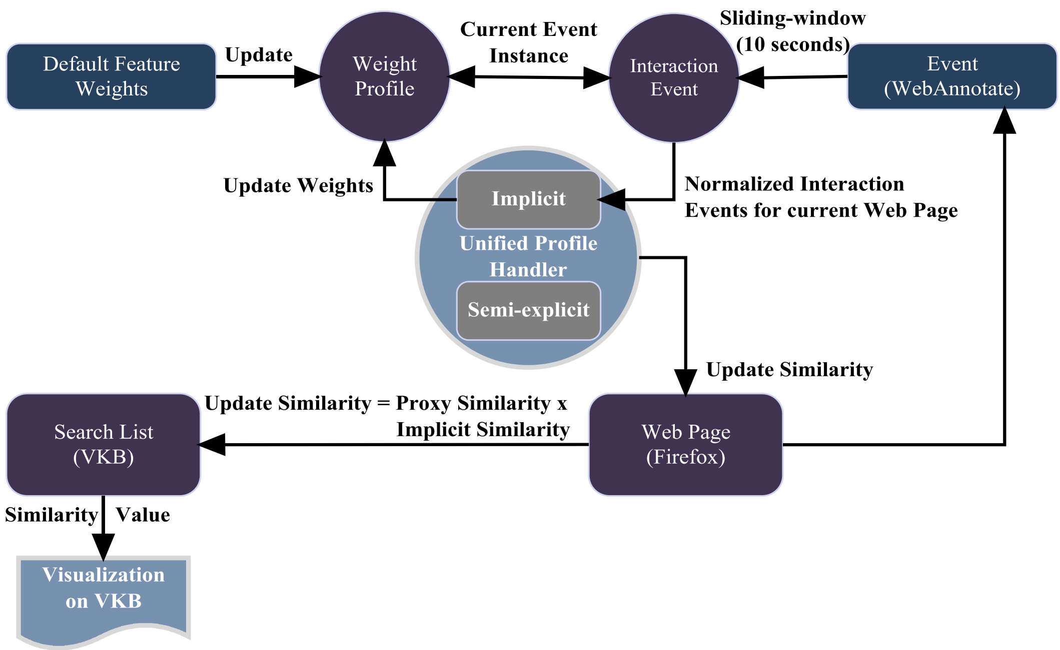

Figure 12: Implicit Relevance Feedback System Architecture

where $R_E(i)$ is the similarity score estimated from semi-explicit feedback model, $R_I(i)$ is an implicit feedback estimated from the equation 6, and $\mu = 0.8$ is a heuristically tuned scaling factor representing the relative importance of the implicit feedback.

5. USER EVALUATIONS AND RESULTS*

Three evaluations were performed to answer research questions aimed at improving the design and implementation of multi-application interest modeling techniques integrating implicit and semi-explicit feedback and assessing their performance in supporting human information activities. The first study (Section 5.1) explores how to reduce the problem of sparsity of content that occurs due to limited textual content being examined and assessed. This is a problem that is common early in information tasks (e.g. the cold start problem for user modeling) and limits the usefulness of content-based models due to alternative vocabularies in different documents. The first study explores how alternative topic modeling approaches affect the interest model's ability to accurately assess document relevance. The second study (Section 5.2) examines the relative value of features associated with users' past activity with content across multiple applications in predicting user assessment of that content. This study provides insight into which features from which types of applications are most valuable in including in a user modeling system. It also provides data useful for comparing the potential increase in performance of models using semi-explicit and implicit feedback. Finally, the third study (Section 5.3) compares the performance of multi-application user modeling approaches using semi-explicit feedback and unified feedback in real user tasks. The third

---

* Jayarathna, S., Patra, A., and Shipman, F. "Unified Relevance Feedback for Multi-Application User Interest Modeling," Proceedings of the 15th ACM/IEEE-Cs Joint Conference on Digital Libraries, pp.129-138, © 2015 Association for Computing Machinery, Inc. Reprinted by permission. http://doi.acm.org/10.1145/2756406.2756914

study also provides data informing the potential for such tasks to be improved via by learning personalized or task-specific weights for user activity features.

5.1 User Study 1 - Topic Modeling (2013)

In this section we first discuss user experiments we have done to evaluate our proposed topic modeling approach. We evaluated alternative topic modeling approaches within our context to determine how well they would work with the type of data available (a small collection of small and large segments of annotated or authored text).

To assess the quality of the topic modeling alternatives, we used each of the user-selected text segments to predict the remainder of that user's selections based on the similarity metrics. We first describe our evolution metrics, and then experimental setup.

5.1.1 Similarity Metrics

We applied LDA to compute the probability distributions of topics for two or more selections of textual content. We then used three distance measures of the divergence between these probability distributions and compared those assessments to the user-provided assessments and Top-N distance measure. The three distance measures are: the Hellinger Distance (H), the Kullaback-Leibler divergence (KL), and the Jensen-Shannon divergence (JSD). In addition, we also evaluated the performance of a Non-negative Matrix Factorization (NMF) model to the three LDA-based techniques.

In our experiments with LDA models, we will create similarity matrices to compare the user-generated annotations (Source S) to document content (Target T); hence we define proposed measures as similarities. The following four measures and NMF have been evaluated in our experiments.

5.1.2 Similarity Models

LDA + Hellinger Distance: The Hellinger distance is computed over two positive vectors Since we are dealing with probability distributions in document-topic distribution, we chose Hellinger distance (Rao 1995) to measure their divergence. The main idea of our approach is to use the Hellinger distance between document topic distributions to find the similarity of target T to the user generated source S.

$$D_{LDA+H}(S||T) = \sqrt{\frac{1}{2}\sum_{i=1}^{K}\left(\sqrt{s_i} - \sqrt{t_i}\right)^2} \quad \textbf{(9)}$$

where S is a K-dimensional multinomial topic distribution and s_i is the probability of the i^{th} topic.

LDA + Kullaback-Leibler Divergence: KL divergence is a non-symmetric measure of the difference between two probability distributions. In our LDA+KL model, the association of source and target in the document topic distribution can be measured using the KL-divergence. The smaller the score is, the stronger the associated similarity is. For two probability distributions, from target to the user generated source, KL divergence is calculated as follows:

$$D_{LDA+KL}(S||T) = \sum_{i=1}^{K} s_i \log_2 \frac{s_i}{t_i} \quad \textbf{(10)}$$

LDA + Jensen-Shannon Divergence: We use Jensen-Shannon divergence (JSD) measure as a smoothed and symmetric alternative to the KL divergence. The measure is 0 only for identical distributions and approaches infinity as the two differ more and more. Formally it is defined as the average of the KL divergence of each distribution to the average of the two distributions (Hall, Jurafsky and Manning 2008).

$$D_{LDA+JSD}(S||T) = \frac{1}{2}D_{KL}(S||R) + \frac{1}{2}D_{KL}(T||R) \quad \textbf{(11)}$$

$$R = \tfrac{1}{2}(S + T)$$

Non-Negative Matrix Factorization: NMF is the task of approximating the matrix $X \in \mathbb{R}^{\geq 0, m \times n}$ by the product of two reduced-dimensional matrices $W \in \mathbb{R}^{\geq 0, m \times k}$ and $H \in \mathbb{R}^{\geq 0, k \times n}$ so that $X \approx WH^T$. Dimensions of W and H are $m \times k$ and $k \times n$ respectively, where k is the select number of topics for $0 < k \ll \min(m, n)$ (Smaragdis and Brown 2003). Then, the minimization problem can be stated as,

$$\min_{s.t\ W \geq 0, H \geq 0} f(W, H) := \|X - W \cdot H\|_F^2 \quad \textbf{(12)}$$

where $\|\cdot\|_F$ is the Frobenius norm. We note that other objective functions can be used to measure the error of the approximation instead of the Frobenius norm, but it is the most appropriate when errors are normally distributed (Gonzalez and Zhang 2005).

The H is initialized to zero and W to some randomly generated matrix where each $W_{ij} > 0$ and these initial estimates are updated with alternating iterations of NMF multiplicative update rules (Lee and Seung 1999). The NMF algorithm successively updates H and W which fixing the other, by taking a step in weighted negative gradient direction for the $f(W, H)$.

$$W_{ij} \leftarrow W_{ij} - \zeta_{ij} \left[\frac{\partial f}{\partial H}\right]_{ij} \equiv W_{ij} + \zeta_{ij}(XH^T - WHH^T)_{ij} \quad \textbf{(13)}$$

$$H_{ij} \leftarrow H_{ij} - \eta_{ij} \left[\frac{\partial f}{\partial H}\right]_{ij} \equiv H_{ij} + \eta_{ij}(W^T X - W^T WH)_{ij} \quad \textbf{(14)}$$

where ζ_{ij} and η_{ij} are individual weights for the corresponding gradient elements with following weight values,

$$\zeta_{ij} = \frac{(W)_{ij}}{(WHH^T)_{ij}}, \qquad \eta_{ij} = \frac{(H)_{ij}}{(W^T WH)_{ij}}$$

Now, we can define the updating formulas:

$$W_{ij} \leftarrow W_{ij} \frac{(XH^T)_{ij}}{(WHH^T)_{ij}} \quad \textbf{(15)}$$

$$H_{ij} \leftarrow H_{ij} \frac{(W^T X)_{ij}}{(W^T WH)_{ij}} \quad \textbf{(16)}$$

5.1.3 Confusion Matrix for Similarity Evaluation

How can we evaluate the effectiveness of our proposed methods? Given that our primary goal is to learn the user's preference from her explicit feedback and use these user generated annotation results to visualize relevant document content, we may consider the standard information retrieval domain evaluation metrics such as precision, recall, accuracy, F1 measure, false positive and true positive.

Table 2: Confusion Matrix for System Evaluation

		User Generated	
		Annotated	**Not-Annotated**
System Generated	**Underlined**	TP	FP
	Not-Underlined	FN	TN

Precision is the ratio of correctly underlined as a class to the total document content as the class. For example, the precision (P) of the underlined class in is $tp/(tp + fp)$. Recall (R) is the ratio of correctly underlined document content as a class to the actual user generated annotations in the class. The recall of the underlined class in the table is $tp/(tp + fn)$. Accuracy is the proportion of the total number of underlines that were correct. The accuracy in the Table 2 is $(tp + tn)/(tp + fp + fn + tn)$. F1 is a measure that trades off precision versus recall. F1 measure of the underlined class is $2PR/(P + R)$.

5.1.4 Ground Truth Data Collection

Since our approaches are based on annotated document contents, we need to collect user's annotations for a set of search tasks. In the meantime, users are required to supply a set of annotations using the WebAnnotate tool that reflects relevance to the main idea of the given search tasks. The data was composed of five search tasks and twenty web documents. Documents were preprocessed and removed graphics and annotations before experiments. We recruited 17 students to annotate the documents relevant to the given search tasks. Users were told to make annotations freely which reflects the main idea of the given task and relevance to the given documents. To compare these approaches, we collected a set of text annotations from the given web documents that indicated relevance to given search tasks. The data was based on 17 participants selecting the relevant paragraphs (text segments) from a set of 20 pre-selected web documents for each of five different information gathering tasks. This resulted in a total of 1267 text segments being selected across the 100 documents.

A number of subcomponents of our approach to unified relevance feedback for multi-application user interest modeling were evaluated. We used data from this ground truth data collection activity that included annotations and post-task relevance assessments to test the feasibility of alternative topic modeling and similarity techniques.

5.1.5 User Study 1 – Results

5.1.6 Topic Modeling Approach Selection

We evaluated alternative topic modeling approaches within our context to determine how well they would work with the type of data available (a collection of small and large segments of annotated or authored text).

We first evaluate the sensitivity to the similarity threshold (between topic-probability distributions of two text units) in the LDA+H, LDA+KL and LDA+JSD. Figure 13 shows how the model threshold influences the performance. As the threshold increases from 0.1 to 0.5, the performance keeps on improving and reaches the average optimal value at 0.45 for all three models. For the experiments beyond this point, we use value of 0.45 as the similarity threshold.

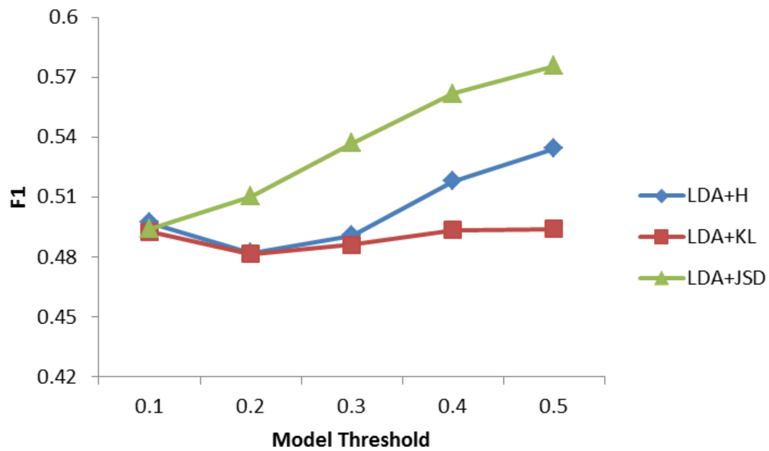

Figure 13: Impact of Varying the Threshold in Topic Models

We next applied LDA to compute the probability distributions of topics for two or more selections of textual content. We then used three distance similarity measures of the divergence between these probability distributions and compared those assessments to the user-provided assessments. The three distance measures are: the Hellinger Distance (H), the Kullaback-Leibler divergence (KL), and the Jensen-Shannon divergence (JSD). In addition, we also evaluated the performance Non-negative Matrix Factorization (NMF) model and TF-IDF with cosine similarity compared to the three LDA-based techniques.

Table 3: Performance Comparison of 5 Similarity Measures

	Precision	Recall	F1	Accuracy
LDA+H	0.944	0.367	0.499	**0.722**
LDA+KL	**0.954**	0.350	0.485	0.719
LDA+JSD	0.736	**0.548**	**0.576**	0.713
NMF	0.814	0.418	0.500	0.692
TF-IDF	0.247	0.396	0.287	0.237

To assess the quality of the topic modeling alternatives, we used each of the user-selected text segments to predict the remainder of that user's selections based on the similarity metrics. When the user-selected paragraph reached a similarity value of 0.5 (experimentally chosen to have reasonable performance) it was assumed to be recommended by the system. When a system-generated recommended by the system was indeed one of that user's other selections, it was counted as a true positive. When a paragraph in the text did not reach that threshold it was counted as a true negative. Table 3 presents the resulting average precision, recall, F-measure and accuracy across the 5 search tasks. This result indicates LDA-based models outperform both classic TF-IDF method as well as stat-of-art NMF method in-terms of Precision, Recall and Accuracy.

We also examined the effect of varying the number of latent topics in the LDA model on performance. Figure 14 shows the overall accuracy of the different distance measure for 5, 10, 15, 20, 25 topics across the 5 information selection tasks. From these results, we first observe that the effect on the final performance is consistent for all three LDA models.

The Figure 15 shows the overall performance of all four algorithms. The improvement on recall and F1 of all three LDA-based models are very significant. This is very encouraging since recall is a more important factor in generating user interest models to provide relevant content as suggestions/recommendations. The results demonstrate that the LDA models consistently outperform the NMF method in terms of hit recall and F1 measure. From this comparison, it can be concluded that the proposed approach is capable of making accurate and effective search suggestions.

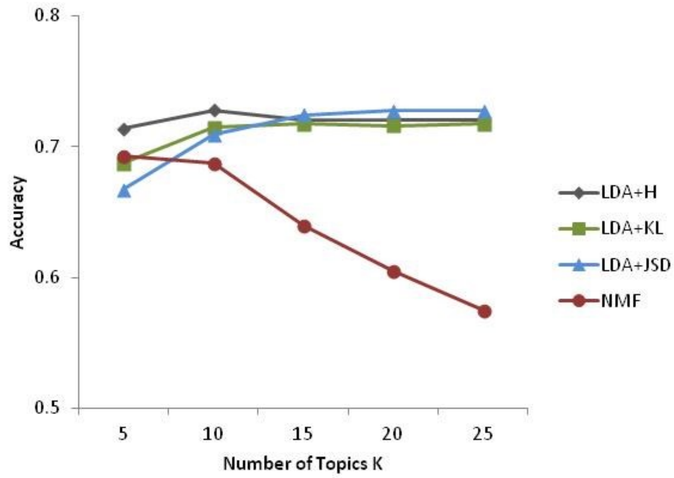

Figure 14: Impact of Varying the Number of Latent Topics

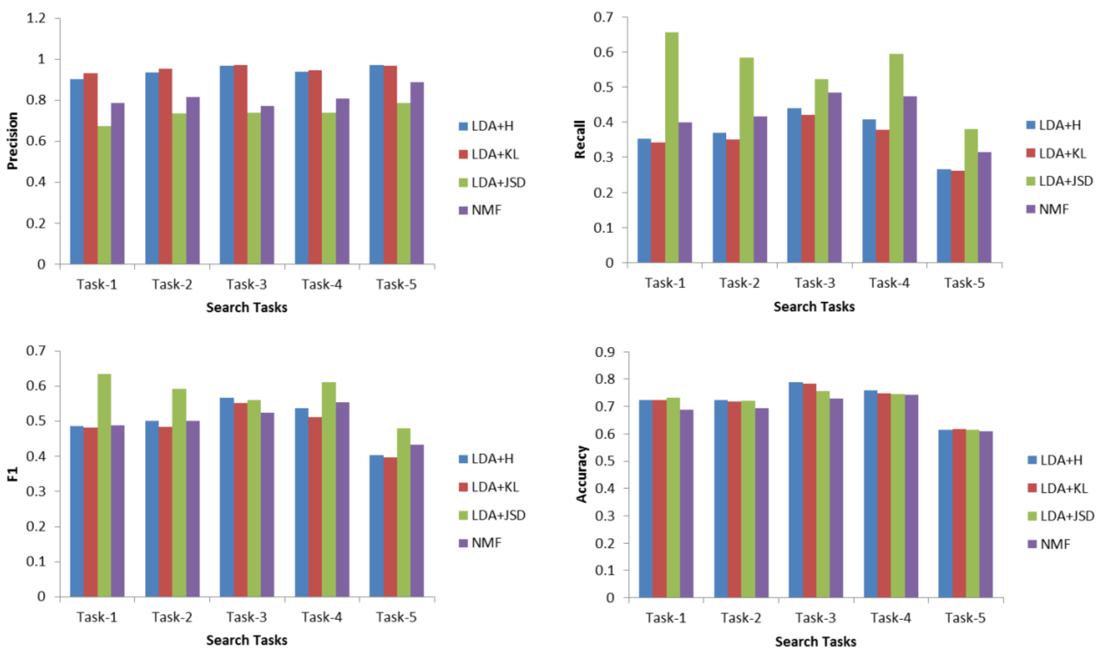

Figure 15: Performance Comparisons of Different Models

5.2 User Study 2 – Unified Model Feasibility Study (2014)

31 undergraduate and graduate students (ages 21 to 40) were recruited to perform a set of four tasks requiring the use of the Firefox web browser with the WebAnnotate extension, Microsoft Word and Microsoft PowerPoint. All participants reported spending at least 1-3 hours daily browsing the Internet. None of the participants had any prior experience with WebAnnotate.

Participants were given the task of writing summaries and generating short slide presentations on topics in four different domains (technology, science, finance, and sports; shown in) based on a set of eight web resources per domain. The instructions suggested that each task would take about 30 minutes, but that they could continue working as long as they needed to.

The resources provided were selected from the top documents returned from a Google query on the topic and were chosen to include pages with varying degrees of relevance to each task. Table 4 includes the average and variance of post-task relevance scores assigned by participants for the documents per task. It shows that each task contained both relevant and non-relevant web pages in similar proportions.

Table 4: Task Topics with Post-Task Document Relevance Assessments

Task No	Task Name	Relevance Score Mean and Variance
1	How does Google Glass work?	3.55 ± 0.96
2	What is mars one project?	3.23 ± 1.11
3	How to improve your credit score?	3.53 ± 0.98
4	What are the rules of American football?	3.52 ± 1.01

User activity data in the three applications and post-task relevance assessments of each document were collected. Activity data collected during the tasks included all the features originally described (in Table 1). Due to experimental setup, this data required preprocessing. For example, as it is expected due to the data collection process, document features such as last access time, creation time, and last write time features are not informative because each individual task lasted approximately 30 minutes. Thus, these features are not considered during the evaluation process. In total, the data captured includes 34 potentially useful features out of 48 features.

In addition to the post-task page level assessments of relevance, each participant was requested to annotate and rate individual segments of documents, so that each segment in a page could be considered as a unique piece of content with the goal of the interest model learning to identify relevant segments in web pages. Pre-processing of the data assumes any segment that was not explicitly annotated and rated by a participant was irrelevant ($C = 1$). At the end of the tasks we conducted a survey about participant's prior knowledge of the applications involved, understanding of tasks and other details. The average score for the question "How comfortable were you doing the tasks" is 4.35 on a scale from 1 to 5 (1 being Lowest & 5 being Highest). This indicates that participants did not have many issues comprehending the topics.

Small segments were also removed from consideration; any segments with less than 10 words are ignored from the data set to avoid noise. We ignored data collected for tasks when participants did not generate the requested document or slides and for participants that did not annotate at least fifty paragraphs across the four tasks. Finally, since the web pages shown to the participants are real web pages and there may be some unwanted segments (comments, page headers) in the content. We removed 6247 such data instances during data filtering stage. Final dataset includes 33212 data instances across 108 tasks available for model evaluation.

We explored the use of Weighted K Nearest Neighbor (WKNN) to assign weights to the various features in our unified model to predict the likelihood of interest. The feature weight values are obtained after averaging 200 iterations of the WKNN classifier. The training data set is generated by randomly selecting 70% data points from the entire data set and the remaining 30% is treated as test data for each iteration. The optimal parameter K=5 for the WKNN is selected based on performance after a 5-fold cross validation.

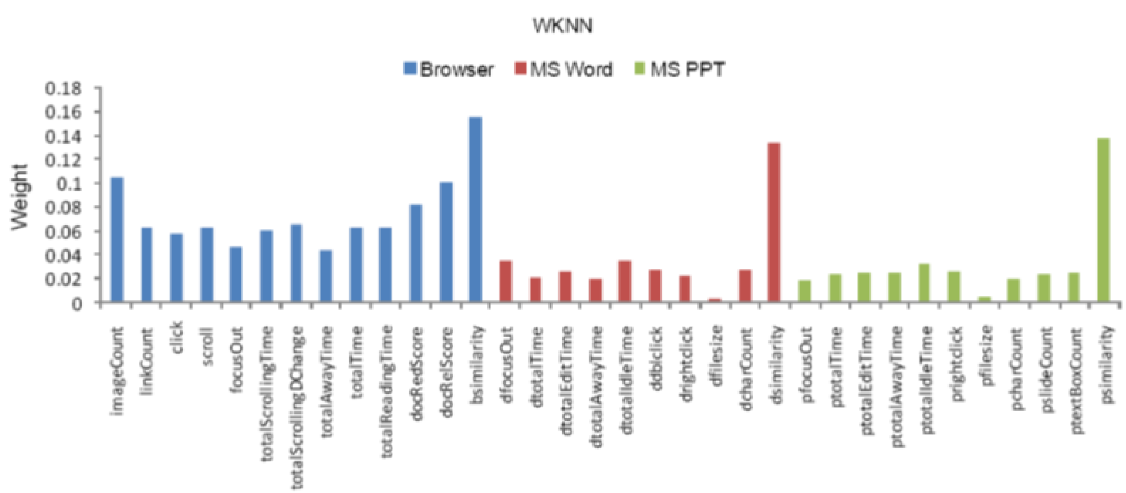

Figure 16: Comparison of Feature Weights Computed from WKNN

In WKNN, features computed (see Figure 16) from the content-consumer applications have higher weights than the features from the content-producer applications except for content similarity. One interpretation of this is that similarity to content being produced by the user is such a strong signals that other features from content-production applications are not needed to help interpret that assessment.

The same cannot be said of content consumption applications. While content similarity is also the strongest feature for the browser, many other features also (including measures of clicks, scrolling, and reading) have strong weights. As opposed to the results from the content production applications, this shows that when assessing activity in the browser, it is important to gauge just how much interest the user has in the content, not just that the content was visited. Each of the three applications contributed one of the three highest strength features. This reinforces the potential for multi-application interest models to improve personalized information delivery via visualizations or recommendations. Feature weighting also indicated that while content similarity is important across all applications, content consumption applications benefit considerably from additional features in order to interpret the perceived value of that content.

We evaluate our models by examining their performance in interest prediction in both page-level and paragraph-level interest modeling. We use Root Mean Square Error (RMSE) to measure the rating prediction quality where a smaller RMSE value indicates better performance. Once the particular topic modeling and evidence weighting schemes were determined based on the results in Sections 6.1 and 6.2, the overall user modeling approach could be examined. The central question is being how the unified user model would perform relative to simpler models. To compare the performance of semi-explicit and unified feedback we compared the performance of classifiers provided with the different sets of features and report on the resulting classifications. We performed our evaluation on page-level user interest estimation by running each user data through the three levels of interest models from baseline-LDA (text edited from production applications), semi-explicit (data from previous model + text annotated from consumption application), and unified (data from previous two + implicit relevance feedback).

Each evaluator provided RMSE on the relevance of each page. The RMSE results for the 4 tasks were computed by averaging the values obtained per each task performance (see Table 5). Although baseline-LDA ($M=1.31$, $SD=0.14$) and semi-explicit models ($M=1.29$, $SD=0.05$) are quite close; $t(3)=0.9459$, $p=0.414$, there was a significant difference in the RMSE for baseline and unified ($M=1.21$, $SD=0.12$); $t(3)= 8.2641$, $p=0.0037$, and semi-explicit and unified; $t(3)= 3.9641$, $p=0.0287$. In all cases the unified relevance model improvement over the semi-explicit relevance models is statistically significant. This demonstrates the importance of implicit relevance feedback indicators in interest predictions.

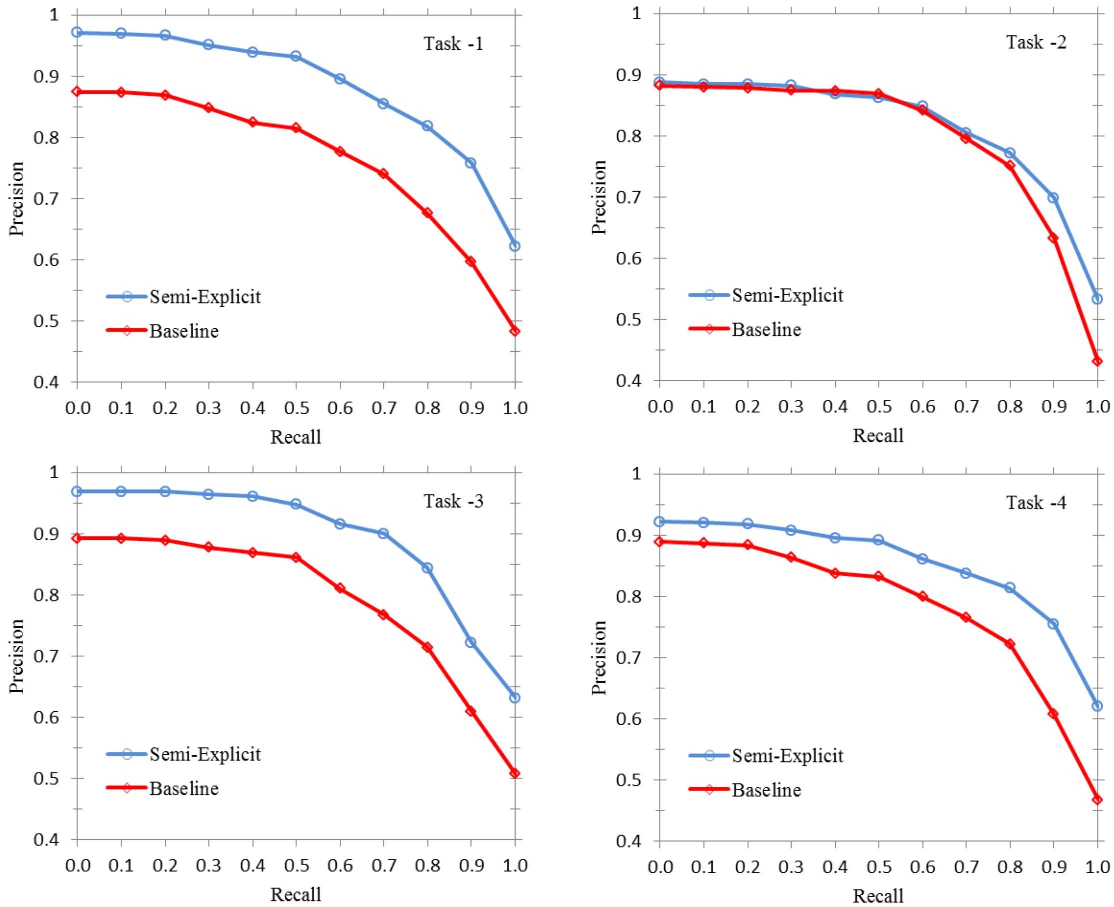

Figure 17: Precision-Recall Segment-level Performance Comparison

Table 5: Page-Level Performance of Interest Models

	Page-Level RMSE			
	Task-1	**Task-2**	**Task-3**	**Task-4**
Baseline-LDA	1.180	1.315	1.239	1.515
Semi-explicit	1.126	1.326	1.258	1.463
Unified	**1.097**	**1.198**	**1.162**	**1.388**

Given that our primary goal is to learn the user's preference from her relevance feedback and use these to identify relevant document content, we consider the standard information retrieval domain evaluation metrics such as precision, recall, harmonic mean (F1), and mean average precision (MAP) to compare the performance of alternative user modeling techniques. MAP gives us an overall sense of how well we identify relevant estimations to recommend from sent of annotation content.

Table 6: Segment-Level Performance of Semi-Explicit Models

	Segment-Level							
	Task-1		**Task-2**		**Task-3**		**Task-4**	
	MAP	F1	MAP	F1	MAP	F1	MAP	F1
Baseline	0.6276	0.5308	0.6371	0.5486	0.6586	0.5739	0.6293	0.5376
Semi-explicit	**0.7827**	**0.6208**	**0.6943**	**0.5568**	**0.7912**	**0.6391**	**0.7488**	**0.5804**

Clearly the unified approach was of value when locating whole resources of interest. But being able to identify relevant segments within the pages is also important for personalized information delivery. We were thus particularly interested in these models performance in this respect.

To examine this segment-level performance we compared the ordering of the segments' similarity to the user models for each task performed by each user to that user's ordered rating of those segments. We calculate MAP and F1 for each task, judging a segment as relevant when it was annotated by the user (see Figure 17).

Unfortunately, the implicit data captured is limited to page-level analysis (we do not know what particular content was being presented when users performed each recorded event). Therefore we only compare the baseline model and the model including semi-explicit content. Table 6 points out the benefit of exploiting paragraph-level user interest via user annotations. MAP improvement of semi-explicit model is both substantial and significant over the baseline-LDA.

5.3 User Study 3 – Dynamic System (2016)

For the third study, there are 3 different system modes depending on the availability of recommendations: baseline system without any recommendations, using semi-explicit system (user annotations), and unified system (implicit + semi-explicit), respectively. Table 7 shows evaluation groups which are all permutations of three different system modes (considering the order of 3 system modes).

Table 7: User Study Groups

	Tasks 1	**Tasks 2**
Group 1	Mode 1	Mode 2
Group 2	Mode 2	Mode 1
Group 3	Mode 1	Mode 3
Group 4	Mode 3	Mode 1
Group 5	Mode 2	Mode 3
Group 6	Mode 3	Mode 2

The participants were randomly assigned to one of the groups. In each group, two system modes were evaluated and the same two tasks were assigned to the participants in each system mode. The entire assignments to each group had equal numbers of the participants to be balanced. In brief, after learning about the system, the participants were asked to perform the two tasks in each system mode according to their group. They completed initial demographic survey (Question set 1), after completion of each task another survey dependent on the system mode (Question sets 2, and 3), and finally a general survey about the overall system (Question set 4). We define the following System Mode Configurations based on the number of application available and the availability of recommendation support:

System Mode 1: *All applications available. No recommendations*

System Mode 2: *All applications available. Only recommendations based on semi-explicit relevance feedback (user annotations and authored text)*

System Mode 3: *All applications available. Complete unified recommendations from both implicit and semi-explicit relevance feedback.*

5.3.1 User Tasks Procedures

The participants involved in the study spent about 60 minutes with the several everyday applications (VKB, Web Browser, MS Word, MS PowerPoint, and Adobe Acrobat PDF). The participants were given a task (Please see Appendix B "Task Sheet" for task definition) to read and identify the relevant content through web search in the VKB application.

Table 8: User Ratings for All the Participants

	Task 1 - Documents								Task 2 - Documents							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
User-10	5	5	4	4	3	3	2	1	5	4	2	5	5	1	2	5
User-11	5	5	1	1	1	1	2	2	5	5	2	5	4	1	1	5
User-12	5	5	1	1	1	1	1	1	5	4	1	4	3	5	1	4
User-13	5	4	4	2	2	3	1	3	5	5	2	3	5	3	2	5
User-14	5	5	3	3	3	2	4	3	5	5	3	4	4	3	3	5
User-15	5	5	3	2	4	3	2	3	5	5	3	5	3	3	3	4
User-16	5	5	2	3	3	4	4	5	5	5	3	5	5	2	2	5
User-17	4	4	2	1	2	2	1	1	4	3	3	4	3	4	3	2
User-18	5	5	2	4	2	2	3	3	5	5	3	4	3	1	4	5
User-19	5	3	4	5	4	3	2	3	4	5	2	5	3	1	2	5
User-20	5	3	2	3	4	2	1	1	5	5	1	3	4	1	1	5
User-21	5	4	2	3	1	1	1	1	4	4	3	4	3	2	2	4
User-22	5	3	3	2	3	2	2	1	5	5	2	3	4	1	2	3
User-23	4	3	2	2	1	1	2	2	4	3	1	4	4	1	1	4
User-24	5	4	3	3	2	3	1	1	5	4	5	4	5	1	2	5
User-25	5	4	2	2	2	1	1	1	5	5	3	4	4	1	1	5
User-26	5	5	5	5	5	4	2	2	5	5	4	5	5	3	3	5
User-27	5	3	3	4	3	4	3	3	5	4	5	5	5	4	3	5
User-28	5	5	3	4	3	3	3	3	5	5	5	5	4	2	2	5
User-28	5	3	4	3	1	1	1	1	5	5	3	3	4	2	2	5
User-29	5	4	2	3	1	1	3	3	5	4	2	2	5	1	2	5
User-30	5	5	4	4	4	4	3	2	5	5	5	5	5	3	4	5
User-31	5	4	3	3	3	4	4	4	5	5	3	4	4	4	3	4
User-32	4	5	3	1	3	1	5	4	5	5	4	5	4	3	4	5
User-33	4	4	3	3	3	3	2	3	5	5	2	4	4	3	3	5
User-34	5	4	2	2	3	1	3	5	1	2	1	4	3	1	1	5
User-35	3	2	4	4	3	5	4	4	5	4	3	4	3	4	3	5
User-**37**	5	5	3	3	2	3	3	2	5	4	3	4	3	3	3	4
User-38	5	5	3	3	3	2	4	3	3	4	3	3	4	4	2	3
User-39	5	4	4	4	3	2	1	2	5	3	2	3	2	3	3	1
Avg.	**4.8**	**4.2**	**2.9**	**2.9**	**2.6**	**2.4**	**2.4**	**2.4**	**4.7**	**4.4**	**2.8**	**4.1**	**3.9**	**2.4**	**2.3**	**4.4**
Std.	**0.5**	**0.9**	**0.9**	**1.1**	**1.0**	**1.2**	**1.2**	**1.2**	**0.8**	**0.8**	**1.2**	**0.8**	**0.8**	**1.2**	**0.9**	**1.0**

The participants were asked to highlight and annotate using the WebAnnotate browser plug-in tool, the relevant content in Mozilla Firefox web browser and Adobe Acrobat Writer via the PDFPad plug-in tool. Simultaneously, they were asked to prepare a Microsoft Word document and Power Point presentation related to the task. After the task-completion, they were given a task-specific questionnaire which was related to their experience of using our applications. After completion of the two tasks, users were asked to rate each web document (1 to 5) given in both tasks (see Table 8).

5.3.2 User Tasks Definitions and Instructions

Task 1 (about 30 minutes)

What is Mars One Project? *Find information related to Mars One project and prepare a summary Word Document and PowerPoint presentation.*

Task 2 (about 30 minutes)

How to improve your credit score? *Find information related to this topic and prepare a summary Word Document and PowerPoint presentation.*

Task Instructions (for Mode 1)

- Complete the given survey (Question set 1).
- Look at the list of documents given in VKB application (8 web documents). To further view each, you can right click on the document and select open from option menu. This will open the document in Mozilla Firefox web browser.
- You can also utilize the given PDF documents (2 PDF documents) to find information related to the task.
- Prepare a summary Word document and PowerPoint document using given templates.
- You can copy/paste or write in your own words a summary (few paragraphs) and couple of slides in PowerPoint.
- Save and close both Word and PowerPoint.
- Now complete the given survey (Question sets 2).

- Also rate each of the given 8 web documents (in VKB Document list) by 1-5, 1- least relevant and 5- most relevant.

Task Instructions (for Mode 2 and Mode 3 systems)

- Complete the given survey (Question set 1).
- Look at the list of documents given in VKB application (8 web documents). To further view each, you can right click on the document and select open from option menu. This will open the document in Mozilla Firefox web browser.
  - If you need automatic recommendations for your documents, click Ctrl+S in Word or PowerPoint.
  - Also if find any relevant content after opening the web browser document, you can utilize WebAnnotate tool to highlight paragraphs using any color.
- You can also utilize the given PDF documents (2 PDF documents) to find information related to the task.
  - If you need automatic recommendations for your documents, use highlight tool in PDF and click on the submit button to find relevant content from web documents in VKB or in browser.
- Prepare a summary Word document and PowerPoint document using given templates.
- You can copy/paste or write in your own words a summary (few paragraphs) and couple of slides in PowerPoint.
- Save and close both Word and PowerPoint.
- Now complete the given survey (Question sets 3 and 4).

- Also rate each of the given 8 web documents (in VKB Document list) by 1-5, 1bing least relevant and 5 being most relevant.

5.3.3 Task Documents and User Ratings

In addition to the post-task questionnaire, we asked participants to rate the relevance of each of the 8 task-specific web pages. These ratings are on a scale from 1 to 5 (1 being Least Relevant & 5 being Most Relevant). Table 8 shows complete list of user ratings from the user study evaluations. Table 9 shows the number of pages in each document and word count per document (approximate) for both task 1 and task 2.

Table 9: Task-wise Page and Word Count

	Task 1	**Task 2**
Total Page and Word Count	**38 (17656)**	**38 (12926)**
Document 1	2 (686)	3 (1347)
Document 2	3 (705)	6 (2002)
Document 3	3 (986)	3 (902)
Document 4	2(329)	3 (1158)
Document 5	4 (1779)	2 (805)
Document 6	2 (1184)	3 (1260)
Document 7	2 (775)	2 (559)
Document 8	2 (570)	2 (776)
Document 9 (PDF)	10 (6318)	8 (1809)
Document10 (PDF)	8 (4324)	6 (2308)

5.3.4 User Interest Shift (Sub-Tasks)

We are also interested in investigating the modeling of changes in user interest in the current task environment. After exposure to different types of information during the tasks, a user's interest may shift or expand to include new areas of interest that may be in contrast to the current activity or become more specific. Depending on the user, these changes may be rapid or take place gradually.

In the current context, we are interested in rapid changes in information need but not drastic changes with respect to the task objective. Therefore, to validate the interest drift and to test our user models (Mode 2 and Mode 3); we added such a change to the tasks to verify this effect of the interest shift by including a sub-task activity. In particular, users were interrupted in the current task and given following sub-tasks in each Mode 2 (semi-explicit) and 3 (unified) to simulate this behavior:

Sub-Task 1 (about 2 minutes)

What is the name of the recent academy award nominated movie about Mars exploration? Highlight this information in the web page using WebAnnotate "Green" color. Write a short sentence about this movie in your word and PowerPoint documents.

Sub-Task 2 (about 2 minutes)

What are the 3 main credit reporting agencies? Highlight this information in the web page using WebAnnotate "Green" color. Write a short sentence about 3 credit report agencies in your word and PowerPoint documents.

5.3.5 Study Participants

This study was conducted to evaluate the final dynamic system with unified feedback and to compare it with the other two system modes (baseline and semi-explicit). This provides data concerning whether the identified unified interests indicators are effective in recognizing user interests during information gathering tasks. The study took place at Texas A&M University. A total of 30 subjects were recruited via social media and other contacts. 21 respondents were male and 9 were female. Ages of respondents ranged from 20 or younger to 50 or older but the majority (57%) were from the 21-25 age group, with 17% from 26-30. Participants came from variety of ethnic origins (see Figure 18). Most of the respondents had work experience while 50% had already received a graduate degree (MS, MPhil, PhD), the rest of reported a Bachelor's degree or currently enrolled in a Bachelor's degree program. 46% of the participants had an engineering background (Computer Science, Computer Engineering, and Electrical Engineering) and the others were from diverse areas (2 from Mathematics, 1 from Statistics, 1 from Molecular Biology and 1 from Agricultural Biology).

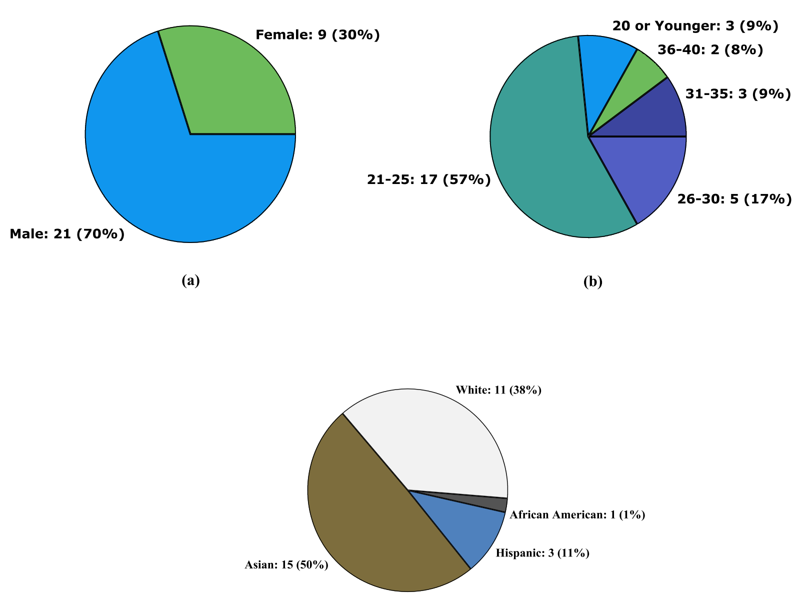

Figure 18: User Study Participants (a) Sex, (b) Age and (b) Ethnic Origin

All participants reported using computers daily and have used mostly Personal Computers (PC), Laptop/Notebook or tablets in their daily activities. 80% of participants reported using the computers in their home and school environments and among them about 40% of them using in their daily work environments. They were highly internet literate with 93% of respondents reporting Heavy computer usage (20 hours or more per week).

5.4 User Study 3 – Results

5.4.1 Perception of Participants

We first assess the how often the participants feel like they find larger amount of information for consumption than their devices are capable of providing in a reasonable manner. Survey results show that information overload is common and finding suitable information for consumption is an issue for almost 90% of respondents (see Figure 19). Among the number of users who find that they locate more information than they can evaluate, about 86% use PC, 84% use Laptop / Notebook, 92% use tablet and about 95% use cell phone.

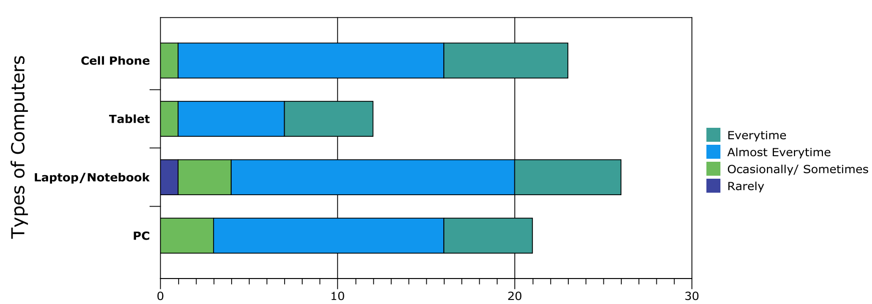

Figure 19: Information Overload across Types of Computing Devices Usages

Regarding perceptions from the task, we also investigated whether the participants felt overwhelmed by the number of applications available for the given task (see Figure 20).

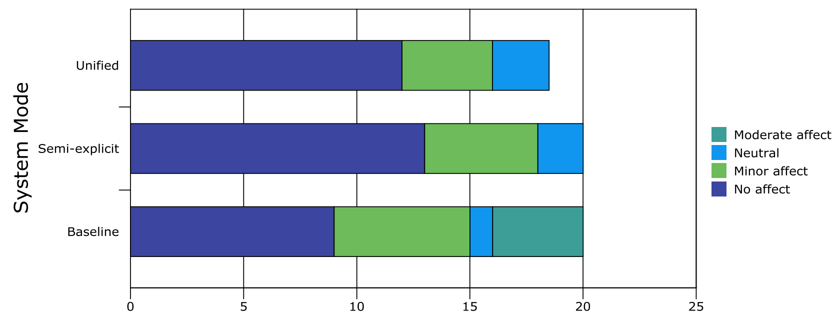

Figure 20: Number of Applications Available for the Given Task

We found that users were relatively comfortable (baseline: 45%, semi-explicit:65%, and unified: 65%) with the number of given applications for the task procedures. Only few of them (20%) found the application environment is moderately overwhelming, but interestingly, this was evident only in the Baseline system mode (without either types of recommendations support from semi-explicit and unified).

5.4.2 Multi-Application Environments and Privacy Issues

We also examined how easy (or difficult) it was for the participants to use multiple applications for the given tasks (semantic-differential question). The overall consensus is that using multiple applications in the current task environments is easy or somewhat easy. There are higher neutral responses from baseline group than both the semi-explicit and unified system configurations (see Figure 21).

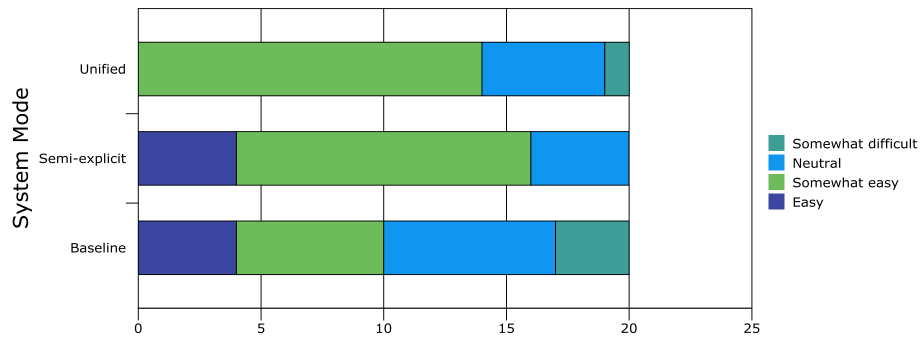

Figure 21: How Easy for the User to Use Multiple Applications

Our user study environment is based on dual-monitor system and entirely PC based system configuration with participants allowed to layout the given software applications in their preferred view. We believe that this is a main reason for a higher percentage of participants finding both the number of applications available for the task environment as not overwhelming. This level of comfort may also indicate that participants have experience with similar activities in their regular computer use.

At the end of the user questionnaire, we asked participants whether they are comfortable having a system monitoring their activities (in background) in daily activities and interactions with everyday applications (see Figure 22): *"Given the nature of the tasks, I didn't not feel my privacy was breached. As long as the data used for the recommendations is only shared with in the application I think it is reasonable. Even when credit scores, the browser is already looking at the history s this isn't any less promising than normal browsing. However, if the task was more sensitive such as medical in nature I might feel otherwise."*

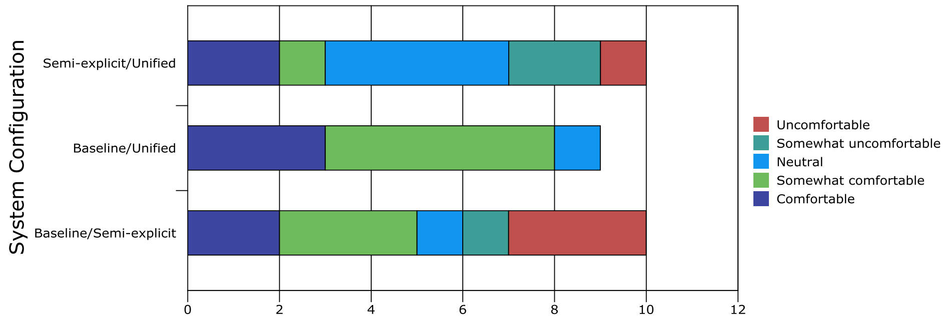

Figure 22: User-System Interactions and Monitoring

Other participants find that because the system resides on their local system, they do not find it breaching their privacy but they would be if it was online or more of a web-based system without the control over monitoring the data: *"No but I would feel so if it was an online system where I had no control over monitored data"*

5.4.3 Document Relevance

We examined how relevant the given documents (in VKB search list and 2 PDF documents) for the two tasks assigned. Figure 23 shows the result from each system mode. All participants find the document lists as relevant to the given tasks and participants report that it is easy to identify relevant web pages from the given VKB document list.

Q1 List of documents (in VKB and PDF documents) given for the task are relevant

Q2 It was easy to identify relevant web pages from VKB document list

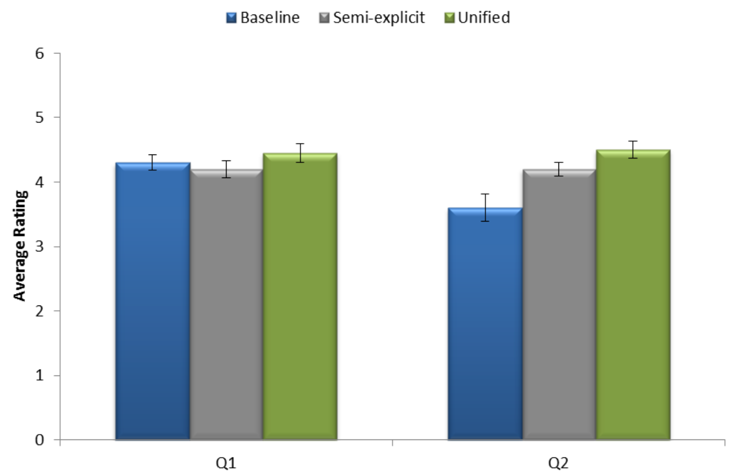

Figure 23: Document Relevance for Each Tasks

5.4.4 Task-Wise Interactions

It is also important to identity issues relevant to the statistical interaction potentially arises when there are two given search tasks and testing the user preferences of recommendations. Table 10 shows the ratings from each user that took each task and the relevant system mode and the average rating and standard error of the ratings. Under the task 1, the ratings of recommendations are not significantly different. Under the task 2, there is a minimal difference (3.6 versus 4.3). Therefore, we don't have any evidence of interaction in this study.

5.4.5 Model Comparisons

We also asked the participants which system configuration helped them to find relevant content while working on the tasks (see Figure 24). Overall consensus is that when compared to the baseline environment, the system with the recommendation support (via authored/annotated text or combined with implicit relevance feedback) is superior.

This was expected and obvious. Interestingly when the system configuration is semi-explicit and unified, a majority of the participants in this group configuration (Mode 2-3), find that both systems performs adequately. Users also found that the multi-application environment provided was helpful in finding interesting content from long list of documents and during search tasks.

Table 10: Task-Wise Interactions of Recommendations

	Semi-explicit (Average ± Standard Error)	**Unified** (Average ± Standard Error)
Task 1	4.1 ± 0.2	4.2 ± 0.2
Task 2	3.6 ± 0.4	4.3 ± 0.2

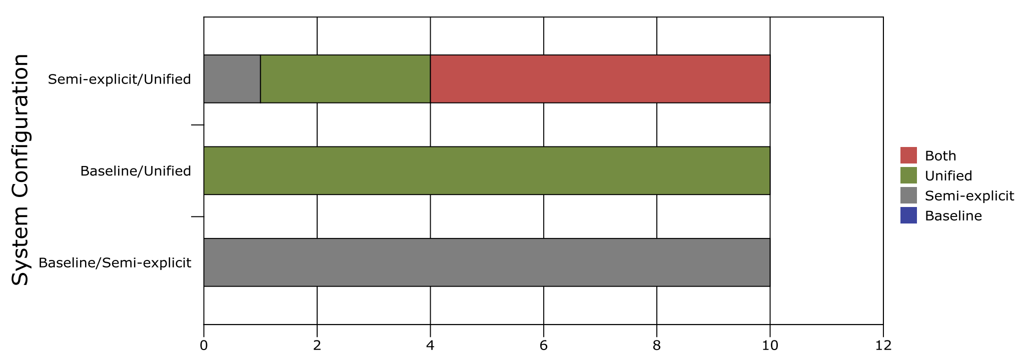

Figure 24: Which Model Helps to Find the Relevant Content?

5.4.6 Participant Task Activities

While the users were performing the task activities, user actions in each application (VKB, Web Browser, Word, PowerPoint, and Acrobat PDF) were logged. The log of task active time includes the start of the first application and the end of the session by closing the last application. For the purpose of this study, a task-session is defined by a continuous series of logged interactions that refers to the start and end of system server application (IPM).

Table 11 and Figure 25 show the time spent on each task and the total time for both tasks based on the system mode assigned for each participant. The time in each task is in seconds. The average task time for Baseline participants is (1514.15), Semi-Explicit participants is (1540.7) and Unified participants is (1442.05), which are not significantly different from each other (Modes 1,2 $p > 0.8$, Modes 1,3 $p > 0.5$ and Modes 2,3 $p > 0.5$ for two-tailed t-test).

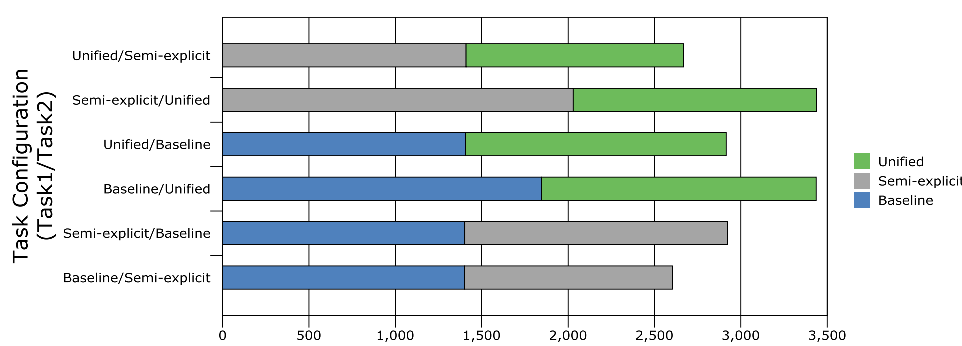

Figure 25: Time Spent on Each Task

Table 11: Quality of User Summary Based on 5 Reviewers Average Rating

	Quality Rating	Mode	Task 1 time	Task 2 time	Total Time Spent
User -10	4.00	2,1	1381	785	2166
User -11	4.70	2,1	2142	1916	4058
User -12	2.70	1,2	1127	1300	2427
User -13	3.40	2,1	1274	1604	2878
User -14	2.35	2,1	930	1577	2507
User -15	4.95	2,1	1874	1129	3003
User -16	2.90	1,2	1409	882	2291
User -17	3.65	1,2	1122	1298	2420
User -18	3.20	1,2	2065	1251	3316
User -19	3.25	1,2	1282	1282	2564
User -20	3.65	2,3	1725	888	2613
User -21	2.95	3,2	901	2679	3580
User -22	2.05	2,3	2188	2169	4357
User -23	2.85	3,2	1043	967	2010
User -24	4.15	2,3	2259	1793	4052
User -25	2.55	3,1	1656	1622	3278
User -26	4.70	3,2	1482	871	2353
User -27	4.50	3,2	1815	1352	3167
User -28	3.75	2,3	2702	1283	3985
User -29	3.80	3,1	1890	1363	3253
User -30	3.80	1,3	2025	1951	3976
User -31	3.35	3,1	2080	1538	3618
User -32	4.60	1,3	2087	1400	3487
User -33	3.85	1,3	1660	1422	3082
User -34	2.60	3,1	833	1026	1859
User -35	3.10	1,3	1853	1683	3536
User -36	3.65	3,2	1062	1178	2240
User -37	2.95	2,3	1279	909	2188
User -38	2.75	1,3	1612	1490	3102
User -39	4.45	3,1	1091	1481	2572
Average			**1514.15**	**1540.70**	**1442.05**
Standard Error			**81.67**	**128.46**	**95.62**

Thus it appears that the task duration is not significantly affected by the different system configurations. On average, unified system participants took marginally less time than the other two configurations.

To assess the quality of the results each pair of word and PowerPoint documents was assessed by five reviewers (senior Ph.D. students in CS). Figure 26 shows that the quality rating received from the 5 reviewers for the Word and PowerPoint summary prepared by each participant. The general trend is that the quality of the Word and PowerPoint summary reflects on more time spent in preparing the documents.

Figure 26: Quality of User Summary Based on 5 Reviewers Average Rating

5.4.7 User Model Performance with User Ratings

In order to compare the semi-explicit and unified system performance, we compared on the user's post-task ratings for each of the 8 web documents and the computed ratings from both semi-explicit and unified models. To examine the performances, we compared the semi-explicit, unified and unified* model by calculating the RMSE for each of the 8 web documents at each task level.

5.4.8 Unified* User Model

Our user evaluation includes 10 participants from (Groups 5 and 6) system Mode 2, 3 and system Mode 3,2 combination where each Mode number specifies the task (task 1 or task 2) in which the participant was assigned first during the two task procedures. Our initial user study is based on the μ (0.8, 0.2) and we learn coefficients for the equation 8 for target-document ratings based on a regression analysis (see Figure 27) from the inferred semi-explicit and implicit relevance feedback from the user study data. There was a significant difference of unified* model performance at the $p < 0.05$ compared to user ratings [$F (2, 20408) = 144.096$, $p=0.00$]. Unstandardized model coefficients are 0.264 and 0.269 respectively (before normalization) for semi-explicit and implicit x proxy-similarity and the values are significantly different based on t-test.

Table 8 shows the user ratings for all the participants and Figure 28 shows the average RMSE for all 10 participants from the semi-explicit, unified and unified*. Furthermore, Table 12 , Figure 28 and Figure 29 summarizes the resultant data from the 10 participants with semi-explicit, unified and unified* for each document (8 web pages) from task 1 and task 2.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.118[a]	.014	.014	.3466415	1.333

a. Predictors: (Constant), Implicit_X_proxy_similarity, Semi_explicit

b. Dependent Variable: UserRating

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	34.629	2	17.315	144.096	.000[b]
	Residual	2452.232	20408	.120		
	Total	2486.861	20410			

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	.481	.011		45.309	.000	.460	.502		
	Semi_explicit	.264	.017	.106	15.149	.000	.229	.298	.978	1.022
	Implicit_X_proxy_similarity	.269	.050	.038	5.349	.000	.170	.367	.978	1.022

Figure 27: Unified* Model Parameter Learning through Regression Analysis

Table 12: RMSE for 10 Participants from User Models

		Participant #									
	Doc #	**20**	**21**	**22**	**23**	**24**	**26**	**27**	**28**	**36**	**37**
Semi-explicit	**1**	0.51	0.29	0.37	0.37	0.47	0.49	0.45	0.60	0.41	0.65
	2	0.16	0.31	0.13	0.10	0.20	0.55	0.24	0.61	0.53	0.44
	3	0.31	0.08	0.15	0.38	0.10	0.27	0.48	0.15	0.12	0.03
	4	0.17	0.34	0.27	0.37	0.17	0.52	0.36	0.40	0.10	0.04
	5	0.31	0.11	0.17	0.32	0.26	0.54	0.42	0.19	0.22	0.17
	6	0.25	0.20	0.20	0.42	0.11	0.10	0.34	0.16	0.17	0.10
	7	0.53	0.26	0.28	0.32	0.56	0.08	0.12	0.31	0.20	0.43
	8	0.50	0.31	0.42	0.34	0.59	0.52	0.48	0.15	0.56	0.38
	Avg.	**0.34**	**0.24**	**0.25**	**0.33**	**0.31**	**0.38**	**0.36**	**0.32**	**0.29**	**0.28**
	Std.	**0.15**	**0.10**	**0.11**	**0.10**	**0.20**	**0.2**	**0.13**	**0.20**	**0.18**	**0.23**
Unified	**1**	0.73	0.73	0.61	0.43	0.63	0.67	0.68	0.72	0.59	0.71
	2	0.72	0.48	0.67	0.24	0.41	0.69	0.23	0.72	0.19	0.48
	3	0.23	0.15	0.11	0.11	0.64	0.65	0.23	0.70	0.46	0.25
	4	0.26	0.23	0.21	0.11	0.35	0.64	0.40	0.73	0.20	0.50
	5	0.50	0.28	0.47	0.27	0.63	0.71	0.25	0.52	0.31	0.28
	6	0.29	0.30	0.34	0.29	0.34	0.46	0.49	0.08	0.35	0.24
	7	0.28	0.31	0.12	0.09	0.15	0.14	0.25	0.12	0.38	0.28
	8	0.70	0.29	0.22	0.10	0.63	0.14	0.27	0.75	0.37	0.50
	Avg.	**0.46**	**0.35**	**0.34**	**0.21**	**0.47**	**0.51**	**0.35**	**0.54**	**0.36**	**0.41**
	Std.	**0.22**	**0.18**	**0.22**	**0.12**	**0.19**	**0.24**	**0.16**	**0.28**	**0.13**	**0.17**
Unified*	**1**	0.44	0.54	0.43	0.19	0.45	0.57	0.53	0.58	0.47	0.35
	2	0.46	0.31	0.48	0.12	0.24	0.60	0.14	0.59	0.15	0.21
	3	0.53	0.27	0.23	0.31	0.52	0.47	0.11	0.59	0.32	0.20
	4	0.17	0.16	0.15	0.22	0.21	0.56	0.21	0.63	0.20	0.29
	5	0.31	0.47	0.34	0.50	0.51	0.62	0.14	0.44	0.37	0.20
	6	0.53	0.49	0.48	0.52	0.45	0.38	0.36	0.16	0.43	0.20
	7	0.53	0.49	0.21	0.30	0.29	0.23	0.16	0.22	0.45	0.17
	8	0.48	0.46	0.17	0.30	0.48	0.20	0.16	0.66	0.42	0.26
	Avg.	**0.43**	**0.4**	**0.31**	**0.31**	**0.39**	**0.45**	**0.23**	**0.48**	**0.35**	**0.24**
	Std.	**0.13**	**0.14**	**0.14**	**0.14**	**0.13**	**0.17**	**0.15**	**0.2**	**0.12**	**0.06**

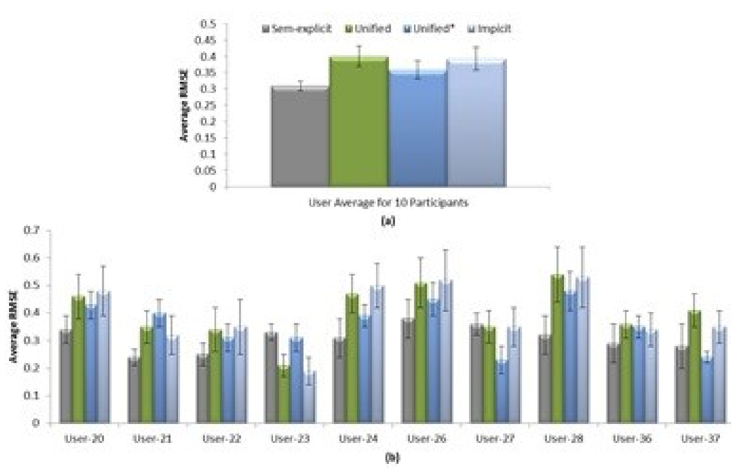

Figure 28: Average RMSE (a) Aggregated Average RMSE (b) All 10 Participants

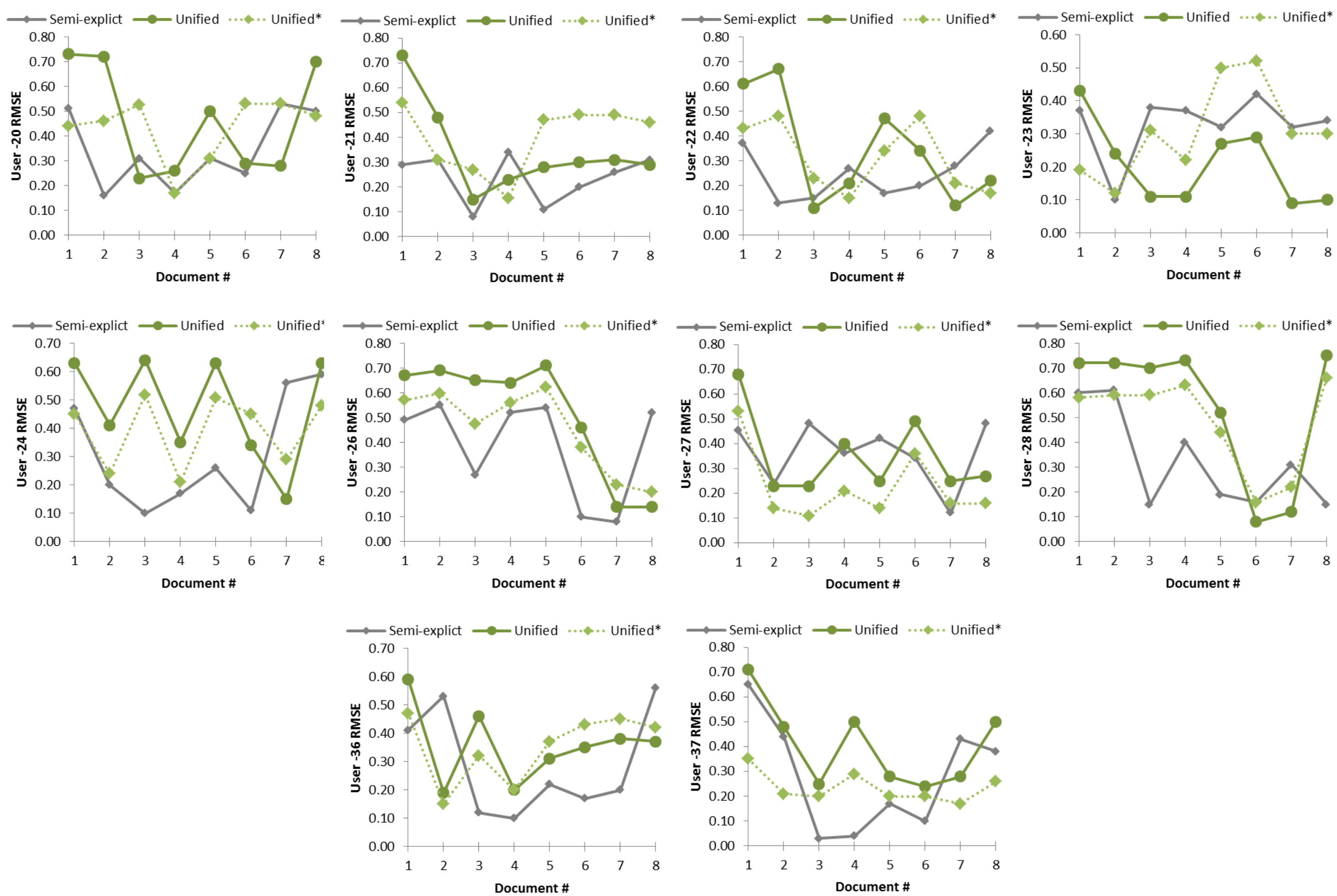

Figure 29: RMSE Values for 10 Participants from User Models

We calculate the difference within the 3 groups of models based on 1-way ANOVA between participants to compare the difference of performance between semi-explicit, unified and unified*. The semi-explicit performance at the $p > 0.05$ level for the other two system modes [F (2, 27) = 3.197, p=0.057]. We further investigate the 3 models revealed in ANOVA by a Post-Hoc test with the help of SPSS multiple comparisons.

ANOVA

RMSE

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.041	2	.020	3.197	.057
Within Groups	.171	27	.006		
Total	.212	29			

Multiple Comparisons

Dependent Variable: RMSE

	(I) Model	(J) Model	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Tukey HSD	1	2	-.09000*	.03564	.045	-.1784	-.0016
		3	-.04900	.03564	.368	-.1374	.0394
	2	1	.09000*	.03564	.045	.0016	.1784
		3	.04100	.03564	.492	-.0474	.1294
	3	1	.04900	.03564	.368	-.0394	.1374
		2	-.04100	.03564	.492	-.1294	.0474
Dunnett T3	1	2	-.09000	.03419	.059	-.1832	.0032
		3	-.04900	.03071	.335	-.1319	.0339
	2	1	.09000	.03419	.059	-.0032	.1832
		3	.04100	.04122	.691	-.0671	.1491
	3	1	.04900	.03071	.335	-.0339	.1319
		2	-.04100	.04122	.691	-.1491	.0671

*. The mean difference is significant at the 0.05 level.

Figure 30: 1-way ANOVA Multiple Comparisons for 3 Models

Multiple comparisons did not reveal a significant difference in performance between semi-explicit and unified* models with ($p > 0.05$). Tukey's HSD shows a significant difference between model performance between the semi-explicit and unified ($p < 0.05$). In Figure 30, samples include following notations; 1= semi-explicit, 2 = unified, 3 = unified*. The overall ANOVA asks a question about the whole independent variable and its relation (or lack thereof) to the dependent variable. The pairwise comparisons ask about differences among pairs. Then the p-value looks at the statistical significance of each of these, with the pairwise adjusted for multiple comparisons (in this case, using Tukey's HSD and Dunnett T3 methods).

We also evaluate the resultant data from the same 10 participants (see Figure 31) by calculating the RMSE with the average rating across all 30 participants. 1-way ANOVA between the 3 models (between 3 models with average user ratings) was conducted to compare the difference of performance between system modes semi-explicit, unified and unified*. There is no significant difference of semi-explicit performance at the $p > 0.05$ level for the models [$F(2, 27) = 0.201$, $p=0.819$]. Table 13 and Figure 32 show the resultant RMSE for the 10 participants in each model with average user ratings. We compare these 3 models of average ratings with previous 3 models (see Figure 28 and Figure 31). There is no significant difference between the user ratings and average user ratings in semi-explicit model ($p > 0.05$). There is a significant difference between the user ratings and average user ratings in unified model and unified* (for models, $p < 0.05$, t-test). Unified and unified* show 25% and 23% performance improving respectively by using average user ratings for RMSE calculations.

Table 13: RMSE for 10 Participants from User Models with Average User Ratings

		Participant #									
	Doc #	**20**	**21**	**22**	**23**	**24**	**26**	**27**	**28**	**36**	**37**
Semi-explicit	**1**	0.28	0.50	0.14	0.37	0.25	0.25	0.21	0.35	0.16	0.40
	2	0.16	0.47	0.13	0.34	0.20	0.31	0.24	0.13	0.28	0.06
	3	0.12	0.50	0.15	0.14	0.10	0.09	0.06	0.15	0.11	0.03
	4	0.17	0.43	0.11	0.37	0.17	0.27	0.18	0.24	0.21	0.04
	5	0.20	0.42	0.20	0.32	0.26	0.29	0.19	0.17	0.22	0.17
	6	0.25	0.44	0.20	0.18	0.28	0.18	0.22	0.21	0.17	0.35
	7	0.53	0.50	0.50	0.12	0.56	0.29	0.32	0.69	0.20	0.93
	8	0.51	0.47	0.42	0.34	0.59	0.29	0.24	0.38	0.31	0.62
	Avg.	**0.28**	**0.47**	**0.23**	**0.27**	**0.3**	**0.25**	**0.21**	**0.29**	**0.21**	**0.33**
	Std.	**0.16**	**0.03**	**0.15**	**0.11**	**0.18**	**0.07**	**0.07**	**0.19**	**0.06**	**0.32**
Unified	**1**	0.49	0.48	0.37	0.43	0.38	0.42	0.44	0.47	0.36	0.46
	2	0.47	0.24	0.43	0.24	0.41	0.21	0.23	0.48	0.19	0.48
	3	0.26	0.27	0.23	0.25	0.16	0.20	0.23	0.24	0.23	0.25
	4	0.50	0.35	0.44	0.30	0.35	0.19	0.17	0.49	0.20	0.50
	5	0.50	0.28	0.47	0.09	0.38	0.13	0.11	0.52	0.11	0.52
	6	0.12	0.30	0.14	0.09	0.11	0.14	0.10	0.08	0.14	0.09
	7	0.12	0.31	0.11	0.30	0.15	0.35	0.30	0.12	0.38	0.08
	8	0.46	0.29	0.45	0.30	0.38	0.31	0.27	0.50	0.37	0.50
	Avg.	**0.37**	**0.32**	**0.33**	**0.25**	**0.29**	**0.24**	**0.23**	**0.36**	**0.25**	**0.36**
	Std.	**0.17**	**0.07**	**0.15**	**0.11**	**0.13**	**0.1**	**0.11**	**0.18**	**0.11**	**0.19**
Unified*	**1**	0.23	0.31	0.23	0.19	0.22	0.33	0.30	0.37	0.34	0.16
	2	0.25	0.16	0.26	0.12	0.24	0.16	0.14	0.38	0.35	0.21
	3	0.26	0.24	0.15	0.10	0.13	0.12	0.11	0.23	0.22	0.20
	4	0.29	0.58	0.30	0.14	0.21	0.21	0.15	0.42	0.40	0.29
	5	0.31	0.47	0.34	0.26	0.28	0.21	0.21	0.45	0.44	0.32
	6	0.30	0.49	0.26	0.29	0.22	0.22	0.20	0.16	0.16	0.38
	7	0.30	0.49	0.21	0.54	0.29	0.45	0.45	0.23	0.22	0.30
	8	0.28	0.46	0.33	0.53	0.25	0.39	0.41	0.43	0.42	0.26
	Avg.	**0.28**	**0.4**	**0.26**	**0.27**	**0.23**	**0.26**	**0.25**	**0.33**	**0.32**	**0.26**
	Std.	**0.03**	**0.07**	**0.06**	**0.11**	**0.05**	**0.12**	**0.13**	**0.11**	**0.11**	**0.07**

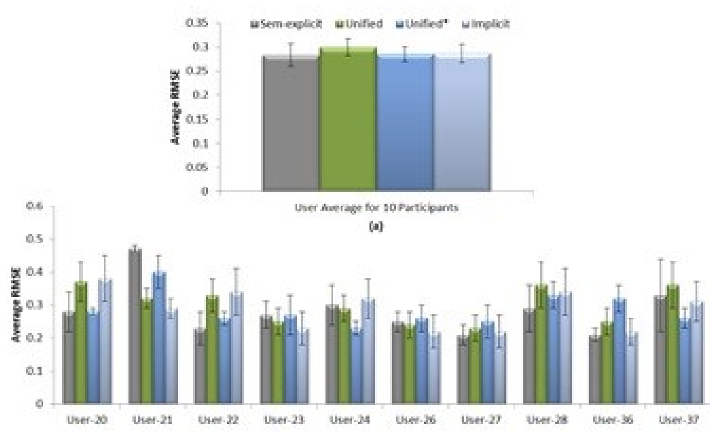

Figure 31: Average RMSE with Average User Ratings (a) Aggregated Average RMSE (b) All 10 Participants

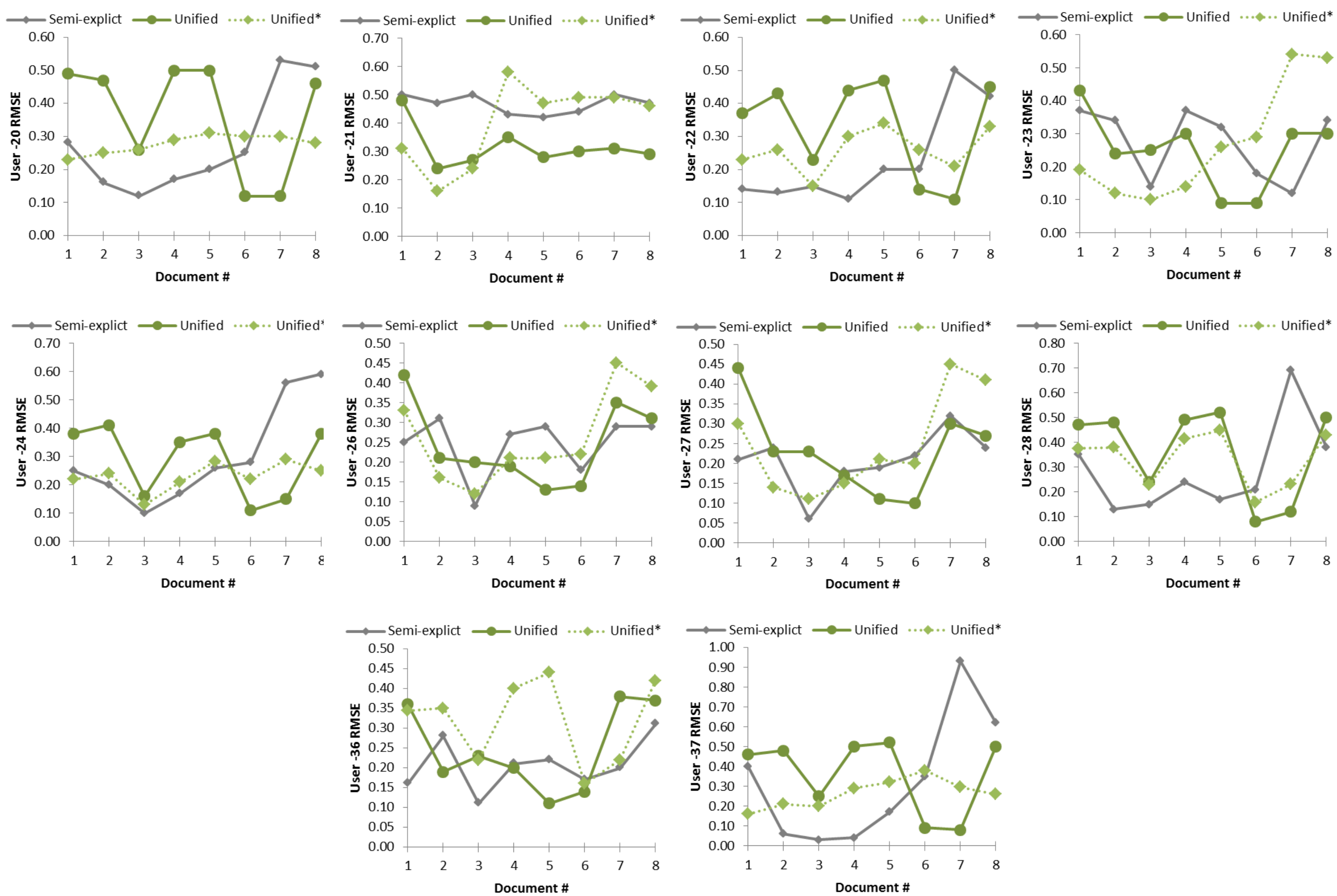

Figure 32: RMSE Values for 10 Participants for 3 Models with Average User Ratings

5.4.9 Qualitative Analysis of Recommendations

Examining the relationships is the centerpiece of the qualitative analysis in the user recommendations and user interest shift process. We employ a decision matrix (see Table 14 and Table 15) to capture how many different concepts from each user are connected to examine further the qualitative content from participant's questionnaires.

When the participants were assigned with either semi-explicit or unified model, we asked from participants: "*Did the recommendations help you to find interesting content relevant to the given task?*"

Table 14: Decision Matrix for System Recommendations Qualitative Analysis

Number of Open-ended responses in each category		
Favorable (semi-explicit, unified)	Neutral (semi-explicit, unified)	Negative (semi-explicit, unified)
34(16,18)	4 (3,1)	2(1,1)

We evaluated favorable outcomes from the participant's open-ended responses from both semi-explicit and unified models with two-tailed t-test and the two system models are not statistically different ($p > 0.13$). Participants find both models equally capable of providing recommendations to support search task completion.

Comments in the open-ended questions related to the recommendation support confirmed that both semi-explicit and unified models helped users in order to complete their tasks. Participants from system mode 2 (semi-explicit) find that the annotations helped them to isolate relevant content for the given tasks by underlining pertinent information:

- *"The recommendations helped me find the important information much more quickly"*
- *"I was impressed by the in document relevance - i.e. when I highlighted something I was shown related things within the open web page. This was useful. I did not find any other recommendation methods to be very useful."*
- *"The recommendations brought my attention to certain links that were more relevant and once on the relevant pages, it brought my attention to the paragraphs that held important information."*
- *"I was able to use the underlined passages to quickly find information related to the topic. I could make my highlights different colors to separate my concerns in the browser and these were reflected in the vkb recommendations. I could open a new webpage link and easily scan it for recommendations."*
- *"It helped me find content a lot faster that I normally would have."*

Interestingly, when the participants are focused on the task at hand, they didn't notice the changes in either VKB document list nor in the Web Browser: *"I'm not sure when the recommendations happened. The list or order might have changed while I was not looking."*

Some participants find the recommendations from annotations are too many to be useful. They prefer to have only the relevant titles recommended so that they can gauge what type of information available in the rest of the content:

- *"It recommends too much to be useful. I liked being able to annotate content, but I could usually gauge what type of content the article has given the title."*

- *" They helped. But there were a bit too many highlights that came up. I didn't find all of the highlighted areas that important, so when they did come up, they were a bit distracting from the rest of the article."*
- *"Not really, I would have liked for it to highlight relevant titles so I could then read what the titles contain or just highlight a few relevant sentences."*

When the participants are given with the unified model in their task 2 (and semi-explicit mode in preceding task), they find that underlined recommendations are more reasonable: *"I liked underlined more this time. It seemed reasonable. , Recommendations did help in tracking down information relevant to task at hand. Located related useful information"*

In addition we asked participant to rate recommendations received during each of the tasks according to following 4 statements. Results are favorable (see Figure 33) for unified model in all 4 statements.

Q3: The recommendations on the VKB web documents list are relevant to the task

Q4: The recommendations on the browser web document paragraphs are relevant to the task

Q5: The visualization provided for recommendations were sufficient

Q6: I was satisfied with the recommendation frequency

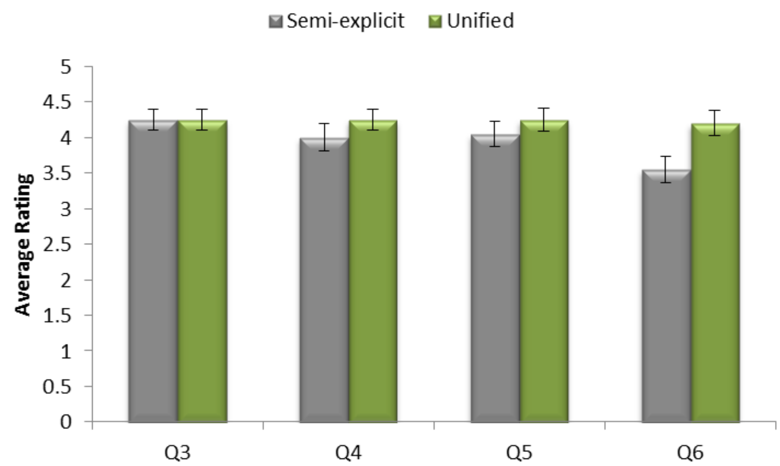

Figure 33: Participants Average Ratings for Recommendations Quality

5.4.10 Qualitative Analysis of Interest Shift

When the participants were assigned with either semi-explicit or unified model, we asked from participants: "*When you change your intent (interest shift), did the recommendations changed accordingly?*"

Results show that the responses are mostly favorable for interest shift for both models.

Table 15: Decision Matrix for User Interest Shift Qualitative Analysis

Number of Open-ended responses in each category		
Favorable (semi-explicit, unified)	Favorable (semi-explicit, unified)	Favorable (semi-explicit, unified)
24(14,10)	8(2,6)	6(3,3)

Comments in the open-ended questions related to the interest shift confirmed that both semi-explicit and unified models support users when their interests shift in ad-hoc manner:

- "*Model changed according to what I highlighted there for it was easier to change my topics quickly and find out relevant information*"
- *"I decided to include that information in my word document and so a few docs containing that info was highlighted"*
- *"It only highlighted in the respective color that related to the sub-task"*
- *"I had been previously focused on finding information of Mars One as a project the sub task highlighted pages that were more people centric and gave more access to new information"*
- *Recommendations changed accordingly, it showed reasons related to the sub-task. This helped in exploring in the direction of things portrayed in sub-task.*

Interestingly, when the participants are focused on the task at hand, they didn't notice the changes in either VKB document list nor in the Web Browser:

- *"I didn't pay attention honestly didn't see recommendations"*
- *"I felt like they didn't change very much. Not too entirely helpful."*
- *"No, I did not notice that as I was working for something related but different"*

Also when the recommendations are not closely relevant to what the user expected to receive, the users tend to move on and rely on the other applications and recommendation methods to obtain new information: *"I changed the intent mostly through saving the Word and PowerPoint documents. I made some annotations in the beginning,* *but they seemed to highlight the entire document, so I relied more on the document saving function."* Participants also find that interest shift is detected but when the recommendation is more general they find it as not entirely as helpful: *"it often annotates too much content to sort through to be more useful than just scanning the page."*

We evaluated favorable outcomes (High, Moderate, and Low) from both semi-explicit and unified models with two-tailed t-test and the difference is statistically significant ($p < 0.03$). Participants find the unified model highly favorable in terms of ad-hoc interest shift.

5.5 Summary

From the three user studies, we have learned that by incorporating topic modeling for representing interests in user models, we can achieve best recall with LDA-JSD similarity method and best precision with either LDA-KL or LDA-H methods. Also, incorporation of semi-explicit data improves performance of segment-level assessment over baseline model. The recommendations based on semi-explicit feedback were viewed the same as those from unified feedback and the semi-explicit feedback was comparable to those from unified feedback in terms of matching post-task document assessments.

6. CONCLUSION AND FUTURE WORK

The work presented in this dissertation addresses a rarely investigated topic: the potential of aggregating activity across multiple applications for user interest modeling. While there are theoretical or software frameworks for distributed user modeling, assessments of modeling techniques are almost always reported in terms of single applications. In this work, we present and evaluate a multi-application modeling technique that combines implicit and semi-explicit feedback across multiple everyday applications.

Our system and tool set supports a wide range of potential applications communicating with the user interest server. To affect the contents of the user interest model an application must be augmented to capture some information about content and its usage. The features described are occasionally specific to the applications (e.g. MS Word and PowerPoint, Firefox) but similar features would be available in most content producer and consumer applications involving text. Thus, the overall architecture and approach will generalize across a wide range of software applications. To the best of our knowledge, this is the first software framework designed to share semi-explicit and implicit relevance feedback among applications.

The evaluation of the alternative modeling techniques involved collecting activity data and post-task relevance assessments for a common type of activity: rapidly browsing/reading content and writing a report or presentation based on that content. While other types of information tasks exist, this is a frequent and broad enough category of task to warrant investigation. There is considerable effort involved in creating an interest model server capable of communicating integrating with real-world applications like Word, PowerPoint, Adobe Acrobat and Firefox. While there is always more that can be done, we believe this infrastructure is substantial and at a reasonable point for assessment.

We have evaluated the effectiveness of the recommendation support from both semi-explicit model (authored/annotated text) and unified model (implicit + semi-explicit) and have found that they are successful in allowing users to locate the content easily because the relevant details are selectively highlighted and recommended documents (in VKB Search List) and passages within documents (in Firefox web browser) based on what the user has indicated interest in already and based on subtle changes of user's indirect interest indicators.

The experimental results show that incorporating implicit feedback in page-level user interest estimation resulted in significant improvements when there is only indirect evidence available for user modeling. Furthermore, incorporating semi-explicit content (e.g. annotated text) with the authored text is effective in identifying segment-level relevant content. Although the study was not designed to test the VKB architecture or basic capabilities, we found that it performed well during the study as a search support interface. We find that the unified model is reasonable in assessing the document value when the semi-explicit (authored/annotated text) data is not available and comparable with semi-explicit only model when both types of feedback are available for inferring user interests. Participants find that the recommendations helped them in locating documents (in VKB Search List) that were more relevant and, once the relevant document is displayed in the Firefox browser, it helped them find paragraphs that held relevant information. Participants also find that both semi-explicit and unified models are reasonable in help them locate content when their interests shift. We find that there is no significant difference between semi-explicit and unified models for supporting interest shifts, and we believe that when the participants are focused on the task-at-hand, they rarely noticed the subtle changes in the VKB Search List. We are interested in investigating how to incorporate interest shift for passage level interactions in the future.

Our results open up many possibilities for using unified feedback in medium and long-term information tasks, especially in the context of personalization of information delivery. Since we have a model that relates the unified feedback to ratings, we can use methods designed for explicit feedback on the unified data. In the future, we plan to study how semi-explicit feedback can be combined with implicit feedback for segment-level assessment and in additional personalized information delivery contexts.

Accurate models of user interest are valuable in personalizing the presentation of the often large quantity of information relevant to a query or other form of information request. Our current software framework helps by capturing user activity across multiple applications and combining this activity data in a user interest model to aid information delivery. In the future, we are interested in extending this user modeling framework based on the non-visible anatomical structure and its characteristics of the human eye.

REFERENCES

Abel, F., R. Baumgartner, A. Brooks, C. Enzi, G. Gottlob, N. Henze, M. Herzog, M. Kriesell, W. Nejdl and K. Tomaschewski (2005). "*The personal publication reader*," International Semantic Web Conference, Springer Berlin Heidelberg, pp.1050-1053.

Abel, F., D. Heckmann, E. Herder, J. Hidders, D. Krause, E. Leonardi and K. Van Der Slujis (2009). "*A framework for flexible user profile mashups*," Workshop on Adaptation and Personalization for Web 2.0, Trento, Italy, pp.1-11.

Abel, F., N. Henze, D. Krause and D. Plappert (2008). "*User Modeling and User Profile Exchange for SemanticWeb Applications*." Computer Law & Security Review **448**(1): pp.4-9.

Agichtein, E., E. Brill and S. Dumais (2006). "*Improving web search ranking by incorporating user behavior information*," Proceedings of the ACM SIGIR conference on Research and development in information retrieval, New York, USA, pp.19-26.

Assad, M., D. J. Carmichael, J. Kay and B. Kummerfeld (2007). "*PersonisAD: Distributed, active, scrutable model framework for context-aware services*," International Conference on Pervasive Computing, Springer Berlin Heidelberg, pp.55-72.

Badi, R., S. Bae, J. M. Moore, K. Meintanis, A. Zacchi, H. Hsieh, F. Shipman and C. C. Marshall (2006). "*Recognizing user interest and document value from reading and organizing activities in document triage*," Proceedings of the 11th international conference on Intelligent user interfaces, Sydney, Australia, pp.218-225.

Bae, S., D. Kim, K. Meintanis, J. M. Moore, A. Zacchi, F. Shipman, H. Hsieh and C. C. Marshall (2010). "*Supporting document triage via annotation-based multi-application visualizations*," Proceedings of the 10th annual joint conference on Digital libraries, Surfer's Paradise, Australia, pp.177-186.

Barla, M. (2011). "*Towards social-based user modeling and personalization*." Information Sciences and Technologies Bulletin of the ACM Slovakia **3**(1): pp.52-60.

Bennani, N., M. Chevalier, E. Egyed-Zsigmond, G. Hubert and M. Viviani (2012). "*Multi-application profile updates propagation: a semantic layer to improve mapping between applications*," Proceedings of the 21st international conference companion on World Wide Web, Perth, Australia, pp.949-958.

Bennett, P., R. White, W. Chu, S. Dumais, P. Bailey, F. Borisyuk and X. Cui (2012). "*Modeling the impact of short-and long-term behavior on search personalization*," Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval, Portland, USA, pp.185-194.

Berkovsky, S., T. Kuflik and F. Ricci (2008). "*Mediation of user models for enhanced personalization in recommender systems*." User Modeling and User-Adapted Interaction **18**(3): pp.245-286.

Bishop, C. (2007). "*Pattern Recognition and Machine Learning,*" Springer, New York.

Blei, D., A. Ng and M. Jordan (2003). "*Latent dirichlet allocation*." Machine Learning Research **3**(1): pp.993-1022.

Brusilovsky, P., S. Sosnovsky and O. Shcherbinina (2005). "*User modeling in a distributed e-learning architecture*." User Modeling 2005. A. Liliana, Springer Berlin Heidelberg. **3538:** pp.387-391.

Buscher, G., A. Dengel and L. Van Elst (2008). "*Query expansion using gaze-based feedback on the subdocument level*," Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval, Singapore, pp.387-394.

Buscher, G., L. Van Elst and A. Dengel (2009). "*Segment-level display time as implicit feedback: a comparison to eye tracking*," Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval, Boston, USA, pp.67-74.

Carmagnola, F., F. Cena and C. Gena (2011). "*User model interoperability: a survey*." User Modeling and User-Adapted Interaction **21**(3): pp.285-331.

Carpineto, C. and G. Romano (2012). "*A survey of automatic query expansion in information retrieval*." ACM Computing Surveys **44**(1): pp.1-50.

Casquero, O., J. Portillo, R. Ovelar, J. Romo and M. Benito (2008). "*iGoogle and gadgets as a platform for integrating institutional and external services*," Proceedings of 1st Workshop of Mash-Up Personal Learning Environments, pp.37-41.

Chen, L. and K. Sycara (1998). "*WebMate: a personal agent for browsing and searching*," Proceedings of the second international conference on Autonomous agents, Minneapolis, USA, pp.132-139.

Chen, M., J. Anderson and M. Sohn (2001). "*What can a mouse cursor tell us more?: correlation of eye/mouse movements on web browsing*," Extended abstracts on Human factors in computing systems, Seattle, USA, pp.281-282.

Claypool, M., P. Le, M. Wased and D. Brown (2001). "*Implicit interest indicators*," Proceedings of the 6th international conference on Intelligent user interfaces, Santa Fe, USA, pp.33-40.

Cooke, L. (2006). "*Is the Mouse a" Poor Man's Eye Tracker"?*," Annual Conference-Society for Technical Communication, Las Vegas, USA, pp.252-255.

Dim, E. and T. Kuflik (2012). "*User models sharing and reusability: a component-based approach*," In Proceedigns of User Modeling and Personalization Workshops, Montreal, Canada,

Dumais, S. (2005). "*Latent semantic analysis*." Annual Review of Information Science and Technology **38**(1): pp.188-230.

Fox, S., K. Karnawat, M. Mydland, S. Dumais and T. White (2005). "*Evaluating implicit measures to improve web search*." ACM Transactions on Information Systems **23**(2): pp.147-168.

Gauch, S., M. Speretta, A. Chandramouli and A. Micarelli (2007). "*User profiles for personalized information access*." The Adaptive Web, Springer Berlin Heidelberg**:** pp.54-89.

Gerard, S. (1971). "*The SMART retrieval system: Experiments in automatic document processing*." IEEE Transactions on Professional Communication **15**(1): pp.17-18.

Germanakos, P., N. Tsianos, Z. Lekkas, C. Mourlas and G. Samaras (2008). "*Capturing essential intrinsic user behaviour values for the design of comprehensive web-based personalized environments*." Computers in Human Behavior **24**(4): pp.1434-1451.

Gonzalez, E. and Y. Zhang (2005). "*Accelerating the Lee-Seung algorithm for non-negative matrix factorization."* Houston, USA, Rice University**:** pp.1-13.

Grudin, J. (1994). "*Groupware and social dynamics: eight challenges for developers*." Communications of the ACM **37**(1): pp.92-105.

Guo, Q. and E. Agichtein (2008). "*Exploring mouse movements for inferring query intent*," Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval, Singapore, pp.707-708.

Guo, Q. and E. Agichtein (2012). "*Beyond dwell time: estimating document relevance from cursor movements and other post-click searcher behavior,"* Proceedings of the 21st international conference on World Wide Web. Lyon, France**:** pp.569-578.

Haiqin, Z., M. Zheng and C. Qingsheng (2003). "*A study for documents summarization based on personal annotation,"* Proceedings of the Text summarization workshop. Stroudsburg, USA**:** pp.41-48.

Hall, D., D. Jurafsky and C. Manning (2008). "*Studying the history of ideas using topic models*," Proceedings of the conference on empirical methods in natural language processing, Honolulu, USA, Association for Computational Linguistics, pp.363-371.

Harper, D. J. and D. Kelly (2006). "*Contextual relevance feedback*," Proceedings of the 1st International Conference on Information Interaction in Context, Copenhagen, Denmark, pp.129-137.

Harvey, M., F. Crestani and M. Carman (2013). "*Building user profiles from topic models for personalised search*," Proceedings of the 22nd ACM international conference on Conference on information & knowledge management, Burlingame, USA, pp.2309-2314.

Hijikata, Y. (2004). "*Implicit user profiling for on demand relevance feedback*," Proceedings of the 9th international conference on Intelligent user interfaces, Madeira, Portugal, pp.198-205.

Houben, D. K., E. Leonardi and K. Van Der Slujis (2009). "*A Framework for Flexible User Profile Mashups*," Adaptation and Personalization for Web 2.0, Trento, Italy, pp.1-10.

James, P., S. Hinrich , C. Todd , C. Rob , T. Don , E. Andy , A. Eytan and B. Thomas (2002). "*Personalized search*." Communications of the ACM **45**(9): pp.50-55.

Jansen, B., A. Spink and T. Saracevic (2000). "*Real life, real users, and real needs: a study and analysis of user queries on the web*." Information processing and management **36**(2): pp.207-227.

Jawaheer, G., M. Szomszor and P. Kostkova (2010). "*Comparison of implicit and explicit feedback from an online music recommendation service*," proceedings of the 1st international workshop on information heterogeneity and fusion in recommender systems, Barcelona, Spain, pp.47-51.

Jayarathna, S., A. Patra and F. Shipman (2015). "*Unified Relevance Feedback for Multi-Application User Interest Modeling*," Proceedings of the 15th ACM/IEEE-CS Joint Conference on Digital Libraries, Knoxville, USA, pp.129-138.

Joachims, T., L. Granka, B. Pan, H. Hembrooke, F. Radlinski and G. Gay (2007). "*Evaluating the accuracy of implicit feedback from clicks and query reformulations in web search*." ACM Transactions on Information Systems **25**(2): pp.1-27.

Jones, R. and K. L. Klinkner (2008). "*Beyond the session timeout: automatic hierarchical segmentation of search topics in query logs*," Proceedings of the 17th ACM conference on Information and knowledge management, Napa Valley, USA, pp.699-708.

Kantor, P. B., E. Boros, B. Melamed, V. Meñkov, B. Shapira and D. J. Neu (2000). "*Capturing human intelligence in the net*." Communications of the ACM **43**(8): pp.112-115.

Kelly, D. (2009). "*Methods for evaluating interactive information retrieval systems with users*." Foundations and Trends in Information Retrieval **3**(1): pp.1-224.

Kelly, D. and N. J. Belkin (2001). "*Reading time, scrolling and interaction: exploring implicit sources of user preferences for relevance feedback*," Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval, New Orleans, USA, pp.408-409.

Kelly, D. and N. J. Belkin (2002). "*A user modeling system for personalized interaction and tailored retrieval in interactive IR*." Proceedings of the American Society for Information Science and Technology **39**(1): pp.316-325.

Kelly, D. and N. J. Belkin (2004). "*Display time as implicit feedback: understanding task effects*," Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval, Sheffield, UK, ACM, pp.377-384.

Kelly, D. and J. Teevan (2003). "*Implicit feedback for inferring user preference: a bibliography*," ACM SIGIR Forum, New York, USA, pp.18-28.

Kim, J. Y., K. Collins-Thompson, P. N. Bennett and S. T. Dumais (2012). "*Characterizing web content, user interests, and search behavior by reading level and topic*," Proceedings of the fifth ACM international conference on Web search and data mining, Seattle, USA, pp.213-222.

Kobsa, A. (2007). "*Generic user modeling systems*." The adaptive web, Springer Berlin Heidelberg**:** pp.136-154.

Kobsa, A. and J. Fink (2006). "*An LDAP-based user modeling server and its evaluation*." User Modeling and User-Adapted Interaction **16**(2): pp.129-169.

Kong, W., E. Aktolga and J. Allan (2013). "*Improving passage ranking with user behavior information*," Proceedings of the 22nd ACM international conference on Conference on information & knowledge management, Burlingame, USA, pp.1999-2008.

Kríž, J. (2012). "*Keyword extraction based on implicit feedback*." Information Sciences and Technologies Bulletin of the ACM Slovakia **4**(2): pp.43-46.

Krulwich, B. and C. Burkey (1997). "*The InfoFinder agent: Learning user interests through heuristic phrase extraction*." IEEE Expert: Intelligent Systems and Their Applications **12**(5): pp.22-27.

Lee, D. D. and H. S. Seung (1999). "*Learning the parts of objects by non-negative matrix factorization*." Nature **401**(6755): pp.788-791.

Leonardi, E., F. Abel, D. Heckmann, E. Herder, J. Hidders and G.-J. Houben (2010). "*A flexible rule-based method for interlinking, integrating, and enriching user data*," 10th International Conference on Web Engineering, Vienna Austria, Springer Berlin Heidelberg, pp.322-336.

Limbu, D. K., A. Connor, R. Pears and S. MacDonell (2006). "*Contextual relevance feedback in web information retrieval*," Proceedings of the 1st International Conference on Information interaction in Context, København, Denmark, pp.138-143.

Liu, J. and N. J. Belkin (2010). "*Personalizing information retrieval for multi-session tasks: The roles of task stage and task type*," Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval, Geneva, Switzerland, pp.26-33.

Liu, N. N., E. W. Xiang, M. Zhao and Q. Yang (2010). "*Unifying explicit and implicit feedback for collaborative filtering*," Proceedings of the 19th ACM international conference on Information and knowledge management, Toronto, Canada, pp.1445-1448.

Lu, Z., D. Agarwal and I. S. Dhillon (2009). "*A spatio-temporal approach to collaborative filtering*," Proceedings of the third ACM conference on Recommender systems, New York, USA, pp.13-20.

Majumder, A. and N. Shrivastava (2013). "*Know your personalization: learning topic level personalization in online services*," Proceedings of the 22nd international conference on World Wide Web, Raleigh, USA, pp.873-884.

Marchiori, E. (2013). "*Class dependent feature weighting and k-nearest neighbor classification*." Pattern Recognition in Bioinformatics, Springer Berlin Heidelberg. **7986:** pp.69-78.

Marios, B., P. Efi, G. Panagiotis and S. George (2013). "*Modeling users on the World Wide Web based on cognitive factors, navigation behavior and clustering techniques*." Journal of Systems and Software **86**(12): pp.2995-3012.

Marshall, C. C. and F. M. Shipman (1995). "*Spatial hypertext: designing for change*." Communications of the ACM **38**(8): pp.88-97.

Martinez-Villaseñor, M. d. L., M. Gonzalez-Mendoza and N. Hernandez-Gress (2012). "*Towards a ubiquitous user model for profile sharing and reuse*." Sensors **12**(10): pp.13249-13283.

McCallum, A. K. (2002). "*Mallet: A machine learning for language toolkit*." from mallet.cs.umass.edu.

Mehrotra, R. (2015). "*Topics, Tasks & Beyond: Learning Representations for Personalization*," Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, Shanghai, China, pp.459-464.

Micarelli, A., F. Gasparetti, F. Sciarrone and S. Gauch (2007). "*Personalized search on the world wide web*." The adaptive web, Springer Berlin Heidelberg. **4321:** pp.195-230.

Morita, M. and Y. Shinoda (1994). "*Information filtering based on user behavior analysis and best match text retrieval*," Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval, Dublin, Ireland, pp.272-281.

Moshfeghi, Y., L. R. Pinto, F. E. Pollick and J. M. Jose (2013). "*Understanding relevance: An fMRI study*." Advances in Information Retrieval, Springer Berlin Heidelberg. **7814:** pp.14-25.

Mueller, F. and A. Lockerd (2001). "*Cheese: tracking mouse movement activity on websites, a tool for user modeling*," Extended abstracts on Human factors in computing systems, Seattle, USA, pp.279-280.

Navrat, P. (2012). "*Cognitive traveling in digital space: from keyword search through exploratory information seeking*." Open Computer Science **2**(3): pp.170-182.

Nichols, D. (1998). "*Implicit Rating and Filtering*," Proceedings of the Fifth DELOS Workshop on Filtering and Collaborative Filtering, Budapest, pp.31-36.

Oard, D. W. and J. Kim (2001). "*Modeling information content using observable behavior*," Proceedings of the 64th Annual Conference of the American Society for Information Science and Technology, Washington, USA, pp.481-488.

Paliouras, G., M. Alexandros, C. Ntoutsis, A. Alexopoulos and C. Skourlas (2006). "*PNS: personalized multi-source news delivery*." Knowledge-Based Intelligent Information and Engineering Systems. Bournemouth, UK, Springer Berlin Heidelberg. **4252:** pp.1152-1161.

Pasi, G. (2010). "*Issues in Personalizing Information Retrieval*." IEEE Intelligent Informatics Bulletin **11**(1): pp.3-7.

Pasi, G. (2014). "*Implicit Feedback through User-system Interactions for Defining User Models in Personalized Search*." Procedia Computer Science **39**(1): pp.8-11.

Porteous, I., D. Newman, A. Ihler, A. Asuncion, P. Smyth and M. Welling (2008). "*Fast collapsed gibbs sampling for latent dirichlet allocation*," Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, Las Vegas, USA, pp.569-577.

Price, M., B. Schilit and G. Golovchinsky (1998). "*XLibris: The active reading machine*," CHI 98 Cconference Summary on Human Factors in Computing Systems, Los Angeles, USA, pp.22-23.

Rafter, R. and B. Smyth (2001). "*Passive profiling from server logs in an online recruitment environment*," Workshop on Intelligent Techniques for Web Personalization at the the 17th International Joint Conference on Artificial Intelligence, Seattle, USA, pp.1-8.

Raman, K., P. N. Bennett and K. Collins-Thompson (2013). "*Toward whole-session relevance: Exploring intrinsic diversity in web search*," Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval, Dublin, Ireland, pp.463-472.

Rao, R. (1995). "*The use of Hellinger Distance in graphical displays of contingency table data*." New trends in probability and statistics **3**(1): pp.143-161.

Renda, M. E. and U. Straccia (2005). "*A personalized collaborative digital library environment: a model and an application*." Information processing and management **41**(1): pp.5-21.

Rocchio, J. J. (1971). "*Relevance feedback in information retrieval*." The SMART Retrieval System: Experiments in Automatic Document Processing **1**(1): pp.313-323.

Rodden, K. and X. Fu (2007). "*Exploring how mouse movements relate to eye movements on web search results pages,"* Web Information Seeking and Interaction. Amsterdam, The Netherlands**:** pp.29-32.

Rudinei, G., C. Renan, C.-G. Jose, I. Valter and P. Maria (2004). "*Interactive multimedia annotations: enriching and extending content,"* Proceedings of the 2004 ACM symposium on Document engineering. Milwaukee, USA**:** pp.84-86.

Ruotsalo, T., J. Peltonen, M. Eugster, D. Głowacka, K. Konyushkova, K. Athukorala, I. Kosunen, A. Reijonen, P. Myllymäki and G. Jacucci (2013). "*Directing exploratory search with interactive intent modeling*," Proceedings of the 22nd ACM international conference on Conference on information & knowledge management, Burlingame, USA, pp.1759-1764.

Ruthven, I. and M. Lalmas (2003). "*A survey on the use of relevance feedback for information access systems*." The Knowledge Engineering Review **18**(02): pp.95-145.

Salton, G. and C. Buckley (1997). "*Improving retrieval performance by relevance feedback*." Readings in information retrieval **24**(5): pp.355-363.

Salton, G., A. Wong and C.-S. Yang (1975). "*A vector space model for automatic indexing*." Communications of the ACM **18**(11): pp.613-620.

Saracevic, T. (2007). "*Relevance: A review of the literature and a framework for thinking on the notion in information science* " Journal of the American Society for Information Science and Technology **58**(13): pp.2126-2144.

Sarwar, B., J. Konstan, A. Borchers, J. Herlocker, B. Miller and J. Riedl (1998). "*Using filtering agents to improve prediction quality in the grouplens research collaborative filtering system*," Proceedings of the 1998 ACM conference on Computer supported cooperative work, Seatle, USA, pp.345-354.

Schiaffino, S. and A. Amandi (2004). "*User–interface agent interaction: personalization issues*." International Journal of Human-Computer Studies **60**(1): pp.129-148.

Sela, M., T. Lavie, O. Inbar, I. Oppenheim and J. Meyer (2015). "*Personalizing news content: An experimental study*." Journal of the Association for Information Science and Technology **66**(1): pp.1-12.

Shahnaz, F., M. W. Berry, V. P. Pauca and R. J. Plemmons (2006). "*Document clustering using nonnegative matrix factorization*." Information Processing and Management **42**(2): pp.373-386.

Shen, X., B. Tan and C. Zhai (2005). "*Implicit user modeling for personalized search*," Proceedings of the 14th ACM international conference on Information and knowledge management, Bremen, Germany, pp.824-831.

Shipman, F., H. Hsieh, P. Maloor and M. Moore (2001). "*The visual knowledge builder: a second generation spatial hypertext*," Proceedings of the 12th ACM conference on Hypertext and Hypermedia, Aarhus, Denmark, pp.113-122.

Shipman, F., M. Price, C. Marshall and G. Golovchinsky (2003). "*Identifying useful passages in documents based on annotation patterns*." Research and Advanced Technology for Digital Libraries, Springer Berlin Heidelberg. **2769:** pp.101-112.

Shipman, F. M., H. Hsieh, P. Maloor and J. M. Moore (2001). "*The visual knowledge builder: a second generation spatial hypertext*," Proceedings of the 12th ACM conference on Hypertext and Hypermedia, Aarhus, Denmark, pp.113-122.

Sieg, A., B. Mobasher and R. Burke (2007). "*Web search personalization with ontological user profiles*," Proceedings of the sixteenth ACM conference on Conference on information and knowledge management, Lisbon, Portugal, pp.525-534.

Smaragdis, P. and J. Brown (2003). "*Non-negative matrix factorization for polyphonic music transcription*," IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, USA, pp.177-180.

Song, Y., N. Nguyen, L.-w. He, S. Imig and R. Rounthwaite (2011). "*Searchable web sites recommendation*," Proceedings of the fourth ACM international conference on Web search and data mining, Hong Kong, pp.405-414.

Speretta, M. and S. Gauch (2005). "*Personalized search based on user search histories*," The 2005 IEEE/WIC/ACM International Conference on Web Intelligence, Compiegne, France, pp.622-628.

Stamou, S. and A. Ntoulas (2009). "*Search personalization through query and page topical analysis*." User Modeling and User-Adapted Interaction **19**(1-2): pp.5-33.

Stevens, K., P. Kegelmeyer, D. Andrzejewski and D. Buttler (2012). "*Exploring Topic Coherence over many models and many topics*," Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, Jeju, Korea, pp.952-961.

Tang, N. and V. R. Vemuri (2005). "*User-interest-based document filtering via semi-supervised clustering*." Foundations of Intelligent Systems, Springer Berlin Heidelberg. **3488:** pp.573-582.

Teevan, J., S. Dumais and E. Horvitz (2010). "*Potential for personalization*." ACM Transactions on Computer-Human Interaction **17**(1): pp.4.

Teevan, J., S. T. Dumais and D. J. Liebling (2008). "*To personalize or not to personalize: modeling queries with variation in user intent*," Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval, Singapore, pp.163-170.

Tsiriga, V. and M. Virvou (2004). "*A Framework for the Initialization of Student Models in Web-based Intelligent Tutoring Systems*." User Modeling and User-Adapted Interaction **14**(4): pp.289-316.

Vassileva, J., G. McCalla and J. Greer (2003). "*Multi-agent multi-user modeling in I-Help*." User Modeling and User-Adapted Interaction **13**(1-2): pp.179-210.

Viviani, M., N. Bennani and E. Egyed-Zsigmond (2010). "*A survey on user modeling in multi-application environments*," Advances in Human-Oriented and Personalized Mechanisms, Technologies and Services, Nice, France, pp.111-116.

Vu, T. T., D. Song, A. Willis, S. N. Tran and J. Li (2014). "*Improving search personalisation with dynamic group formation*," Proceedings of the 37th international

ACM SIGIR conference on Research & development in information retrieval, Gold Coast, Australia, pp.951-954.

Wang, B., M. Rahimi, D. Zhou and X. Wang (2012). "*Expectation-Maximization collaborative filtering with explicit and implicit feedback*." Advances in Knowledge Discovery and Data Mining, Springer Berlin Heidelberg. **7301:** pp.604-616.

White, R. W. and G. Buscher (2012). "*Text selections as implicit relevance feedback,"* Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval. Portland, Oregon, USA**:** pp.1151-1152.

White, R. W., W. Chu, A. Hassan, X. He, Y. Song and H. Wang (2013). "*Enhancing personalized search by mining and modeling task behavior*," Proceedings of the 22nd international conference on World Wide Web, Rio de Janeiro, Brazil, International World Wide Web Conferences Steering Committee, pp.1411-1420.

White, R. W. and D. Kelly (2006). "*A study on the effects of personalization and task information on implicit feedback performance*," Proceedings of the 15th ACM international conference on Information and knowledge management, Arlington, USA, pp.297-306.

Xu, W., X. Liu and Y. Gong (2003). "*Document clustering based on non-negative matrix factorization*," Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval, Toronto, Canada, pp.267-273.

Yudelson, M., P. Brusilovsky and V. Zadorozhny (2007). "*A user modeling server for contemporary adaptive hypermedia: An evaluation of the push approach to evidence propagation*," In Proceedings of the 11th International Conference on User Modeling Corfu, Greece, Springer Berlin Heidelberg, pp.27-36.

Zhang, B.-T. and Y.-W. Seo (2001). "*Personalized web-document filtering using reinforcement learning*." Applied Artificial Intelligence **15**(7): pp.665-685.

Zigoris, P. and Y. Zhang (2006). "*Bayesian adaptive user profiling with explicit & implicit feedback*," Proceedings of the 15th ACM international conference on Information and knowledge management, Arlington, USA, pp.397-404.

APPENDIX A

USER STUDY 1 (2013)

A-1. Task procedure

The purpose of this study was to collect ground-truth data in order to evaluate the initial topic modeling, similarity of text content and design of the software application system. The participants were recruited for this study as a **relevance assessor** to read through a set of tasks and identify the relevant content from the given set of documents.

A-2. Task Sheet

Objective: You will be recruited for this study as a **relevance assessor** to read through a set of tasks (5 tasks) and identify the relevant content from the given set of documents (total of 20 web documents). Please be advised that there are no risks associated with participation in this session.

1. Please read the given Search Task and click on the relevant document link.
2. Each document will be opened in the Mozila Firefox Web browser.
3. Use the given WebAnnotate browser plug-in to select color specified per each Search Task
4. Read the document content (read how you normally read a document retrieved from web search).
5. Find the content relevant to the given Search Task from the document content, highlight the section which is relevant and then annotate using the WebAnnotate browser plug-in.
6. Follow the same procedures 1-5 for all the Search Tasks in given list.

	Task	Annotation Color	Document URL
1	What is the current Graduate Program ranking of the Texas A&M College of Engineering?	Green	http://www.theeagle.com/news/local/article_6516f72b-457c-5020-9555-16f47eeee571.html http://en.wikipedia.org/wiki/Texas_A%26M_University http://ogs.tamu.edu/prospective-students/why-am/ http://tamutimes.tamu.edu/2013/03/12/texas-am-engineering-programs-continue-to-rise-in-u-s-news-rankings-other-programs-also-ranked/
2	How the Texas weather is feels like in winter?	Blue	http://en.wikipedia.org/wiki/Climate_of_Dallas http://www.texassegp.org/climate-in-texas.php http://en.wikipedia.org/wiki/Climate_of_Texas http://wwwp.dailyclimate.org/tdc-newsroom/2011/11/winter-weather
3	Python Programming Language	Red	http://en.wikipedia.org/wiki/Object-oriented_programming http://en.wikipedia.org/wiki/Web_development http://en.wikipedia.org/wiki/Programming_language http://www.realpython.com/build-your-own-website-python-vs-ruby/
4	What is Texas A&M 25 by 25 program?	Yellow	http://engineering.tamu.edu/25by25 http://www.theeagle.com/news/local/article_6516f72b-457c-5020-9555-16f47eeee571.html http://engineering.tamu.edu/news/2013/01/23/texas-am-announces-initiative-to-increase-engineering-enrollment-to-25-000-students http://texas.construction.com/texas_construction_news/2013/0123-texas-am-launches-25-by-25-initiative.asp
5	what are the hurricanes that hit Texas after year 1990	Orange	http://en.wikipedia.org/wiki/List_of_Texas_hurricanes_(1980–present) http://www.livescience.com/9594-hurricane-history-texas-top-target.html http://usatoday30.usatoday.com/weather/hurricane/history/whtexas.htm http://www.bounceenergy.com/articles/weather/historical-hurricanes-in-texas

A-3. Consent form

Project Title: (UIMaP) User Interest Modeling & Personalization

You are invited to take part in a research study being conducted by Ukwatta Jayarathna, a researcher from Texas A&M University under the direction of Dr. Frank Shipman at Computer Science & Engineering. The information in this form is provided to help you decide whether or not to take part. If you decide to take part in the study, you will be asked to sign this consent form. If you decide you do not want to participate, there will be no penalty to you, and you will not lose any benefits you normally would have.

Why Is This Study Being Done?

The purpose of this study is to collect ground-truth data in order to evaluate the design of the software application system. You will be recruited for this study as a **relevance assessor** to read through a set of tasks (5 tasks) and identify the relevant content from the given set of documents (total of 20 web documents). Please be advised that there are no risks associated with participation in this session.

Why Am I Being Asked To Be In This Study?

There are no specific selection criteria to be able to participate in this study. You can choose to participate or not participate. This research is voluntary and you have the choice whether or not to be in this research study. However, you must be age 18 or older to participate.

How Many People Will Be Asked To Be In This Study?

30 people will be invited to participate in this study, which will be done at room 232, HRBB, TAMU.

What Are the Alternatives to being in this study?

The alternative to being in the study is not to participate.

What Will I Be Asked To Do In This Study?

You will be given a set of tasks (5 tasks) to read and identify the relevant content from a set of documents (total of 20 web documents). You will be asked to highlight and annotate (using online browser plug-in tool) the relevant content from the given set of documents. Your participation in this study will last up to two hours and includes only one visit.

Are There Any Risks To Me?

The things that you will be doing are no greater than risks that you would come across in everyday life.

Will There Be Any Costs To Me?

Aside from your time, there are no costs for taking part in the study.

Will I Be Paid To Be In This Study?

You will not be paid for being in this study

Will Information From This Study Be Kept Private?

The records of this study will be kept private. No identifiers linking you to this study will be included in any sort of report that might be published. In addition, our study will not contain identifiable information (name, location, contact details etc.). Research records will be stored securely and only the investigators (Ukwatta Jayarathna, Frank Shipman) will have access to the records.

Your name or personal information will not be collected in this study nor will your name or other personal information be associated with any session data collected from you during this study. An anonymous identification number will be assigned to your study session data and these will be kept confidential to the extent permitted or required by law. People who have access to user study data include the Principal Investigator and research study personnel. Representatives of regulatory agencies such as the Office of Human Research Protections (OHRP) and entities such as the Texas

A&M University Human Subjects Protection Program may access user study data records to make sure the study is being run correctly and that information is collected properly.

Who may I Contact for More Information?

You may contact the Principal Investigator Ukwatta Jayarathna (Doctor of Philosophy student, Computer Science), to tell him about a concern or complaint about this research at UKSJayarathna@tamu.edu or by telephone at 512-665-5480. You may also contact the Co-Investigator, Dr. Frank Shipman at shipman@cse.tamu.edu.

For questions about your rights as a research participant; or if you have questions, complaints, or concerns about the research, you may call the Texas A&M University Human Subjects Protection Program office at (979) 458-4067 or irb@tamu.edu.

What if I Change My Mind About Participating?

This research is voluntary and you have the choice whether or not to be in this research study. You may decide to not begin or to stop participating at any time. If you choose not to be in this study or stop being in the study, there will be no effect on your student status, medical care, employment, evaluation, relationship with Texas A&M University, etc. If for any reason you are uncomfortable during the session and do not want to complete a task, you may say so and we will move on to the next task. In addition, if you do not want to continue, you may end the session and leave at any time.

STATEMENT OF CONSENT

I agree to be in this study and know that I am not giving up any legal rights by signing this form. The procedures, risks, and benefits have been explained to me, and my questions have been answered. I can ask more questions if I want. A copy of this entire consent form will be given to me.

________________________________ ________________________________

Participant's Signature Date

Printed Name Date

INVESTIGATOR'S AFFIDAVIT:

Either I have or my agent has carefully explained to the participant the nature of the above project. I hereby certify that to the best of my knowledge the person who signed this consent form was informed of the nature, demands, benefits, and risks involved in his/her participation.

Signature of Presenter Date

Printed Name Date

APPENDIX B

USER STUDY 2 (2014)

B-1. Task procedures and list

Objective: You will be acting as a researcher for this user study. You will be given a specific number of tasks and web pages related to the task. Read through the web documents given for each task and prepare a word document and power point. Here are the detailed instructions. A training will be given on how to use web annotate tool via video or manual demo.

1. Open the Word document and Power Point template for the task.
2. Please read the given Search Task and click on the document link given for that task.
3. Each document will be opened in new Mozilla Firefox Web browser Tab.
4. Read the document content (Read how you normally read a document retrieved from web search).
5. Find the content relevant to the given Search Task from the document content.
6. A drop down menu is displayed to choose a relevance rating from 1-5 for each annotation.
7. Select the most appropriate relevance score (1 being not relevant at all and 5 being highly relevant) and press ok.
8. Choose any color in the web annotate tool and highlight the section which is relevant and then annotate using the WebAnnotate browser plug-in.
9. Edit the word and power point template with information related to the task.
10. You can select as many paragraphs as you like and put as much as content in power point and word.
11. Once you are done with web page close it. Assign readability and document relevance score to each web page (sliders are given next to each document).
12. Repeat from step3-12 for each document in the task.
13. Once all the web documents per each task are read, next, complete the power point and word.
14. Save and close Microsoft Word and PowerPoint application after the completion.
15. Repeat from step 2-14 for each task.
16. At the end please clicks "submit" button at the end to upload your data.

	Search Task	Document List *
1	How does Google glass Work??	http://www.stateofdigital.com/google-glass-explained/
		http://readwrite.com/2013/09/25/first-100-days-with-google-glass#awesm=~onr5uxslUv97hA
		http://thenextweb.com/google/2012/04/13/google-wants-project-glass-to-work-with-your-prescription-glasses/
		http://www.techlife.net/lifestyle/news/2013/7/how-does-google-glass-work/
		http://www.techlife.net/lifestyle/news/2013/7/how-does-google-glass-work/
		http://www.tomsguide.com/us/google-glass,news-17711.html
		http://www.techradar.com/us/news/video/google-glass-what-you-need-to-know-1078114
		http://www.10news.com/news/san-diego-woman-cecilia-abadie-says-she-was-cited-for-driving-with-google-glass-103013
2	What is Mars One Project??	http://www.mars-one.com/en/about-mars-one/about-mars-one
		http://en.wikipedia.org/wiki/Manned_mission_to_Mars
		http://en.wikipedia.org/wiki/Mars_One
		http://www.heavy.com/news/2013/05/mars-one-project-top-10-facts-you-need-to-know/
		http://forums.bistudio.com/showthread.php?161928-Project-Mars-One
		http://www.business-standard.com/article/current-affairs/1-lakh-people-apply-for-a-one-way-trip-to-mars-mission-to-cost-6-bn-113081100212_1.html
		http://www.slate.com/blogs/the_slatest/2013/09/10/the_mars_one_project_receives_more_than_200_000_applications_for_martian.html
		http://motherboard.vice.com/blog/we-asked-mars-one-applicants-why-they-want-to-leave-this-planet-forever
3	How to improve your credit score??	http://www.huffingtonpost.com/neal-frankle/3-junk_b_3880042.html
		http://money.msn.com/credit-rating/9-fast-fixes-for-your-credit-scores-weston.aspx
		http://money.msn.com/credit-rating/9-fast-fixes-for-your-credit-scores-weston.aspx?page=2
		http://www.csmonitor.com/Business/Saving-Money/2012/1201/Eight-surprising-ways-to-raise-your-credit-score
		http://www.marketwatch.com/story/how-to-improve-a-credit-score-1304923724437
		http://www.experian.com/credit-education/improve-credit-score.html
		http://www.myfico.com/crediteducation/improveyourscore.aspx
		http://www.bbb.org/credit-management/balancing-act/improve-your-credit-score/
4	What are the rules of American football??	http://www.topendsports.com/sport/gridiron/basics.htm
		http://www.rulesofsport.com/sports/american-football.html
		http://en.wikipedia.org/wiki/American_football_rules
		http://liveworktravelusa.com/american-football-rules-for-die-hard-soccer-fans/
		http://www.understanding-american-football.com/football-rules.html
		http://news.bbc.co.uk/sport2/hi/other_sports/american_football/3192002.stm
		http://www.infoplease.com/encyclopedia/sports/football-american-football.html
		http://www.ducksters.com/sports/footballrules.php

B-2. Consent form

Project Title: (UIMaP) User Interest Modeling & Personalization

You are invited to take part in a research study being conducted by Atish Kumar Patra and Ukwatta Jayarathna, a researcher from Texas A&M University under the direction of Dr. Frank Shipman at Computer Science & Engineering. The information in this form is provided to help you decide whether or not to take part. If you decide to take part in the study, you will be asked to sign this consent form. If you decide you do not want to participate, there will be no penalty to you, and you will not lose any benefits you normally would have.

Why Is This Study Being Done?

The purpose of this study is to collect ground-truth data in order to evaluate the design of the software application system. You will be recruited for this study as a **researcher** to read through a set of tasks (4 tasks) and identify the relevant content from the given set of documents (total of 32 web documents). You will also be asked to prepare a Microsoft Word document (approximately half a page) and Microsoft PowerPoint (3 slides) related to the task. At the end of each task you will be asked to provide a separate rating for each page based on the readability of the webpage and relevance to the task. Please be advised that there are no risks associated with participation in this session.

Why Am I Being Asked To Be In This Study?

There are no specific selection criteria to be able to participate in this study. You can choose to participate or not participate. This research is voluntary and you have the choice whether or not to be in this research study. However, you must be age 18 or older to participate.

How Many People Will Be Asked To Be In This Study?

30 people will be invited to participate in this study, which will be done at room 232, HRBB, TAMU.

What Are the Alternatives to being in this study?

The alternative to being in the study is not to participate.

What Will I Be Asked To Do In This Study?

You will be given a set of tasks (4 tasks) to read and identify the relevant content from a set of documents (total of 32 web documents). You will be asked to highlight and annotate (using online browser plug-in tool) the relevant content from the given set of documents and assign a score to it(1~5). You will also be asked to prepare a word document (approximately half a page) and PowerPoint (3 slides) related to the task. At the end of each task you will be asked to provide a separate rating (1~5) for each page based on the readability of the webpage and relevance to the task. Your participation in this study may last up to two hours and includes only one visit.

Are There Any Risks To Me?

The things that you will be doing are no greater than risks that you would come across in everyday life.

Will There Be Any Costs To Me?

Aside from your time, there are no costs for taking part in the study.

Will I Be Paid To Be In This Study?

You will be paid $10 for being in this study

Will Information From This Study Be Kept Private?

The records of this study will be kept private. No identifiers linking you to this study will be included in any sort of report that might be published. In addition, the study data will not be linked to identifiable information (name, location, contact details etc.). Research records will be stored securely and only the investigators (Atish Kumar Patra, Ukwatta Jayarathna, Frank Shipman) will have access to the records.

Your name or personal information will not be associated with any session data collected from you during this study. An anonymous identification number will be assigned to your study session data and these will be kept confidential to the extent permitted or required by law. People who have access to user study data include the Principal Investigator and research study personnel. Representatives of regulatory agencies such as the Office of Human Research Protections (OHRP) and entities such as the Texas A&M University Human Subjects Protection Program may access user study data records to make sure the study is being run correctly and that information is collected properly.

Who may I Contact for More Information?

You may contact the Co-Investigators Atish Patra (Masters, Computer Engineering) at atish.patra@tamu.edu, Tel: 979-571-1704, and Ukwatta Jayarathna (Doctor of Philosophy student, Computer Science) at UKSJayarathna@tamu.edu or by telephone at 512-665-5480, to tell them about a concern or complaint about this research. You may also contact the Principal Investigator, Dr. Frank Shipman at shipman@cse.tamu.edu.

For questions about your rights as a research participant; or if you have questions, complaints, or concerns about the research, you may call the Texas A&M University Human Subjects Protection Program office at (979) 458-4067 or irb@tamu.edu.

What if I Change My Mind About Participating?

This research is voluntary and you have the choice whether or not to be in this research study. You may decide to not begin or to stop participating at any time. If you choose not to be in this study or stop being in the study, there will be no effect on your student status, medical care, employment, evaluation, relationship with Texas A&M University, etc. If for any reason you are uncomfortable during the session and do not want to complete a task, you may say so and we will move on to the next task. In addition, if you do not want to continue, you may end the session and leave at any time.

STATEMENT OF CONSENT

I agree to be in this study and know that I am not giving up any legal rights by signing this form. The procedures, risks, and benefits have been explained to me, and my questions have been answered. I can ask more questions if I want. A copy of this entire consent form will be given to me.

___________________________________ ___________________________________

Participant's Signature Date

___________________________________ ___________________________________

Printed Name Date

INVESTIGATOR'S AFFIDAVIT:

Either I have or my agent has carefully explained to the participant the nature of the above project. I hereby certify that to the best of my knowledge the person who signed this consent form was informed of the nature, demands, benefits, and risks involved in his/her participation.

___________________________________ ___________________________________

Signature of Presenter Date

___________________________________ ___________________________________

Printed Name Date

APPENDIX C

USER STUDY 3 (2016)

C-1. Task procedure

Before conducting given tasks, participants will be given a manual demonstration which includes a hands-on presentation of the VKB, WebAnnotate tool and PDF Highlighter tool. Then, they will have an additional 5-minute trial and learning time to practice how to use the system in each mode.

For the study, there are 3 different system modes depending on the availability of recommendations: baseline system without any recommendations, using semi-explicit system (user annotations), and unified system (implicit + semi-explicit), respectively. Table 1 shows evaluation groups which are all permutations of three different system modes (considering the order of 3 system modes). The participants will be randomly assigned to one of the groups. In each group, two system modes will be evaluated and the same two tasks will be asked to the participants in each system mode. The entire assignments to each group will have equal numbers of the participants to be balanced. In brief, after learning about the system, the participants will be asked to perform the two tasks in each system mode according to the group they belong to. They will also complete initial demographic survey (Question set 1), after completion of each task another survey depends on the system mode (Question sets 2, and 3), and finally a general survey about the overall system (Question set 4).

Interest Shift: Change of interest or interest drift will be tested by simulating a sub-task. User will be given a sub-task in the middle of each task in each Mode 2 and 3 to simulate this behavior.

Table 1. User study groups and System Modes

	Tasks 1	**Tasks 2**
Group 1	Mode 1	Mode 2
Group 2	Mode 2	Mode 1
Group 3	Mode 1	Mode 3
Group 4	Mode 3	Mode 1
Group 5	Mode 2	Mode 3
Group 6	Mode 3	Mode 2

System Mode 1: All software tools available. No recommendations

System Mode 2: All software tools available. Only recommendations based on semi-explicit relevance feedback (user annotations and authored text)

System Mode 3: All software tools available. Complete unified recommendations from both implicit and semi-explicit relevance feedback.

C-2. Task Sheet

Objective: You will be acting as a **research librarian** for this user study. You will be given a task and set of tools (VKB, Mozilla Browser enabled with WebAnnotate, Microsoft Word, Microsoft PowerPoint and Adobe Acrobat Writer) to prepare a summary report and a presentation.

Read the two tasks and prepare a summary word document and power point presentation using the tools provided.

Task 1 (about 30 minutes)

What is Mars One Project? *Find information related to Mars One project and prepare a summary Word Document and PowerPoint presentation.*

Task 2 (about 30 minutes)

How to improve your credit score?

Find information related to this topic and prepare a summary Word Document and PowerPoint presentation.

Task Instructions (for Mode 1 – Baseline System)

1. Complete the given survey (Question set 1)
2. Look at the list of documents given in VKB application (8 web documents). To further view each, you can right click on the document and select open from option menu. This will open the document in Mozilla Firefox web browser.
3. You can also utilize the given PDF documents (2 PDF documents) to find information related to the task.
4. Prepare a summary Word document and PowerPoint document using given templates
5. You can copy/paste or write in your own words a summary (few paragraphs) and couple of slides in PowerPoint
6. Save and close both Word and PowerPoint
7. Now complete the given survey (Question sets 2 and 4)

Task Instructions (for Mode 2 Semi-Explicit and Mode 3 Unified systems)

1. Complete the given survey (Question set 1)
2. Look at the list of documents given in VKB application (8 web documents). To further view each, you can right click on the document and select open from option menu. This will open the document in Mozilla Firefox web browser.
    a. If you need automatic recommendations for your documents, click Ctrl+S in Word or PowerPoint
    b. Also if find any relevant content after opening the web browser document, you can utilize WebAnnotate tool to highlight paragraphs using any color.

3. You can also utilize the given PDF documents (2 PDF documents) to find information related to the task.
    a. If you need automatic recommendations for your documents, use highlight tool in PDF and click on the submit button to find relevant content from web documents in VKB or in browser
4. Prepare a summary Word document and PowerPoint document using given templates
5. You can copy/paste or write in your own words a summary (few paragraphs) and couple of slides in PowerPoint
6. Save and close both Word and PowerPoint
7. Now complete the given survey (Question sets 3 and 4)

C-3. Questionnaire

Question Set 1: Demographics Questions

1. Are you male or Female?
    - Male
    - Female

2. Which category below includes your age?
    ☐ 18 ~ 20
    ☐ 20 ~ 25
    ☐ 26 ~ 30
    ☐ 31 ~ 35
    ☐ 36 ~ 40
    ☐ Over 40

3. What is the highest level of school you have completed or the highest degree you have received?
    - Bachelors
    - Masters
    - Doctorate
    - Other (please specify):

4. Amount of computer use per week
    a. Light (less than 10 hours / week)
    b. Moderate (between 10 - 20 hours / week)
    c. Heavy (more than 20 hours / week)

5. Where do you use computer in an average week
    a. Home
    b. School

c. Work
d. Other

6. Which of the following types of computers you use?
    a. PC
    b. Laptop / Notebook
    c. Tablet
    d. Cell Phone
    e. Other

7. Please list names of application software(s) you use in your daily routines
   Ex: Microsoft Word, Adobe Acrobat, Photoshop or any other software

8. How often do you use multiple of these applications in your daily routines
    a. Never
    b. Rarely
    c. Occasionally / Sometimes
    d. Almost Every time
    e. Every time

Question Set 2: System with "No Recommendations" (baseline)

9. List of documents (in VKB and PDF documents) given for the task are relevant
    a. Strongly disagree
    b. Disagree
    c. Neutral
    d. Agree
    e. Strongly agree

10. It was easy to identify relevant web pages from VKB document list
    a. Strongly disagree
    b. Disagree
    c. Neutral
    d. Agree
    e. Strongly agree

11. Please rate how difficult/easy it was for you to use multiple applications for the given task (semantic differential question)

Difficult	1	2	3	4	5	Easy

12. Did you feel overwhelmed by the number of applications available for the given task?
    a. No affect
    b. Minor affect
    c. Neutral
    d. Moderate affect
    e. Major affect

13. What combination of applications do you think helpful when completing the given task?
    a. Web browser
    b. VKB Document list
    c. PDF documents
    d. MS Word
    e. MS Powerpoint

Question Set 3: System with "Recommendations" (for both semi-explicit and unified Systems)

14. List of documents (in VKB and PDF documents) given for the task are relevant
    a. Strongly disagree
    b. Disagree
    c. Neutral
    d. Agree
    e. Strongly agree

15. It was easy to identify relevant web pages from VKB document list
    a. Strongly disagree
    b. Disagree
    c. Neutral
    d. Agree
    e. Strongly agree

16. The recommendations on the VKB web document list are relevant to the task
    a. Strongly disagree
    b. Disagree
    c. Neutral
    d. Agree
    e. Strongly agree

17. The recommendations on the browser web document paragraphs are relevant to the task
    a. Strongly disagree
    b. Disagree
    c. Neutral
    d. Agree
    e. Strongly agree

18. The visualization provided for recommendations were sufficient.
    a. Strongly disagree
    b. Disagree
    c. Neutral
    d. Agree
    e. Strongly agree

19. I was satisfied with the recommendation frequency
    a. Strongly disagree
    b. Disagree
    c. Neutral
    d. Agree
    e. Strongly agree

20. Did the recommendations help you to find interesting content relevant to the given task? Please explain

____

____

____

____

____

21. When you change your intent (interest shift), did the recommendations changed accordingly? Please explain.

____

____

____

22. Please rate how difficult/easy it was for you to use multiple applications for the given task (semantic differential question)

Difficult	1	2	3	4	5	Easy

23. Did you feel overwhelmed by the number of applications available for the given task?
    a. No affect
    b. Minor affect
    c. Neutral
    d. Moderate affect
    e. Major affect

24. What combination of applications do you think helpful when completing the given task?
    a. Web browser
    b. VKB Document list
    c. PDF documents
    d. MS Word
    e. MS Powerpoint

Question Set 4: General Questions

25. Did you have problems using the system? If Yes, please explain

26. Did you feel privacy of your interactions are breached when using the given system? If Yes, please explain

____

____

____

____

____

27. Please rate how difficult/easy it was for you to use multiple applications for the given task
(semantic differential question)

Comfortable 1 2 3 4 5 Uncomfortable

28. Overall, I think multiple applications are useful to find interesting data content via recommendations
    a. Strongly disagree
    b. Disagree
    c. Neutral
    d. Agree
    e. Strongly agree

29. Which System Configuration helps you to find the relevant content?

☐ None ☐ Task 1 System ☐ Task 2 System ☐ Both

30. Any other comments/feedback?

____

____

____

____

____

C-4. Consent Form

Project Title: (UIMaP) User Interest Modeling & Personalization

You are invited to take part in a research study being conducted by Ukwatta Jayarathna, a researcher from Texas A&M University under the direction of Dr. Frank Shipman at Computer Science & Engineering. The information in this form is provided to help you decide whether or not to take part. If you decide to take part in the study, you will be asked to sign this consent form. If you decide you do not want to participate, there will be no penalty to you, and you will not lose any benefits you normally would have.

Why Is This Study Being Done?

The purpose of this study is to evaluate the design of the software application system. You will be recruited for this study as a **research librarian** to read a task (two tasks) and prepare summary report (word document) and summary presentation (power point) related to the task. You will be given a set of tools (VKB, Mozilla Browser with WebAnnotate, MS Word and PowerPoint, Adobe Acrobat PDF) and simultaneously prepare the summary report and power point presentation related to the task. Please be advised that there are no risks associated with participation in this session.

Why Am I Being Asked To Be In This Study?

There are no specific selection criteria to be able to participate in this study. You can choose to participate or not participate. This research is voluntary and you have the choice whether or not to be in this research study. However, you must be age 18 or older to participate.

How Many People Will Be Asked To Be In This Study?

50 people will be invited to participate in this study, which will be done at room 232, HRBB, TAMU.

What Are the Alternatives to being in this study?

The alternative to being in the study is not to participate.

What Will I Be Asked To Do In This Study?

You will be given a questionnaire (8 questions) which you will have to answer which is about demographics and domain knowledge. You will be given a single task to read and will also be asked to prepare a summary report and presentation related to the task. You will be asked to search (using VKB), highlight, and annotate (using online browser plug-in tool). After the task-completion, you will be given a questionnaire (10 questions) which will be related to your experience of using our applications. Your participation in this study will last up to 45 minutes and includes only one visit.

Are There Any Risks To Me?

The things that you will be doing are no greater than risks that you would come across in everyday life.

Will There Be Any Costs To Me?

Aside from your time, there are no costs for taking part in the study.

Will I Be Paid To Be In This Study?

You will not be paid for being in this study.

Will Information From This Study Be Kept Private?

The records of this study will be kept private. No identifiers linking you to this study will be included in any sort of report that might be published. In addition, the study data will not be linked to identifiable information (name, location, contact details etc.). Research records will be stored securely and only the investigators (Ukwatta Jayarathna, Frank Shipman) will have access to the records.

Your name or personal information will not be associated with any session data collected from you during this study. An anonymous identification number will be assigned to your study session data and these will be kept confidential to the extent permitted or required by law. People who have access to user study data include the Principal Investigator and research study personnel. Representatives of regulatory agencies such as the Office of Human Research Protections (OHRP) and entities such as the Texas A&M University Human Subjects Protection Program may access user study data records to make sure the study is being run correctly and that information is collected properly.

Who may I Contact for More Information?

You may contact the Investigator Ukwatta Jayarathna (Doctor of Philosophy student, Computer Science) at UKSJayarathna@tamu.edu or by telephone at 512-665-5480, to tell them about a concern or complaint about this research. You may also contact the Principal Investigator, Dr. Frank Shipman at shipman@cse.tamu.edu.

For questions about your rights as a research participant; or if you have questions, complaints, or concerns about the research, you may call the Texas A&M University Human Subjects Protection Program office at (979) 458-4067 or irb@tamu.edu.

What if I Change My Mind About Participating?

This research is voluntary and you have the choice whether or not to be in this research study. You may decide to not begin or to stop participating at any time. If you choose not to be in this study or stop being in the study, there will be no effect on your student status, medical care, employment, evaluation, relationship with Texas A&M University, etc. If for any reason you are uncomfortable during the session and do not want to complete a task, you may say so and we will move on to the next task. In addition, if you do not want to continue, you may end the session and leave at any time.

STATEMENT OF CONSENT

I agree to be in this study and know that I am not giving up any legal rights by signing this form. The procedures, risks, and benefits have been explained to me, and my questions have been answered. I can ask more questions if I want. A copy of this entire consent form will be given to me.

______________________________ ______________________________

Participant's Signature Date

______________________________ ______________________________

Printed Name Date

INVESTIGATOR'S AFFIDAVIT:

Either I have or my agent has carefully explained to the participant the nature of the above project. I hereby certify that to the best of my knowledge the person who signed this consent form was informed of the nature, demands, benefits, and risks involved in his/her participation.

______________________________ ______________________________

Signature of Presenter Date

______________________________ ______________________________

Printed Name Date