SEMANTICALLY ENHANCED MEDICAL INFORMATION RETRIVAL SYSTEM:A TENSOR FACTORIZATION BASED APPROACH.

N.SRINIVASA RAO [1], **KUMPATLA SAI CHARAN**[2].

[1] Assistant Professor, **DEPT OF MCA, SKBR PG COLLEGE , AMALAPURAM, Andhra Pradesh**

Email:- naagaasrinu@gmail.com

[2]**PG Student of MCA, SKBR PG COLLEGE , AMALAPURAM, Andhra Pradesh**

Email:- kumpatlasaicharan@gmail.com.

Abstract: The medical information relevance model plays a vital role to retrieve accurate results regarding the medical information and also enhancing the decision making power. Integrating the medical knowledge bases has the potential to improve the information retrieval performance through incorporating medical domain knowledge for relevance assessment. In this project, we propose a medical information retrieval system with a two-stage query expansion strategy, which is able to effectively model and incorporate the latent semantic associations to improve the performance. Hence, we use incremental PRF and latent semantic relevance model to retrieve accurate results.

IndexTerms **-IRS, Incremental PRF**

I.INTRODUCTION

With the increase in the medical related information, the retrieval of high quality results is becoming more critical.
The following are the challenges which are overcome in this project such as:

- The inherent complexity of medical languages such as obscure medical terminologies and ambiguous abbreviations.
- The associated variety of information needs from different types of users, such as patients and physician.
- The complexity and ambiguity of medical languages result in vocabulary mismatch between queries and documents.

We present a latent semantic relevance model based on tensor factorization to identify semantic association patterns under sparse settings. Experiments showed that the performance of the proposed system is significantly better than the baseline system, and is comparable with state-of-the-art systems. First, we applied a heuristic approach to enhance the widely used pseudo relevance feedback method for more effective query expansion, through iteratively expanding the queries to boost the similarity score between queries and documents. Second, to improve the retrieval performance with structured knowledge bases, we presented a latent semantic relevance model based on tensor factorization to identify semantic association patterns under
sparse settings. For example, a user submits a query to search for the information of fever treatment. There is a document introducing "paracetamol", a medication used to treat fever. Human experts may judge that this document is relevant to the user's query. However, this relevance could not be automatically identified without the knowledge bases which contain the association in the form of a triple ("paracetamol", "may treat", "fever"). In this, we also present the Knowledge-Based Query Expansion, where the accurate results are fetched. This system works by comparing the documents with the accurate result documents. As some diseases may have similar symptoms, this approach shows all the results where the symptoms are common according to their ranks. This latent semantic document contains the key technical terms which is used as keywords while comparing various documents.

II. LITERATURE SURVEY

The increasing amount of information that is annotated against standardized semantic resources, offers opportunities to incorporate sophisticated levels of reasoning or inference into the retrieval process. In this position paper, we reflect on the need to incorporate semantic inference into retrieval (in particular for medical information retrieval) as well as previous attempts that have been made so far with mixed success. Medical information retrieval is a fertile ground for testing inference mechanisms to augment retrieval. The medical domain offers a plethora of carefully curetted, structured, semantic resources, along with well established entity extraction and linking tools, and search topics that intuitively require a number of different inferential processes (e.g., conceptual similarity, conceptual implication, etc.). We argue that integrating semantic inference in information retrieval has the potential to uncover a large amount of information that otherwise would be inaccessible; but inference is also risky and, if not used cautiously, can harm retrieval. Knowledge-based query expansion to support scenario-specific retrieval of medical free text

In retrieving medical free text, users are often interested in answers pertinent to certain scenarios that correspond to common tasks performed in medical practice, e.g., treatment or diagnosis of a disease. A major challenge in handling such queries is that scenario terms in the query (e.g. treatment) are often too general to match specialized terms in relevant documents (e.g. chemotherapy). In this paper, we propose a knowledge-based query expansion method that exploits the UMLS knowledge source to append the original query with additional terms that are specifically relevant to the query's scenario(s). We compared the proposed method with traditional statistical expansion that expands terms which are statistically correlated but not necessarily scenario specific. Our study on two standard test beds shows that the knowledge-based method, by providing scenario-specific expansion, yields notable improvements over the statistical method in terms of average precision-recall. On the OHSUMED test bed, for example, the improvement is more than 5% averaging over all scenario-specific queries studied and about 10% for queries that mention certain scenarios, such as treatment of a disease and differential diagnosis of a symptom/disease. Using knowledge-based relatedness for information retrieval

Traditional information retrieval (IR) systems use keywords to index and retrieve documents. The limitations of keywords were recognized since the early days, specially when different but closely related words are used in the query and the relevant document. Query expansion techniques like pseudo-relevance feedback (PRF) and document clustering techniques rely on the target document set in order to bridge the gap between those words. This paper explores the use of knowledge-based semantic relatedness techniques to overcome the vocabulary mismatch between the query and documents, both on IR and Passage Retrieval for question answering. We performed query expansion and document expansion using WordNet, with positive effects over a language modeling baseline on three datasets, and over PRF on two of those datasets. Our analysis shows that our models and PRF are complementary; in that, PRF is better for easy queries, and our models are stronger for difficult queries and that our models generalize better to other collections, being more robust to parameter adjustments. In addition, we show that our method has a positive impact in an end-to-end question answering system for Basque and that it can be readily applied to other knowledge bases, as our good results using Wikipedia show, paving the way for the use of other knowledge structures such as medical ontologies and linked data repositories.

III. OVERVIEW

3.1 Existing system:

Sfakianaki et al. proposed a natural language processing framework to automatically transform a clinical research question to a query that contains only terms of biomedical ontology. Their research demonstrated the capability of biomedical ontology and entity annotation algorithms to bridge the gap between clinical questions in natural language and biomedical literature. Mao et al. proposed a new medical IR system enhanced by manually assigned subject terms (Medical Subject Headings, MeSH). The proposed system constructs generative concept models to capture the associations between queries and documents. Otegi et al. performed both query expansion and document expansion using a lexical database on IR tasks for question answering, and showed that their methods are complementary with pseudo-relevance feedback. The traditional way was based on ranks of the document. This was inefficient as there were no accurate or precise results fetched. The next approach used only average of the ranks of the documents and depending on average the related documents were fetched. The proposed system uses incremental PRF algorithm with latent semantic relevance model by which accurate results are fetched.

3.2 Disadvantage:

- However, the performance of knowledge-based approaches was not satisfactory
- The organizers pointed out that the poor performance of existing approaches could be resolved with available training data to tune the parameters

3.3 Proposed System

To develop the knowledge-based medical IR system, we incorporate the domain-specific information extracted from UMLS, a widely used knowledge base in medical domain. In the UMLS, synonymous terms are clustered into concept, and concepts are linked to other concepts in the semantic network. We developed a semantically enhanced medical IR system, which has a two-stage query expansion strategy to integrate the pseudo relevance feedback and the knowledge-based query expansion to improve the performance of retrieving relevant documents for queries. First, we proposed the incremental pseudo relevance feedback (incremental PRF) approach for query expansion to obtain the initial ranking list of retrieved documents. Second, we developed an enhanced knowledge-based query expansion method with a novel latent semantic relevance model. The proposed method will re-rank the documents retrieved by the incremental PRF in the first stage.

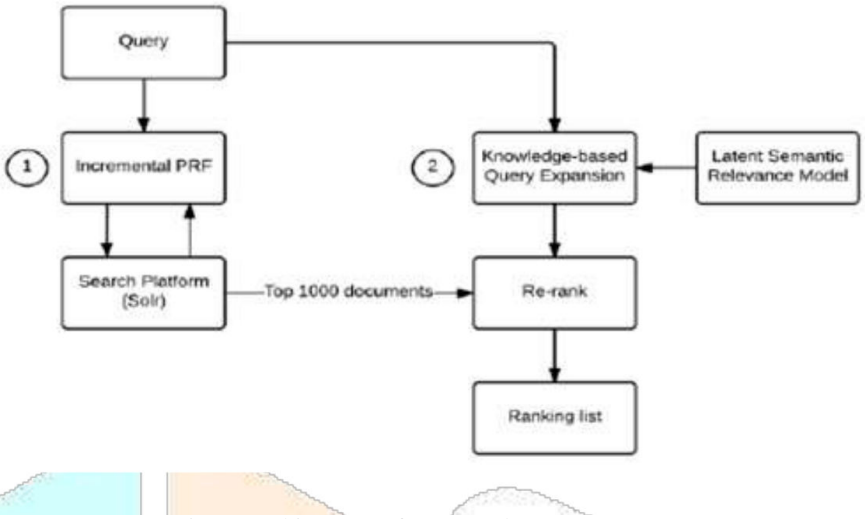

Fig. 1 Architecture of Proposed System

3.4 Module

- **Admin**
- **User**

3.4.1 Admin

In this module, the Admin has to login by using valid user name and password. After login successful he can do some operations such as view all user and their details and authorize them. Admin can add medical records from dataset and update to database. Admin can view total no of users activated.

3.4.1 User

In this module, there are n numbers of users are present. User should register before doing some operations. After successful registration he has to wait for admin to authorize him and after admin authorized him, he can login by using authorized user name and password. After successful Login he can do some operations like search for medical records based on keywords and view search results. User can perform the following operation:

- Incremental Pseudo Relevance Feedback (Incremental PRF)
- Latent Semantic Relevance Model
- Tensor based Semantic Association Representation

IV. METHODOLOGY

The methodology used in this system is implementing incremental PRF algorithm. In this system ranks are calculated using incremental PRF and knowledge based query expansion. In this, keywords are selected from queries which are compared with semantic documents. For e.g. if the query is “fever”, incremental PRF is applied and from the resultant documents the keywords comprising of disease name, symptoms and medicine is compared between semantic documents that contains accurate documents. Hence using this system, we will get accurate and more precise results of documents. It is highly efficient.

The Pseudo Relevance Feedback algorithm uses the documents which are retrieved at the top of the retrieved documents list, as these are more related to the query. From these retrieved documents terms can be taken for the query expansion. Generally a fixed setting is used across different queries, which is not optimal because of variety of queries and feedbacks documents. Proper expansion can fetch more relevant documents. So, optimal approach is used to extract proper and relevant expansion terms from the top-ranked retrieved documents for each query. Hence, in this algorithm, the average rank of the ranked documents retrieved is calculated, followed by retrieving the documents having ranks above the average rank of the documents. From these documents, the terms or the keywords are passed to Expansion and the keywords are compared from the latent Semantic documents. The results are re-ranked which gives an efficient list of documents.

V. RESULTS AND DISCUSSION

The evaluation of the proposed medical information retrieval system follows the standard TREC evaluation method for ad hoc retrieval tasks. Documents with high relevance were selected for experts to decide the relevance as “not relevant”, “possible relevant” and “definitely relevant”. These samples were used to evaluate the performance of the systems. The evaluation is done for the performance of three systems, the baseline of system, the PRF system that incorporated incremental PRF and the proposed system that has both incremental PRF and Knowledge based query expansion system. The results demonstrated that the proposed system is more efficient and accurate than the existing baseline system. The results demonstrated that the proposed system improves the performance of the system and outperforms the existing system. As an active research area a variety of areas are explored recently to optimize the performance of medical information retrieval system.

VI. SCREENSHOTS

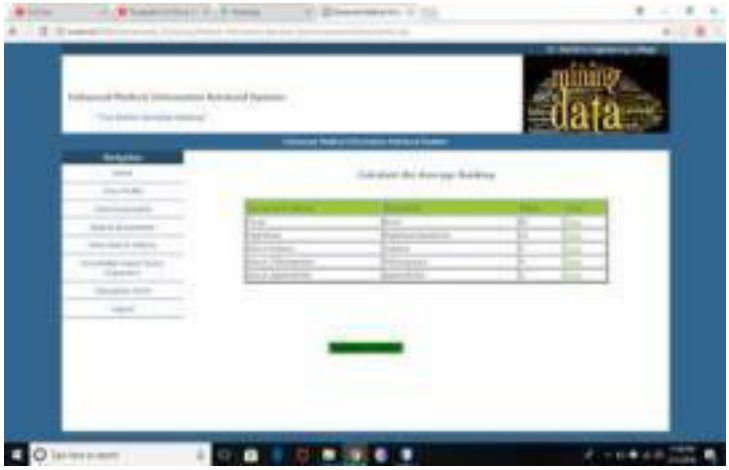

Fig. 3 Screen shot of the proposed system

Fig. 4 Screen shot of the proposed system

VII. CONCLUSION AND FUTURE SCOPE

In this Project, we proposed a medical IR system with a two-stage query expansion strategy, based on the incorporation of semantics from knowledge bases with tensor factorization methods. Experiments with the TREC dataset demonstrated the effectiveness of the proposed system. The proposed system has the potential to be adapted in other machine learning and medical informatics applications, like recommender systems, ontology learning, bioinformatics, etc. In future this project can be implemented in real-world medical decision support applications through the collaboration with doctors and decision makers in local hospitals

REFERENCES

1. M.Nickel,"Scikit-tensor library, available online. URLpi.python.org/pypi/scikit-tensor.," 2013.

2. Q. Zhang and D. Haglin, "Semantic similarity between ontologies at different scales," IEEE/CAA Journal of AutomaticaSinica, vol. 3, pp. 132-140, 2016.

3. M. Nakatsuji, H. Toda, H. Sawada, J. G. Zheng, and J. A. Hendler, "Semantic sensitive tensor factorization," Artificial Intelligence, vol. 230, pp. 224-245, 2016.

4. K.-W. Chang, W.-t. Yih, B. Yang, and C. Meek, "Typed Tensor Decomposition of Knowledge Bases for Relation Extraction," in EMNLP, 2014, pp. 1568-1579.