An eHealth Process Model of Visualization and Exploration to Support Improved Patient Discharge Record Understanding and Medical Knowledge Enhancement

Harvey Hyman[1] and Warren Fridy[2]

[1]Florida Polytechnic University, Lakeland, Florida, USA
[2] H2 & WF3 Research, LLC., Tampa, Florida, USA

hhyman@Floridapolytechnic.org, Warren@H2WF3.com

Abstract. In this paper we examine a two part information retrieval (IR) problem presented in Task 1, of how to design a visual interactive system, to foster better patient understanding of terminologies and vocabularies contained in a discharge summary, and how that system can be used to additionally support the patient's information retrieval need stemming therefrom. To address this problem set, we apply an IR process model, designed to support context learning and knowledge discovery, based on explicit-implicit and exploration-exploitation schemes. We instantiate the process using an IT artifact (RetrivikaTM) designed to support the search of high volume, context oriented IR collections. The artifact has been built to support the process model, and has been previously validated by Hyman and Fridy in the IR domain of eDiscovery [1].

Keywords : Information Retrieval, Medical Retrieval, Exploration, Exploitation, Explicit, Implicit, IR Process Model.

1 Introduction

We initialize the problem space with the operational definition of Information Retrieval (IR) as the process of determining the presence or absence of relevant documents that satisfy an information need [2]. The problem space is motivated by the increased reliance upon digital documentation to record everyday information such as business transactions, agreements, medical records, and other information stored electronically. This increased reliance has led to large volume collections from which relevant documents must be extracted. In this research we are focused upon medical discharge summaries and amplification of patient knowledge and understanding.

Prior research has found that IR domains which are highly context and content dependent can lead to under inclusion of relevant documents and over inclusion of non-relevant documents, resulting in poor performance when using automated methods

alone [3], [4]. We define the problem set in Task 1 as a context and content dependent IR scenario.

2 Approach

We begin our approach by classifying the problem set into two distinct needs: knowledge and explanatory. We describe the first problem of how to improve patient understanding of the discharge summary as a knowledge need. We define a knowledge need as a situation whereby a user possesses information that is required to be better understood. We model this first part as an explicit-implicit knowledge problem.

Explicit knowledge represents information that is common knowledge or readily accessible to the layman. It is easily codified in written form and can be found in manuals, documents, and various web media outlets (links, pages, etc.). Implicit knowledge, on the other hand, represents information that is not commonly known. Its meaning is often based upon specialized knowledge of a narrowly focused community of experts in the area. This type of knowledge is sometimes called tacit knowledge [5]. Examples of terminologies that are implicit in their nature are local vocabularies, jargon and slang expressions, unique to the specific domain of operation [1]. Quite often implicit knowledge is acquired through specialized training and experience within the specified domain.

We categorize the terminologies in the discharge summary as implicit, insofar as their usage is operationalized as common parlance of the experts (doctors, nurses and health professionals) and thereby outside the knowledge base of the layman patient. The system objective here is to convert the implicit to the explicit, to achieve the stated goal of better patient understanding. In this case, expanding the medical terminologies from the discharge summary is accomplished through the use of a codified (explicit) knowledge base: UMLS and SNOMED-CT.

The methodology used for converting the implicit to the explicit is the IR Process Model first proposed by Hyman et al., and the Retrivika™ IT artifact [6], [7]. The model is implemented using a human-computer interface, to facilitate the translation of implicit knowledge recorded in the discharge summary to explicit knowledge for the purpose of fostering better patient understanding.

We describe the second problem of how to support a patient's information retrieval need as explanatory in nature. We define an explanatory need as a situation whereby a user (in this case a patient) desires to amplify information about a specific topic (in this case a condition contained in a discharge summary). We model this second part as an exploration-exploitation problem described in the Foundation section of this paper.

3 Foundation

Exploration is an underlying construct representing the human search behavior [8], [9]; it is operationalized in electronic search as browsing. The concept of exploration

has been associated with learning [10], [11], familiarization [12], and information search [13]. In fact, work done by Berlyne in the 1960s classifies exploration as a "fundamental human activity" [14].

Exploration that is goal directed is classified as extrinsic [15]. Extrinsic exploration typically has a specific task purpose, whereas intrinsic exploration is motivated by learning [15], [14].

The exploration-exploitation dilemma describes the decision to focus attention and commit resources to the current selection versus abandoning it in favor of searching for a new selection; hopefully bettering one's position, but unknown until explored [7].

Browsing as an information seeking process has been established as a method when the information need is ill-defined [16], [17]. Browsing has been described as a fundamental information seeking function [16], [17], [18], [19].

Holschler and Strube, examined the types of knowledge and strategies involved in web-based information seeking [8]. They found that users with higher levels of knowledge were more flexible in their approaches and were better able to tackle search problems than those who were less knowledgeable. They characterize the information space as "diverse and often poorly organized content."

The IR process model and artifact discussed in this paper seek to organize the information need stemming from the discharge summary around the subject matter contain therein. Holschler and Strube's finding that experts can outperform less experienced users is a fundamental assumption for evaluating whether knowledge acquired by exploration can improve a user's ability to tackle the search problem of information amplification. We specifically address this issue in the process model section of the paper.

4 Assumptions

There are several assumptions defined in this case. The first assumption is that the information presented should contain some form of hierarchical clustering method for categorization and sorting of the relevant documents extracted from a large corpus, but not so much that it confuses the layman, who may not be exposed to common clustering and sorting methods such as trees and visualization clusters.

The second assumption is that each document may contain text, images and links that need to be displayed in some rank order method.

The third assumption is that the system should contain a visual interactive display component that allows a user to navigate freely and easily among levels of the hierarchical document clusters.

The fourth assumption is that the user (in this case a patient) has a point of focus from which their information need stems. For example, a discharge summary may contain a diagnosis described using complex implicit terms that the patient wishes to translate to explicit. It is the underlying assumption of a focused starting point that drives the process as described in the next section.

The fifth assumption is that the user (once again, the patient) will follow an exploration-exploitation methodology (as described in the Foundation section) to achieve their goal of better understanding by leveraging external information sources (web sites and links).

Not all assumptions are addressed in this paper. Some are too complex to handle up front and will be addressed in later versions of the artifact.

5 Process Model

We apply an IR process model designed to support user learning and knowledge discovery to achieve an improved visual display to highlight implicit concepts, assist in the explanation of context and support the exploration of medications, conditions and health related topics for possible interactions with everyday items.

The model was originally built upon the IR constructs of uncertainty, context and relevance to support user driven learning, by leveraging explicit knowledge to discover implicit knowledge from a large corpus of documents. In this case, we reverse the model by converting the implicit knowledge contained in the discharge summaries to explicit knowledge, by leveraging the internal, bounded collections of UMLS, SNOMED-CT and external scale free search (web page contents from provided URLs). We instantiate the adapted process model to support the exploration-exploitation system application.

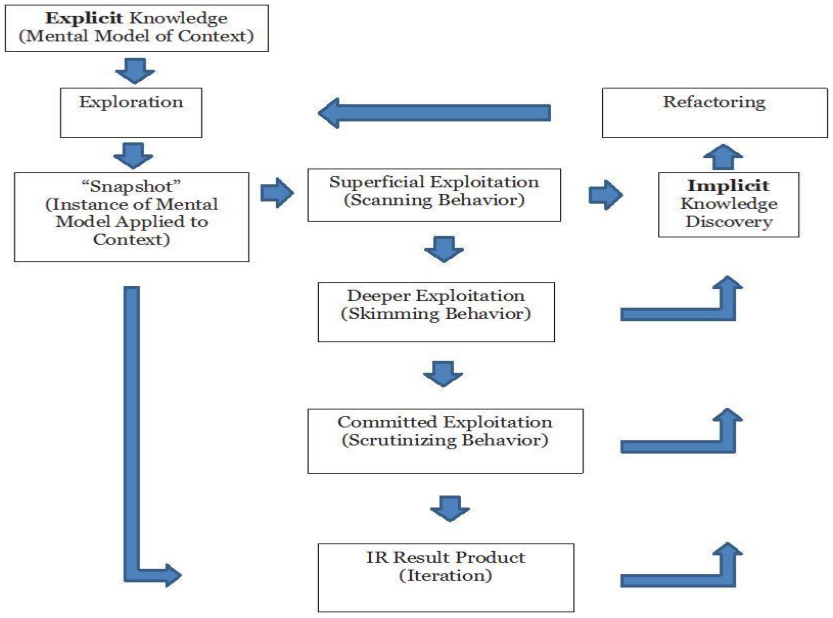

Fig. 1. IR Context Learning Process Model (Hyman et al.)

The IR Process model originally proposed by Hyman et al., describes how a user's mental model of relevance (information sought) can be matched against candidate documents, by applying an iterative and cyclic method of the three levels of exploitation found in search behavior [6]. The iterative process is designed to leverage known, explicit knowledge to discover implicit knowledge found in a bounded information collection.

We have adapted the process to support two activities in this case. The first is to take known, implicit terminologies and compare them against internal lexicons and taxonomies (SNOMED and UMLS) to translate the terms to the explicit. The second is to support user exploration of external information through the use of the Snapshot™ artifact to translate the user's mental model of relevance (in this case a useful document that is informative on the subject), and produce suggested document matches. This is explained in the next section.

6 Original Snapshot™ Method

We have developed a presentation method we call Snapshot™. The method was first proposed by Hyman and Fridy during their development of the commercial artifact Retrivika™. The Snapshot™ presentation makes use of a document list and reading pane, with the user's search structure displayed above. The documents can be clustered by topic or arranged in a hierarchical display. The user scans the document list for the most likely relevant titles. Once a title is selected by the user, he/she may skim the document using a reading pane. The user may become further committed and scrutinize the document, by selecting on search terms presented in the top portion of the screen. The selected search terms are highlighted within the displayed document. Our design emphasizes use of different colors to communicate categories of information across the several window pane displays. The Snapshot™ presentation method illustrated in Figure 2, shows how the query terms, hierarchical listing of documents, and highlighted selections are presented across the several window panes displayed in the presentation screen. Our system design seeks to balance the multiple levels of information amplification with an integrated means for user consumption.

We will now describe the Snapshot™ appearing in Figure 2; it is designed to bring together several dimensions of exploratory search methodologies in one screen. The reader will note that there are two landscape text boxes at the top of the screen display. These text boxes represent the user's current search structure. The search structure is bifurcated into inclusive search terms (indicated with green underline) and exclusive search terms (indicated with red underline). Our prior research has found that the use of exclusive terms is positively correlated with fewer false positives (increased precision in the search result).

The main body of the screen contains two panels. The left panel displays a list of the returned documents by their titles. The right panel displays the document selected from the list. We have enabled a find function so that the user may click on a term in the search structure from above and the term will be highlighted within the selected

document. Our research has found that the use of this find function supports the deeper, scrutinizing behavior described earlier in the paper.

An additional element in the Snapshot™ that has not been carried over to the adaptation here is the relevancy radio buttons. In our eDiscovery implementation, we leveraged relevancy feedback to refactor our results presented in the next iterated Snapshot™.

Here is how the system works. Prior research has shown that, when a user finds multiple documents he or she will tend to switch back and forth, between items; this activity can be supported via an iterative approach to information seeking [17].

Our previous experiments have found that three levels of search described in the literature as exploratory, window, and evolved [20], [9], can be modeled as search behaviors representing scanning, skimming and scrutinizing [6]. Our model further defines these behaviors as: Superficial, Deeper and Committed. The model harmonizes both top-down and bottom-up approaches [21], to provide support for the three levels of search by implementing a multi-tiered and iterative, cyclic method.

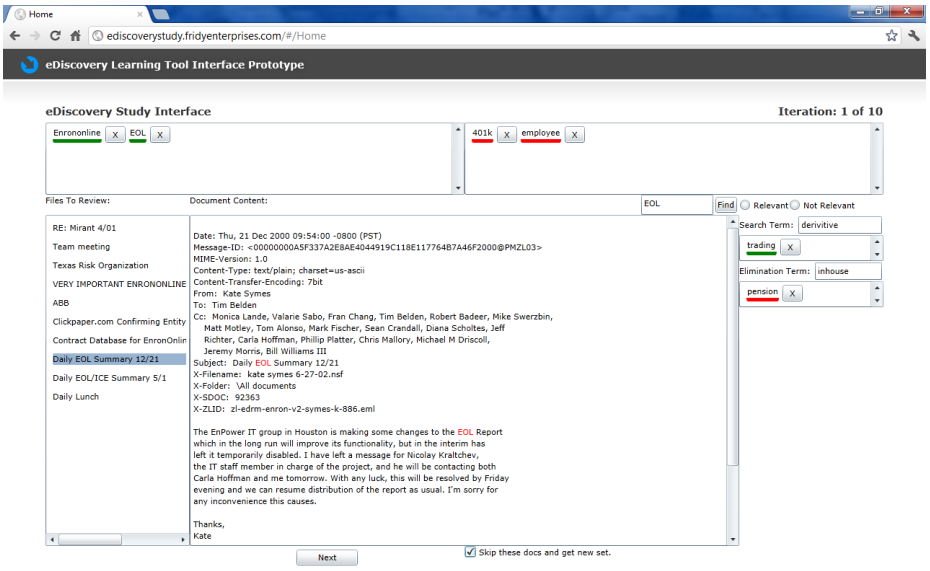

Fig. 2. Original Snapshot™ Method Using Data from the Enron Collection

The Retrivika™ artifact which instantiates the model is based on a method of learning [13], [22], adapted from Active Learning [23], using relevance feedback [24], balancing exploration-exploitation in an iterative cycle. We adapted the learning method for the Snapshot™ approach by shifting the focus of the learner. The traditional active learning technique is based on machine learning -- the system "learns" the patterns and improves performance. In this case, it is the user who is learning; the system simply supports the process.

7 Discussion of Initial Designs

In this section we will take the reader through the development of our approach. We began with several guiding principles for user interface (U/I) design [25], to implement our system display scheme for presentation of information in this case. They are as follows: (1) Functions visible only when the user needs them, (2) Reduced need for horizontal scrolling, (3) Effective use of 'gutter space,' (4) Information to screen ratio, and (5) Minimum clicks. Our initial prototype design screens are depicted in Figures 3 through 8 in this section, along with the narrative descriptions of how we implemented the guiding principles.

To implement this presentation, we wrote a simple program application to load the Clef Task 1 data set into a SQL database. This allowed us to manipulate the rendering of the discharge documents, to include a highlighting feature to support scanning behavior, to assist in the presentation of the embedded medical terminologies within the discharge summaries. We began with a simple and clean window to display the discharge summary with a search box at the top of the screen. This is depicted in Figure 3 below.

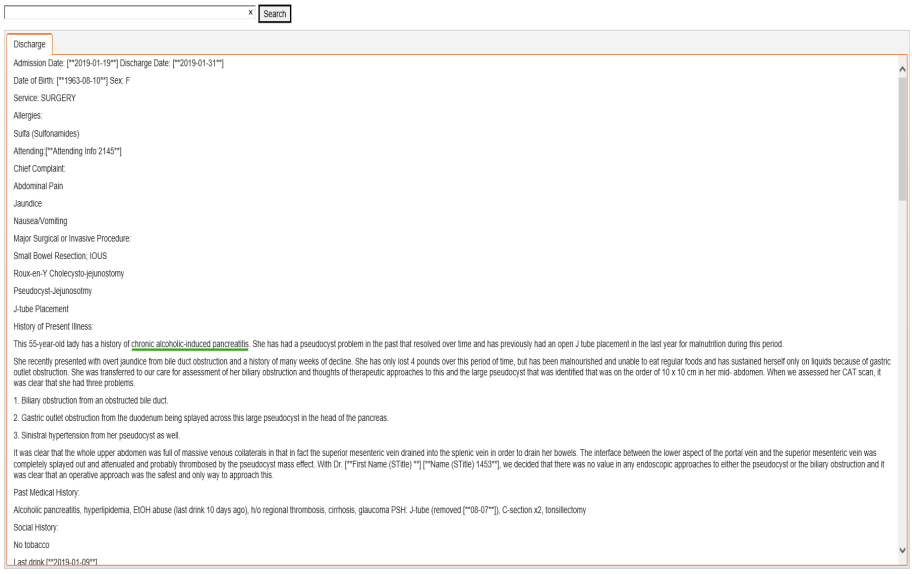

Fig. 3. Initial Rendering of Discharge Summary, Terminology Highlighting and Search Box

Next, we included a hover method to implement a call out feature to provide amplifying information for the selected medical terminology. In the example depicted in Figure 4, we are focusing on the condition of chronic alcohol-induced pancreatitis.

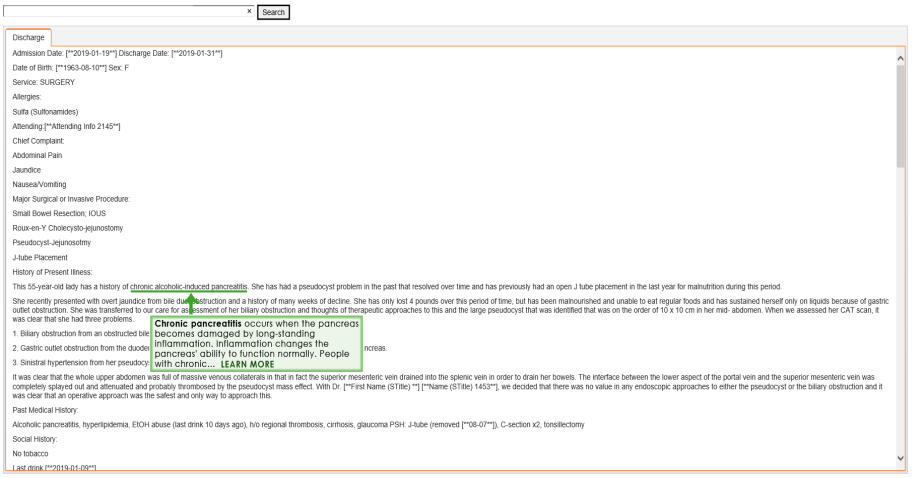

Fig. 4. Discharge Summary with Added Mouse Hover Feature

Next, we expanded the mouse hover feature to include the display of external information links to support the patient's knowledge acquisition (amplification) need. This added feature is displayed in Figure 5.

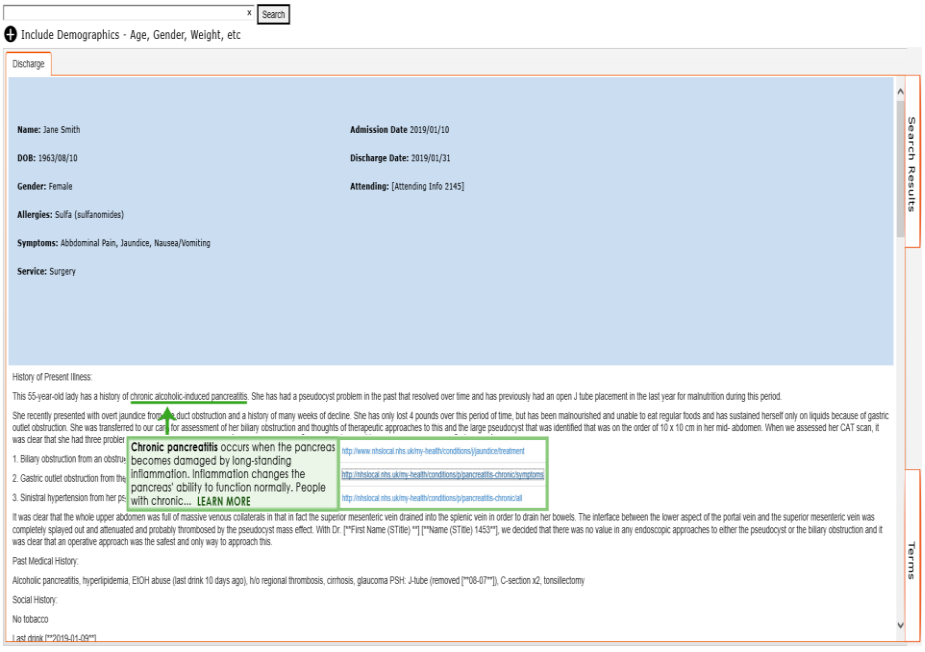

Fig. 5. Mouse Hover Feature with Incorporation of External Link Information

A close up rendering of the mouse hover feature with the incorporation of the external links is depicted in Figure 6.

Fig. 6. Close up View of Mouse Hover Feature with Additional of External Links

We also experimented with a collapsing window feature to accommodate all information activities on one screen and thereby avoid the need for the user to switch between multiple windows or screens. When the user submits a search request, the display screen reduced the space of the discharge summary display in the window to accommodate simultaneous viewing of the discharge document alongside a window pane containing the clustered, hierarchical list of URLs comprised of potentially amplifying information sources for the user to further select. This is depicted in Figure 7. At this point in our research, we have not yet been able to tackle a ranking method for the list. We will continue to work on that aspect in our next set of experimental designs.

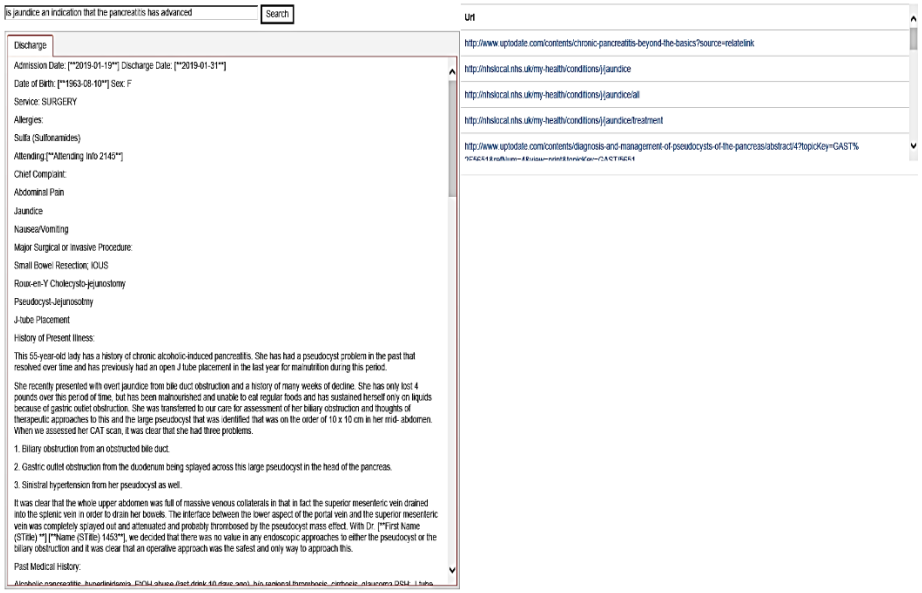

Fig. 7. Display of Discharge Summary, External Search Results, Collapsing Window Effect

Next, we experimented with how to implement the collapsing window feature to support three window panes to display the discharge summary, hierarchical list of documents and the amplified information for a user selected terminology. This is depicted in Figure 8.

Our choice of design in this instance seeks to minimize the need for horizontal scrolling. We implement a single screen display designed to support the user's ability to shift between exploration and exploitation, and back again, without having to navigate to a different screen or negotiate multiple windows. We have found that this reduces the number of clicks needed to acquire information and also provides a balanced ratio of information to screen proportional space.

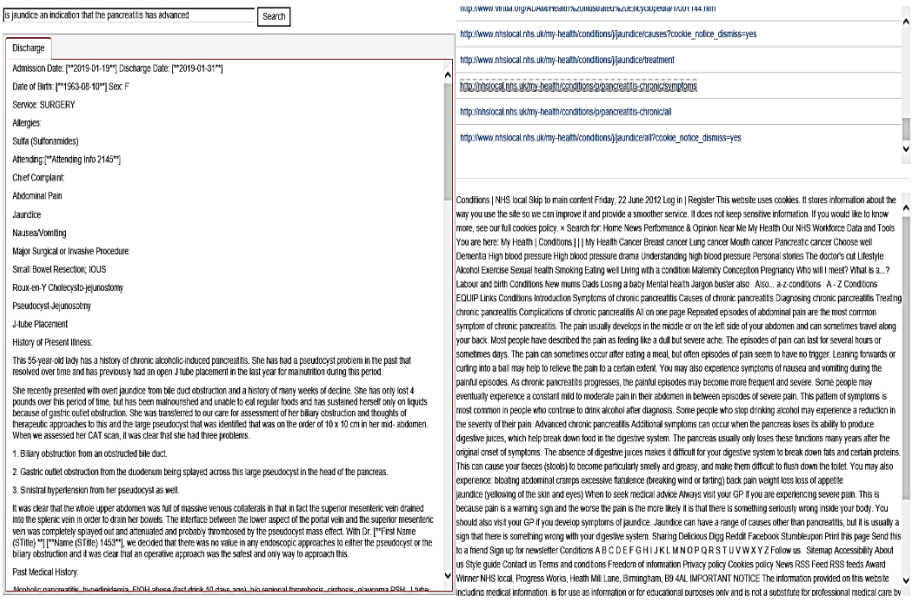

Fig. 8. Display of Three Information Sources using Collapsing Window Approach

We continued to develop the single screen design in Figure 9. This depiction is an illustration of the coordinated information amplification display utilizing the three window pane feature. The left side pane displays the original discharge summary with the highlighted medical term. The upper right pane displays the URL page links in ranked order of significance (ranking method not implemented in this paper). The lower right pane presents the amplifying information for the user (patient) highlighted medical term.

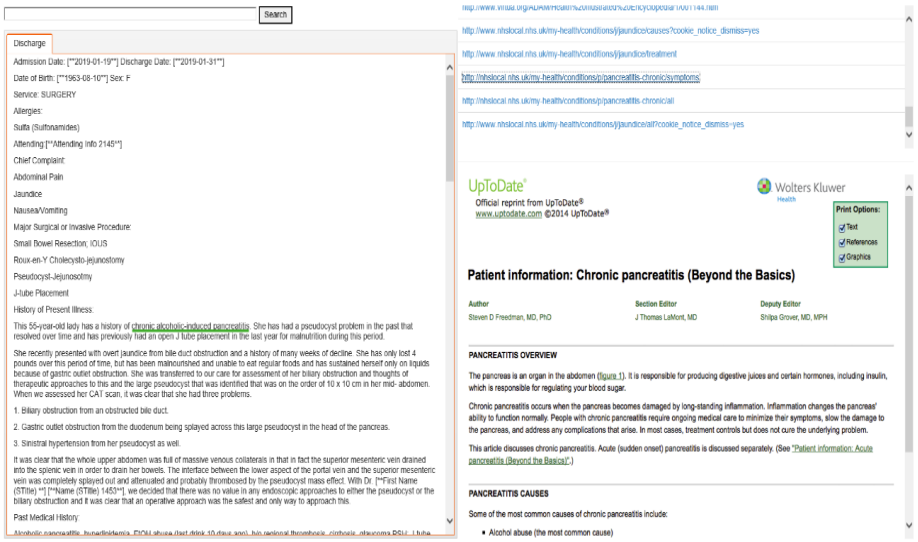

Fig. 9. Display of Discharge Summary, URL Links, Content from User Selected URL Source

8 Adaptation of IR Snapshot™ Method to eHealth

This section will describe our adaptation of the IR Process Model and Snapshot™ to the CLEF Task 1 problem sets A and B, and present exemplars depicted over several figures with accompanying narratives.

Our first adaptation was how we displayed the search structure feature itself. This modification is depicted in Figure 10. The feature was originally developed for semi-expert search of a bounded corpus, where terminologies were not standardized vocabularies. The patient information need in this case is based on standardized vocabularies (medical terminologies from the discharge summary), and the search is bifurcated into internal and external corpora. The internal corpus (SNOMED or UMLS) is bounded, but the external corpus may be scale free (web pages and links).

To address this difference in search structure application, we modified the feature to account for the bifurcated nature of the internal versus the external orientation of the information need by implementing two new functions: Search Results and Terms. In Figure 10, the reader will note the two tabs named Search Results and Terms, located on the right side of the display screen. The Search Results function displays external content to support the patient's information goal of amplification through knowledge acquisition. The Terms function supports the patient's knowledge explanation goal using content from the internal, bounded corpora such as SNOMED and UMLS.

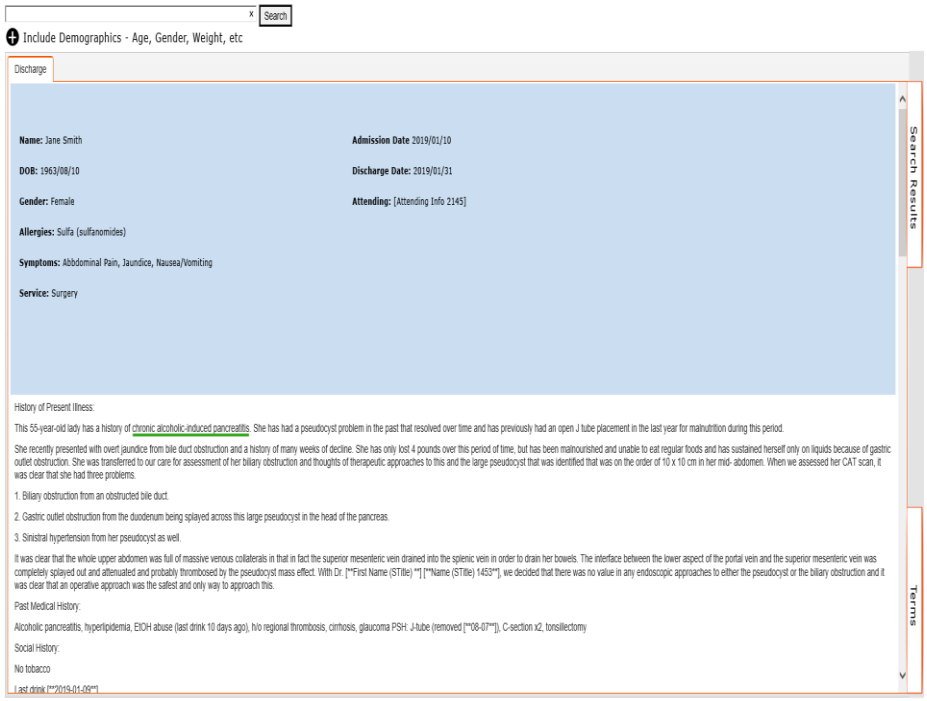

Fig. 10. Modified Snapshot Feature Supports Discharge Summary and Terminology Search

Our second adaptation was the need to account for the information result being effected by the individual attributes of the patient. To account for this, we maintained the inclusive search box feature at the top of the screen and added a feature for the patient to concatenate to the search structure, their individual demographics contained within the discharge summary document. We include a Plus icon, to allow the patient to toggle between including individual attributes and ignoring the attributes.

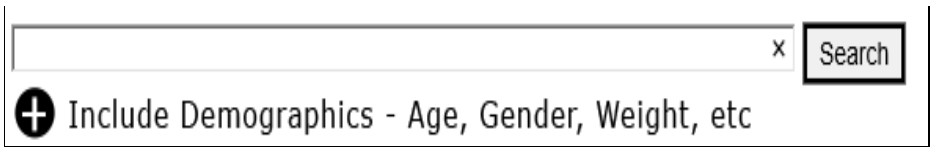

Fig. 11. Modified Search Box to Include "*Plus"* Icon/Toggle Feature for Demographics

Our next adaptation was the use of tab functions for the Search Results information and the Terms information. This was based upon our initial experiments and feedback from reviewers. We continued to adapt the Snapshot™ method to the discharge summary documents in this problem set. We next discuss the evolution of our approach.

As we ran through our simulations we continued to modify our collapsing screen approach. Figure 12 and Figure 13 depict the modified Snapshot™ method implementing the collapsing window approach to display the internal and the external information sources implemented, using the tab functions Search Results and Terms.

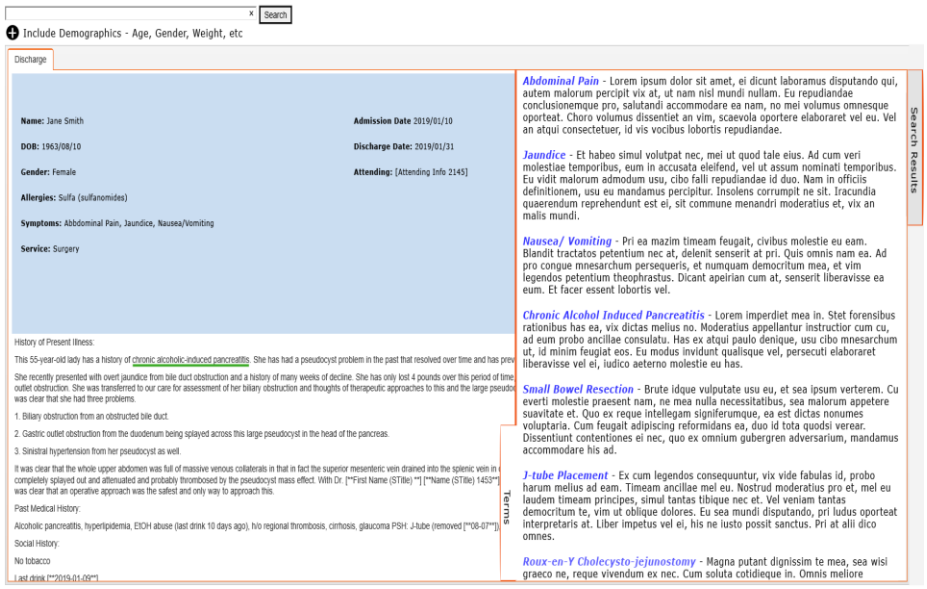

Fig. 12. Snapshot™ Method Adapted for eHealth with Collapsing Window for *Terms*

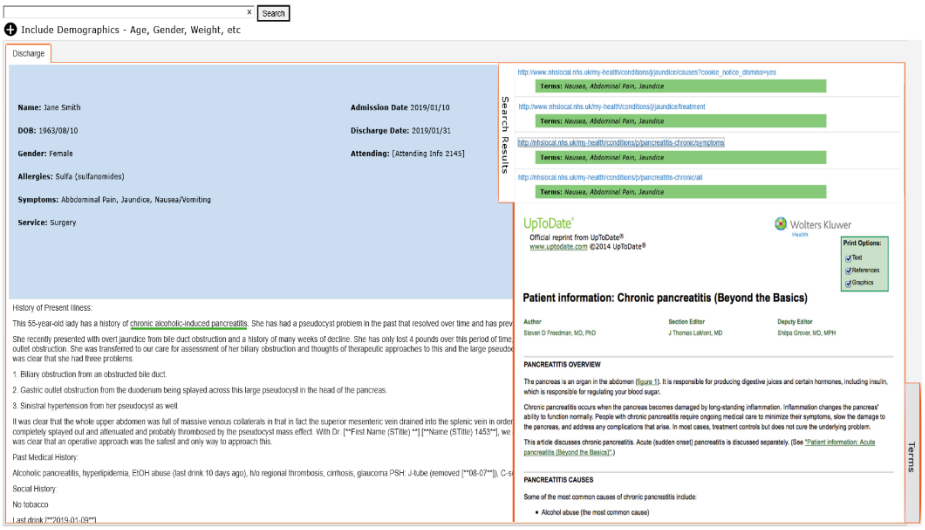

Fig. 13. Snapshot™ Method Adapted for eHealth with Collapsing Window for *Search Results*

Figure 14 and Figure 15 depict the presentation of the Search Results information and the Terms information in their current, modified form. The reader will note that we continue to employ a collapsing window approach to allow the user to remain on one screen and inside a single window, and we have added the tab functions to facilitate the effect.

We believe this supports a more ergonomic method for the patient to keep track of the three forms of information being explored and presented: discharge summary, external search acquisition (amplification) information, internal search explanation information.

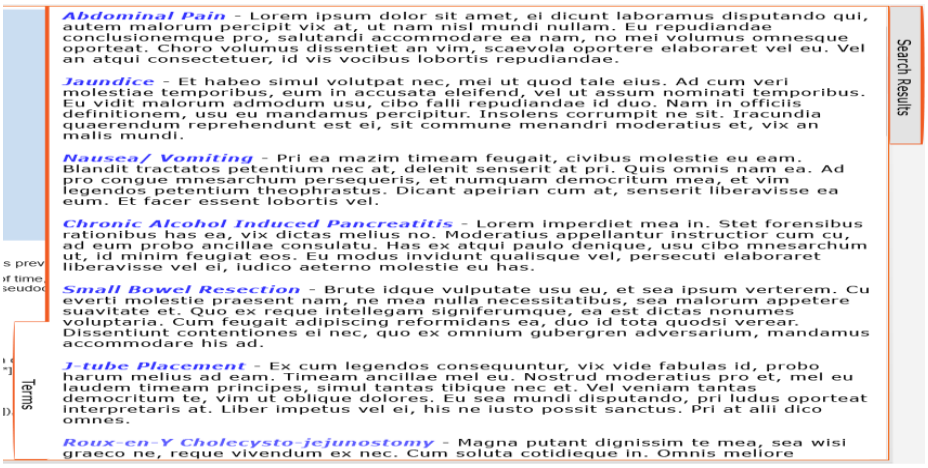

Fig. 14. Expansion of *Terms* Tab, Collapsing Window to Shrink Discharge Summary

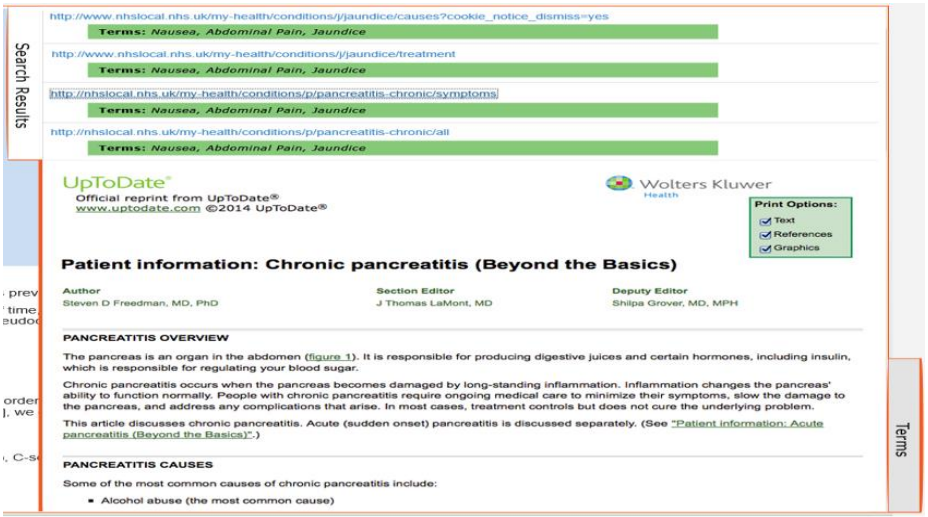

Fig. 15. Expansion of *Search Results* Tab, Collapsing Window to Shrink Discharge Summary

9 Results

Our goal in this paper was to design and test a methodology for a framework to improve patient understanding of the contents of a discharge summary. We divided the goal into two objectives: increasing patient understanding and expanding patient knowledge.

We wanted to expand a patient's knowledge base without losing fidelity in the information retrieved. To accomplish this, we studied how users of the system (patients) might formulate their information need. We found that, in general, a patient will review their discharge document, and when they had come across a term they did not understand, the immediate response was to seek out an explanation. This was achieved through the use of the mouse hover as a presentation technique for the UMLS/SNOMED definitions. If a patient needed more information, they would choose to select on one of the links presented as an associated external source for the term.

Our original studies implementing the Snapshot^{TM} for Legal-IR produced significant results supporting improved document retrieval in bounded collections. In this adaption of the model for Medical-IR, our limited testing conducted thus far has produced encouraging results. We believe further development of this approach may continue to improve patient understanding of information contained in discharge summaries by supporting the patient in conducting external information search to amplify knowledge of a term beyond the explicit definition supplied by internal reference corpora, and thereby explain conditions and medical concepts using external information sources. This two pronged approach addresses the content within the discharge summary and the context of the implicit (tacit) medical terminology requiring explanation AND amplification.

10 Considerations

Our initial results in this study have led us to some further considerations. First, we found that the more we personalized the search feature, the system began to over fit the patient's attributes during the acquisition on external information sources. To address this we measured the retrieval results using the context and attributes from the discharge summary and the retrieval results without using the context and attributes. This allowed us to isolate the patient's ability to decouple the discharge specific content for the external search query. We are still analyzing the data returned, and plan to further study this phenomenon.

The second consideration we found was that future applications of this model need to account for a vetting process for the external links. In this case we used the references common to the CLEF task. To make this system model more generalizable we plan to work on a vetting method to assure reliability of the external information sources.

Another consideration had to do with the callout feature itself. We found in this study that pulling the relevant medical terms from the discharge summary upon opening, was the most effective means of indexing against the internal corpora (SNOMED and UMLS).

11 Conclusion

This paper reports on an IR Process Model and an approach called Snapshot™ which have been adapted from Legal-IR and modified for Medical-IR, to address the CLEF eHealth Evaluation Lab 2014, Task 1, A & B. We introduced the IR process model and Snapshot™ artifact previously implemented in the domain of eDiscovery, and have applied it to the Task 1 problem stated and the data set provided. We welcome feedback and suggestions for how we can improve our approach and methods, and are interested in collaborating with other researchers to continue to address ways to improve patient understanding. Correspondence is best done through the email addresses listed at the beginning of this paper.

12 References

1. Hyman, H. S., Fridy III, W., "Using Exploration and Learning for Medical Records Search: An Experiment in Identifying Cohorts for Comparative Effectiveness Research," NIST Special Publication, Proceedings: Text Retrieval Conference (TREC) 2012.
2. Van Rijsbergen, C. J, Information Retrieval, Butterworth, London, Boston. 1979.
3. Grossman, M., R., Cormack, G., V., "Technology-Assisted Review in E-Discovery Can Be More Effective and More Efficient Than Exhaustive Manual Review," Richmond Journal of Law and Technology, Volume 27, Issue 3 (2011).
4. Oard, D. W., Baron, J. R., Hedin, B., Lewis, D. D., Tomlinson, S., "Evaluation of Information Retrieval for E-discovery," Artificial Intelligence and Law, 18:347 (2010).

5. Polyani, Personal Knowledge: Towards a Post Critical Philosophy, London: Routledge (1958).
6. Hyman, H. S., Sincich, T., Will, R., Agrawal, M., Fridy, W., Padmanabhan, B., "A Process Model for Information Retrieval Context Learning and Knowledge Discovery," (under review).
7. Hyman, H. S., Applied Information Science Approaches for Technology and Business Processes, (Release, 2015).
8. Holscher, C., Strube, G., "Web Search Behavior of Internet Experts and Newbies," Cite as: www9.org/w9cdrom/81/81.html.
9. Muylle, S., Moenaert, R., Despontin, M., "A Grounded Theory of World Wide Web Search Behavior," Journal of Marketing Communications, Available Online (09 Dec 2010).
10. Berlyne, D., E., "Motivational Problems Raised by Exploratory and Epistemic Behavior," Psychology: A Study of Science, Vol. 5, pp. 284-364, New York: McGraw Hill (1963).
11. March, J., G., "Exploration and Exploitation in Organizational Learning," Organizational Science, 2(1), (1991).
12. Barnett, S., A., A Study in Behavior. London: Methuen (1963).
13. Debowski, S., Wood, R., E., Bandura, A., "Impact of Guided Exploration and Enactive Exploration on Self-Regulatory Mechanisms and Information Acquisition Through Electronic Search," Journal of Applied Psychology, Vol. 86, No.6, (2001).
14. Demangeot, C., Broderick, A., J., "Exploration and Its Manifestations in the Context of Online Shopping," Journal of Marketing Management, Vol. 26, No. 13 – 14, (December, 2010).
15. Berlyne, D., E., Conflict, Arousal and Curiosity, New York: McGraw Hill (1960).
16. Kuhlthau, C., C., "Inside the Search Process: Information Seeking from the User's Perspective," Journal of the American Society for Information Science, Vol. 42, 361-371 (1991).
17. McKay, D., Shukla, P., Hunt, R., Cunningham, S., J., "Enhancing Browsing in Digital Libraries: Three New Approaches to Browsing in Greenstone," International Journal of Digital Libraries, (2004).
18. Bates, M., J., "Information Search Tactics," Journal of the American Society for Information Science, July, (1979).
19. Bates, M., J., "The Design of Browsing and Berry Picking Techniques for the Online Search Interface," Online Review, 13 (5), (1989).
20. Broder, A., "A Taxonomy of Web Search," IBM Research, SIGIR Forum, Vol. 36, No. 2, (Fall, 2002).
21. Navarro-Prieto, R., Scaife, M., Rogers, Y., "Cognitive Strategies in Web Searching," Cited as: zing.ncsl.nist.gov/hfweb/proceedings/Navarro-Prieto/index.html. (June 3, 1999).
22. Hills, T., T., "The Central Executive as a Search Process: Priming Exploration and Exploitation Across Domains," Journal of Experimental Psychology, Vol. 139, No. 4, (2010).
23. Zheng, Z., Padmanabhan, B., "Selectively Acquiring Customer Information: A New data Acquisition Problem and an Active Learning-Based Solution," Management Science, Volume 52, Number 5, May, 2006.
24. Schweighofer, E., Geist, A., "Legal Query Expansion Using Ontologies and Relevance Feedback," TREC Conference 2008, Proceedings.
25. Hyman, H. S., System Acquisition, Integration and Implementation for Engineers and IT Professionals, Sentia Publishing, (2014).