Hippocrates: A Context-aware, Collaboration Enabling Search Tool

Georgios Aravanis, Anca Bucur
Precision & Decentralized Diagnostics
Philips Research
Eindhoven, the Netherlands
{georgios.aravanis, anca.bucur}@philips.com

Mykola Pechenizkiy
Department of Computer Science
TU Eindhoven
Eindhoven, the Netherlands
m.pechenizkiy@tue.nl

Abstract— As the volume of medical literature is growing rapidly, search engines implementing the basic information retrieval model often fail to address the information needs of the medical professionals. In this paper we propose a search tool that is oriented in capturing the information need of the oncologist by including concepts from the context in the searching process. Three different approaches were implemented and evaluated; the pre-filtering, the query expansion and the re-ranking. The system also supports Collaborative Information Seeking (CIS) through algorithmic and User Interface (UI) interventions. The returned citations on a given query are re-ranked based on the ratings that team members have assigned for similar cases making use of the scoring proposed by the I-Spy system. The results from the evaluation of the proposed tool show that pre-filtering performed better than the alternative contextualization search approaches in terms of precision.

Keywords- Context aware systems, Collaborative information seeking, Search tool, Information need

I. INTRODUCTION

In the medical domain there is an exponentially increasing number of scientific publications each year [10]. This trend makes it impossible for the health professionals to be updated to all the research developments. On the other hand, systems implementing the traditional information retrieval model, often do not succeed in meeting the information needs raised to a specialist when dealing with a case.

One of the recent trends in search engines is the use of contextual information. A broadly used method is the inclusion of words from previous queries in order to return results with higher precision. Another trend is the utilization of the location of the user in numerous systems such as recommending systems and search engines especially in smartphone applications e.g. Google now.

An important aspect of the modern knowledge worker's workflow is the collaboration with colleagues. E-mail is one of the most widely used technologies for collaboration [11], [12]. Although there are research studies showing the benefits of a collaborative environment for information seeking, in the medical domain, the main sources of citations are oriented towards more traditional approaches.

Hippocrates is a tool designed based on the premises that contextual signals and collaboration could significantly improve the searching process of the medical expert. On that direction and inspired by the contextualization techniques of the Context Aware Recommender Systems (CARS) we implemented three different algorithms; the pre-filtering, the query expansion and the re-ranking. From the experiment we conducted, it can be concluded that the pre-filtering has better performance in both precision and cumulative gain. A CIS algorithm, the *Collaborative Filtering Re-ranking* was also implemented and evaluated, with positive results.

II. RELATED WORK

Researchers have systematically studied the way that employees collaborate through the information seeking process [11], [12]. These studies took place in a different context than the medical domain and the oncologists. The email was the most common communication tool for communicating interesting findings by 46.8% followed by SMS 30.3% and talking over the phone in 27.5% of the cases. Similar results are seen in a study in Collaborative Information Seeking [3]. The use of email for collaboration by oncologists was confirmed during the first evaluation of Hippocrates.

In [16], multiple systems are presented that promotes collaboration in a synchronous way such as in [6] where the members of a team have specific roles, namely Prospector, Miner, domain expert etc. The prospector performs the queries, and the miner rates the results and based on the algorithmic part of the system, recommendations are presented in a common UI. Such systems can be classified based on the location of the team members, Collocated/Remote located, the data synchronization, Synchronously-Asynchronously [14] and the type of system intervention User Interface-Algorithmically.

Other systems that support collaboration through algorithmic integration are the I-Spy and the Haystack. The core assumption on which the I-Spy algorithm was based on is that the documents that are accessed more often are those that are more relevant to a group of users issuing the same or a similar query [2], [18]. The Haystack system [19] uses defined tasks as staks where people can join and rate pages as relevant or not relevant and tag results with keywords. The backend system of Haystack includes a recommender system that promotes relevant results on top of the returned results from a search engine.

One of the first systems that used contextual information is the Medline button [4]. The process starts with the selection of the terms that are relevant to the information need. The system proposes a set of potential relevant questions which are based on the analysis of user questions posed to librarians. The doctor selects the question that best describes the information need. The system matches the relevant terms to the MeSH terms. Then the system provides the references from the Medline. This pattern of manual selection of relevant terms, matching to MeSH terms and retrieving the results is seen in other research efforts with some modifications [8], [13].

Other context aware searching approaches include into the scoring function contextual information from previous queries or click through data from the logs of the user [5], [7], [20] that leads to search personalization. Another approach towards that direction is the use of ontological user profiles [17] and expert-driven adaptive hypermedia [21], [23].

In CARS different contextual signals are used for improving the recommendations [1].

III. Implementation

A. *System's input*

The Hippocrates search tool presented in **Figure 1**, is making use of five sources of information: (i) the Electronic Health Record (EHR) of the patient, (ii) the MEDLINE/PubMed citations that were indexed in the search engine, (iii) the SNOMED Clinical Terms (through the Bioportal interfaces) (iv) the user logs and the ratings of users on citations and (v) a vocabulary linked to a workflow state.

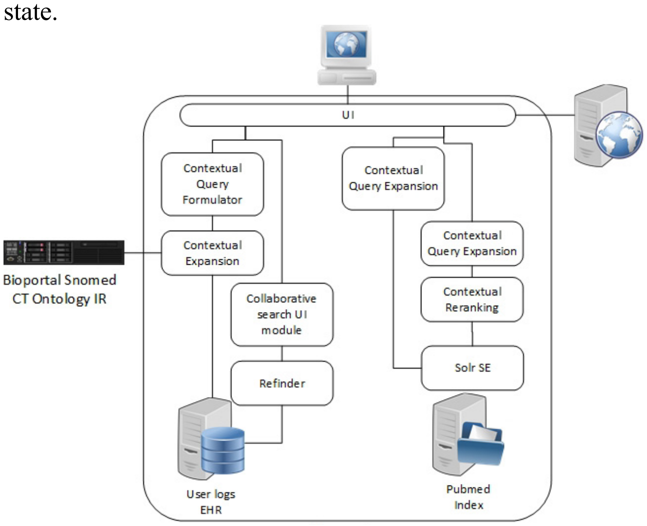

Figure 1. Hippocrates architecture overview

The proposed system requires from the user to enter the patient ID of the case in order to load the information from the patient's EHR and the workflow state of the treatment process. For the evaluation of the proposed contextualized search approaches the user selects one of the four searching methods: (i) the basic search, (ii) the query expansion, (iii) the re-ranking and (iv) the pre-filtering. Furthermore, the user can optionally activate and in combination with any of the previous searching methods, the collaborative re-ranking.

In case the user selects one of the contextualized search options, the *Contextual Expansion* module retrieves patient data from the EHR. The structure of the EHR follows a reduced data model of the HL7 Reference Information Model (RIM). This implementation choice makes the system integration-ready since HL7 is a widely used standard for healthcare interoperability. For evaluation purposes, fictitious cases were inserted in the database with SNOMED CT encoded concepts. The alternative of using real patients EHRs would require a legal and security framework to handle the complicated privacy issues.

B. *Contextual Expansion*

The *Contextual Expansion* component decodes SNOMED CT concepts to terms e.g. 248152002 is translated to the term "female", classifies the patient from the age of birth as adult or child, identifies the diagnosis concept among all the Observations of the Act entity of the HL7 RIM based EHR and retrieves diagnosis' alternative labels. Furthermore, the parent and siblings concepts of the identified *Antineoplastic agent* administered to the patient are retrieved. The component is making use of the Bioportal API for searching SNOMED CT. The corresponded set of terms to the workflow state of the treatment process are retrieved by the module as well.

The retrieved concepts are organized in categories and become available to be used by the contextualized search algorithms. In case the pre-filtering approach is selected those terms are presented to the user in a form of a menu. The UI component serving this functionality is the *Contextual Query Formulator*.

C. *Solr search engine configuration*

The Solr search engine [1] provides the basis for the proposed tool, the index of which consists of approximately 1.6 million citations of the Medline/Pubmed library. The documents were filtered on the MeSH term Neoplasms since the system addresses to oncologists. The fields that can be searched are the *ArticleTitle* and the *AbstractText*. A preprocessing of those fields was applied during indexing i.e. stemming, lowercasing and stop words removal. The same processing is applied to the user queries. A minimum threshold of matching terms between the query and the index, after testing different levels, was set to 80%. The scoring is based on the state of the art in the probabilistic retrieval framework function BM25 (BM25SimilarityFactory class of Solr with default parameters).

$$Score = \sum_{t \in q} max\{BM25(t, k_{Title}), BM25(t, k_{AbstractText})\}$$

where t are words and phrases of a query q. The coefficients k_{Title} and $k_{AbstractText}$ are boost values used in Solr that were set after testing to 1.1 and 1 respectively. Phrases are considered keywords with less than seven terms distance in between.

[1] http://lucene.apache.org/solr/

D. Context aware algorithms and modules

1) Query Expansion

State of the art search tools utilize the user's search history in order to expand the entered queries with additional terms. In the proposed system, the *Contextual Query Expansion* uses concepts extracted from the EHR of the patient. Initially all the concepts were included in the query with lowering the minimum matching criteria from 80% of the terms to 50%. The experimentation with this approach though, suggested that many irrelevant results were retrieved due to very frequent words such as *woman* or *adult*.

Based on a behavior analysis of the users of Pubmed the diagnosis is used by the 20% of the queries [10]. Thus the current setup of this component expands the user query with the *Observations* of the EHR and particularly sets higher coefficient to the diagnosis by using the boost queries functionality of Solr (1.1 for the observation-related terms and 1.2 for the diagnosis).

2) Pre-Filtering

A similar concept implemented in the CARS, the pre-filtering, is one of the contextualized search approaches proposed in this paper. The user is filtering the results by selecting and unselecting terms provided by the *Contextual Query Formulator* menu. From the implementation aspect, those terms must appear in the title or the abstract of the retrieved citations. The aforementioned scoring function is applied for the terms of the user's query, including the concepts selected from the menu.

3) Contextual Re-ranking

The re-ranking component is activated, in case the corresponded option is selected by the user. The first 500 documents with the highest score are assigned to an array. All the concepts provided by the *Contextual Expansion* component are assigned in a different array. Then a *Boolean score* is calculated for each document, which is the number of matched terms from the context that appears in the document. The final score is calculated by:

$$Score = Score + k \cdot BooleanScore$$

Where *Score* is the resulted score from the BM25 based function of the search engine and *k* is a coefficient for the *BooleanScore*. For the evaluation we used *k*=10 since during the internal testing returned the most reasonable results.

E. Collaborative search algorithm

The *Collaborative Filtering Re-Ranking* algorithm is implemented to support the CIS among the members of an oncology team and is based on that of I-Spy. The algorithm is designed to boost documents highly rated for similar queries and cases and to penalize documents with lower ratings. The algorithm presented in the I-Spy system is modified in order to make use of the rating component (1-5 ratings), the tag component (tags assigned to citations) and the contextual information that is available through the EHR and the Hippocrates system. The proposed extension is based on the user ratings instead of assuming the pages that are clicked the most are of higher relevance (as proposed by I-Spy). Additionally, a contextual pre-filtering is applied. The set of documents retrieved, are rated in a context of the same diagnosis and workflow state. Subtracting the standard deviation of the ratings of a movie can improve the accuracy in predicting an unknown rating [5]. A similar approach is implemented in Hippocrates.

The I-Spy equations are expanded in the following way:

$$Sim(q, q') = \frac{|q \cap q'|}{|q \cup q'|} \tag{1}$$

$$Relevance(p_j, q_i) = \frac{H_{ij}}{\sum_{\forall i} H_{ij}} \tag{2}$$

$$RatingScore(p_j, q_i) = \frac{AvgScore_{ij}}{1 + STD_{ij}} \tag{3}$$

$$W\,Rel(p_j, q_i, q_1, \dots, q_n) = \\ = \frac{\sum_{i=1\dots n} Relevance(p_j, q_i) \cdot Sim(q_T, q_i) \cdot RatingScore(p_j, q_i)}{\sum_{i=1\dots n} Exist(p_j, q_i) \cdot Sim(q_T, q_i)} \tag{4}$$

where q and q' in (1) are the sets of terms of the user's query and the set of terms used for tagging a reference respectively. In (2), the relevance of a reference p_j on a set of tags q_i, is given by the number a tag set was assigned on a reference-H_{ij}, over the number the same tag set was assigned on all references. In (3), the rating score of a reference p_j on a tag set q_i is given by the average score that users assigned $AvgScore_{ij}$, over the standard deviation of those scores STD_{ij}.

The algorithm is described in the following steps:

1. Select the articles, the ratings and the tags where the diagnosis is the same as the diagnosis of the current patient and the workflow state is the same as the workflow state of the current case.
2. Calculate the similarity of the tags used for the rating of a document with the user's query (1).
3. Filter out the results with similarity less than a threshold of 60% .
4. Group the documents by tag set.
5. Count the frequency that each document is rated on a given tag set (2).
6. Calculate the Rating Score of a document for each tag set (3).
7. Calculate the score (4) of a document by aggregating the scores of the document for the different tag sets.

The scores calculated from the algorithm, are stored in an external file. Then the score of each document is calculated as following:

$$Score = Score + k \cdot MaxScore \cdot WRel$$

where *MaxScore* is the score of the first result and *k* is a coefficient in Hippocrates, which after the internal testing is set to 0.2. The citations are re-ranked and presented to the user (instead of promoting the top results approach of I-Spy).

F. Other features and UI functionality

The proposed tool besides the contextualized search and collaborative filtering re-ranking algorithms, includes a set of features that supports collaboration through UI intervention and facilitates the user in the searching process. Some of those features are presented in Figure 2. The supporting implemented features are:

- Auto-complete of the users' queries based on the index.
- Sorting by time of publication.
- Recommendation of similar references (tf-idf based).
- Visual notification of new results after query alteration.
- Text highlighting on matched terms between a query and a citation.
- Faceted search on the fields provided by Pubmed: Year of publication, Publication Type, Journal title .
- *Re-finding* of previously highly rated citations and previously entered queries sorted in the chronological order.
- Link to the full paper when available (through Pubmed.

Collaboration support through UI features are:

- E-mail of a specific citation including case and query information.
- Recommendation of the most common queries entered for a specific case on a certain workflow state.
- List of the highly recommended citations from the team sorted by workflow state and then by time.
- List of the highly rated citations from the team sorted on the score of the *Collaborative Filtering Re-Ranking* module.

Figure 2. Hippocrates search tab UI

IV. EXPERIMENT AND TOOL EVALUATION

In order to determine which of the implemented contextualized search approaches performs the best and to assess the performance of the collaborative filtering re-ranking algorithm, we designed and performed an experiment in parallel with a usability evaluation session.

A. *Experimental setup*

The participants were provided with a fictitious case on a form with all the contextual information; patient characteristics, diagnosis, observations, and procedures. The corresponded case was inserted in the EHR of Hippocrates. The use of workflow state related concepts and context was left out of scope.

1) *Contextualized search evaluation*

The participants were asked to judge the relevance of the first 18 results in 2 pages in order to judge the precision@18 (P@p) and the cumulative gain (CG) with the implemented 5 stars rating system. The rating was performed with the "basic search" and with all three contextualized search methods for a specific information need and with a single query; "complete resection side effects". This information need was formulated in collaboration with an oncologist.

P@*p* corresponds to the number of relevant articles (above 3 stars rating) in the first *p* results. The CG of the first *p* results is defined as the sum of the graded relevance values of the results:

$$CG_p = \sum_{i=1}^{P} rel_i$$

where rel_i is the graded relevance of a result at position *i*.

For the third option, that of the query formulation support, the users were asked to use the filtering options derived from the EHR (using the *Contextual Query Formulator* menu) until they reach the most relevant result set. Then they were asked to judge the relevance of the top 18 results once more.

2) *Collaborative search evaluation*

For the collaborative search evaluation we studied the effect of the activation of the module on the cumulative gain compared to the "basic search".

3) *Usability evaluation*

The usability evaluation included three parts; the first contained a general usability satisfaction questionnaire (adapted from [24]), the second referred to the separate usability evaluation of the components of Hippocrates and the third part requested the user's input on the most positive and negative aspect of the tool, new features requests and any additional comments.

V. RESULTS

The participants of the study were of three specialties; an oncologist, an informatician and a data analyst. The experiment setup was adjusted content-wise to the participants domain expertise. The query used for the evaluation of the informatician and data analyst was changed to the more generic "breast cancer". Due to time limitations the oncologist completed half of the experiment on the contextualized search part (first page/nine results), and since the query used was different the results are presented in different figures.

A. *Contextualized Search*

Figure 3 and **Figure 4** show that the contextualized search performed better than the default setting of the search engine with re-ranking and pre-filtering retrieving eight out of nine relevant results. The query expansion retrieved one relevant result less but in total the returned citations were more relevant (Cumulative Gain).

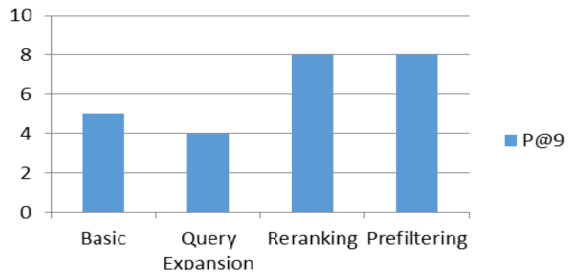

Figure 3. Precision Oncologist

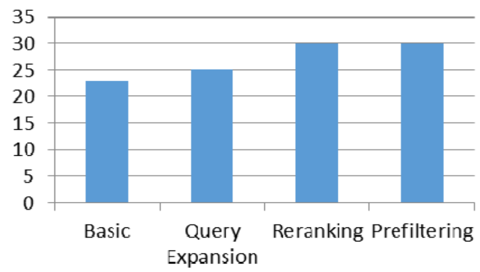

Figure 4. Cumulative Gain Oncologist

The bar charts in **Figure 5** and **Figure 6** show that the informaticians got more relevant in numbers and quality with pre-filtering than with any other method. The results on re-ranking contradict with the findings from the oncologist. That could be due to the different background and query used in each case. Query expansion performed better in terms of precision and cumulative gain to one of the participants.

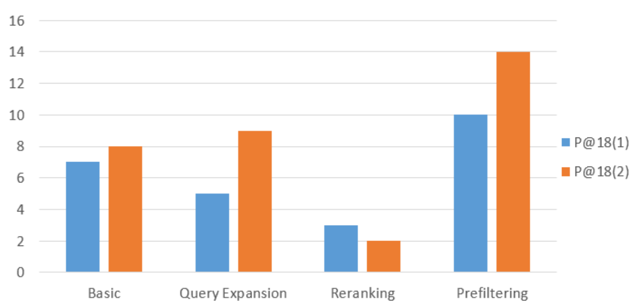

Figure 5. Precision Informaticians

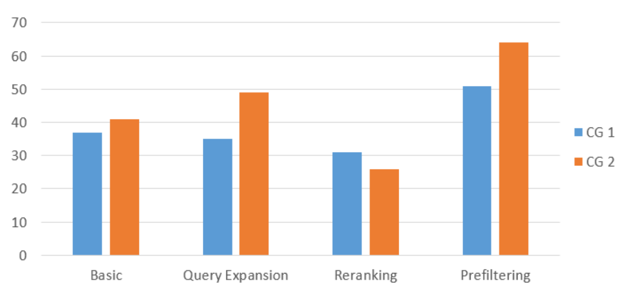

Figure 6. Cumulative Gain Informaticians

B. *Collaborative Filtering Re-ranking*

As seen on **Table 1**, the cumulative gain and the precision of the first results can be both improved by the activation of the collaborative filtering re-ranking algorithm. As an example, the document with PMID 3167232 was ranked higher since it received consistently high ratings for the "breast cancer" query (6x4 stars and 1x5 star). For the rest documents the deviation of the ratings was high enough and/or the scoring differences with the next in rank document was that broad that did not affect the ranking significantly.

Another finding is the necessity to replace the sum of scores divided by the sum of similarities part of the equation (4) of the I-Spy by a more effective one. As seen in **Table 1**, the document with PMID 9250077 was re-ranked in a much lower position. This was resulted due to a single negative rating (1 star) with high relative frequency and a slightly altered tag set that minimized the influence of 13 ratings with average score of 2.33 stars.

Table 1. New ranking after the collaborative ranking is enabled

New Ranking	Old Ranking	PMID	Old Score	I-spy Score	New Score
1	4	3167232	26.76	0.29	27.05
2	2	14574169	26.82	0.02	26.84
3	3	16630739	26.80	-0.06	26.74
4	5	15711624	26.71	-0.29	26.42
5	6	22990110	26.39	0.02	26.41
6	7	21844186	26.38	-0.05	26.33
11	10	7954390	26.29	-0.08	26.21
14	9	15842935	26.34	-0.18	26.16
22	8	19996029	26.38	-0.48	25.90
N/A	1	9250077	27.51	-4.51	23.00

C. *Tool and Modules evaluation*

Figure 7 depicts the positive evaluation in terms of usability of the proposed system by the participants. Notably, the Oncologist rated higher several aspects of the system.

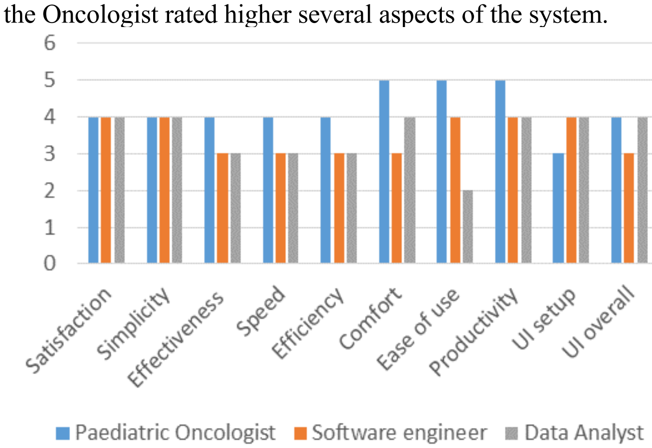

Figure 7. Overall usability evaluation of Hippocrates

Besides the experiment and the usability evaluation, the participants were requested to describe the positive aspects of the system and any areas of improvements. Among the positives aspects, the oncologist emphasized the ease of use.

The software engineer gave feedback on two contextualization approaches namely that of the query expansion for which he positively described the fact that he immediately noticed the results corresponded to a more refined information need, by the inclusion of the diagnosis and that of the pre-filtering option. Fast response was another positive aspect of the tool. The re-finding component was received as a useful feature. Re-finding was mentioned by the data analyst as a positive aspect of Hippocrates who also mentioned the importance of the different types of recommendations and the usefulness of the different searching approaches.

Feedback on aspects of the system that could be improved includes the use of the MeSH vocabulary as an option to filter on the results, the exclusion of citation discussing different types of cancer and the lack of a searching feature for the tagged keywords. Full control to the user on which Journals appear on the faceted menu was also requested as an extra functionality by the oncologist.

VI. Conclusions and Future Work

In this paper we presented Hippocrates, a search tool designed to support both contextualized search and collaborative information seeking. Inspired by and in an analogy to the contextualized approaches of the recommendation systems, we implemented three different algorithms, namely the pre-filtering, the query expansion and the re-ranking that are utilizing the concepts included in the EHR of the patient and the user needs on a specific workflow state of the treatment process. On the collaboration part the proposed system re-uses and extends the scoring functions of the I-Spy CIS system and also provides additional features based on UI interventions. Supporting re-finding of cases is another important aspect of the system.

The experiment has shown that among the contextualization techniques, that of the pre-filtering performed the best among the algorithms in terms of precision and cumulative gain. The study is inconclusive for the re-ranking and the query expansion since only in some cases those approaches performed better than the default search functionality. In addition we presented how the *collaborative filtering re-ranking* module is potentially improving precision and the cumulative gain by including the ratings of citations of other users in the scoring function.

Many directions could be followed for future work and extension of Hippocrates. Further evaluation and validation of the currently implemented algorithms and features is a necessity in combination with fine-tuning of the scoring mechanisms. Inclusion of additional contextual signals such as the workflow state that was excluded from the experiment, is one of the next steps. Better modeling of the current expert's information need [22] would be a possibility too. Another aspect needed to be tested is the efficiency of the combined use of algorithms in the searching process. Furthermore, real patients' EHRs should be used to a next evaluation of the tool. In that case, a complex legal and security framework should be developed that will assure the privacy of the patients and the integrity of their data.

VII. References

[1] Adomavicius, G., & Tuzhilin, A. (2011). Context-Aware Recommender Systems. *Recommender Systems Handbook*, 217-253.

[2] Balfe, E., & Smyth, B. (2007). Case-Based Collaborative Web Search. *Case-Based Reasoning Research and Development* , Volume 4626 pp 329-343.

[3] Böhm, T., Klas, C.-P., & Hemmje, M. (2014). Collaborative Information Seeking in Professional Work-Settings: A Study of Equipment Utilization. *Springer*.

[4] Cimino, J. J., Johnson, S. B., Aguirre, A., Roderer, N., & Clayton, P. (1993). The Medline Button. *AMIA*, 0195-4210/92.

[5] Fonseca, B. M., Golgher, P., Possas, B., Ribiera-Neto, B., & Ziviani, N. (2005). ConceptBased. *ACM*.

[6] Golovchinsky, G., Adcock, J., & Pickens, J. (2008). Cerchiamo: a collaborative exploratory search tool. *In Proc* of CSCW'08.

[7] Jiang, D., Kenneth , L., & Wilfred, N. (2011). Context-Aware Search Personalization with Concept. *ACM*, 978-1-4503-0717-8/11/10.

[8] Jimenez-Castellanos, A., Fernandez, I., Perez-Rey, D., & Viejo, E. (2013). Patient-Based Literature Retrieval and Integration: a use case for Diabetes and Arterial Hypertation. *Communications in Computer and Information Science*, 312-323.

[9] Kwon, Y. (2008). Improving top-n recommendation techniques using rating variance. *In Proc. ACM Conf. Recommender systems*. 307-310.

[10] Lu, Z. (2011). PubMed and beyond: a survey of web tools for searching biomedical literature. *Database*, p.36.

[11] Morri, M. R. (2013). Collaborative Search Revisited. *ACM*, 978-1-4503-1331-5/13/02.

[12] Morris, M. R. (2008). A Survey of Collaborative Web Search Practices. *Proceedings of CHI*, 1657-1660.

[13] Perez-Rey, D., Jimenez-Castellanos, A., Garcia-Remesa, M., Crespo, J., & Maojo, V. (2012). CDAPubMed: a browser extension to retrieve. *BMC Medical Informatics and Decision Making*.

[14] Pickens, J., Golovchinsky, G., Shah, C., Qvarfordt, P., & Back , M. (2008). Algorithmic mediation for collaborative exploratory search. *Proceedings of the Annual ACM*.

[15] Rezarta, I. D., Murray , C. G., Aurelie, N., & Zhiyong, L. (2009). Understanding PubMed user behavor through log analysis. *The Journal of Biological Databases & Curation*.

[16] Shah, C. (2012). *Collaborative Information Seeking*. Springer.

[17] Sieg, A., Mobasher, B., & Burke, R. (2007). Web Search Personalization with Ontological User Profiles. *ACM*, 978-1-59593-803-9/07/0011.

[18] Smyth, B., Balfe, E., Feyne, J., Briggs, P., Coyle, M., & Boydell, O. (2004). Exploiting Query Repetition and Regularity in an Adaptive Community-Based Web Search Engine. *UMUAI* 14: 383–423.

[19] Smyth, B., Coyle, M., & Briggs, P. (2011). On Communities, Collaboration, and Recommender Systems in Personalized Web Search. *Recommender Systems Handbook*, 579-614.

[20] Xiang, B., Jiang, D., Pei, J., Sun, X., Chen, E., & Li, H. (2010). Context-Aware Ranking in Web Search. *ACM*, 978-1-60558-896-4/10/07.

[21] Hendrix M., De Bra P., Pechenizkiy M., Smits D. & Cristea A. (2008) Defining Adaptation in a Generic Multi Layer Model: CAM: The GRAPPLE Conceptual Adaptation Model. In Proc. 3rd European Conf. on Technology Enhanced Learning (EC-TEL'08), 132-143.

[22] Kiseleva J., Lam H.T., Pechenizkiy M., Calders T. (2013). Predicting Current User Intent with Contextual Markov Models. ICDM Workshops 2013: 391-398.

[23] Knutov E., De Bra P. & Pechenizkiy M. (2011). Generic Adaptation Framework: a Process-Oriented Perspective, Journal of Digital Information 12(1).

[24] Lewis J.R. (1995) IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use. International Journal of Human-Computer Interaction 7(1): 57-78