Relevant Evidence Acquisition and Appraisal using Knowledge-intensive Queries

Muhammad Afzal, Sungyoung Lee

Department of Computer Engineering, Kyung Hee University,Seocheon-dong, Republic of Korea, 446-701
{muhammad.afzal, sylee}@oslab.khu.ac.kr

Abstract- **Information needs of the users have grown exponentially with the advent of advancements in information and communication technology. The traditional ways of searching information from the online resources has been evolved and the tendency is geared more towards getting quality contents. In healthcare domain, the clinical researchers and physicians are even more interested to find quality information to use as a clinical evidence in decision making. An increasing number of potential articles in the form of MEDLINE articles are readily available to be retrieved, helps in evidence-based clinical decisions, however, the retrieval methods pose several challenges to clinicians. The first challenge is to automatically reformulate the user query into a knowledge-intensive query in order to acquire articles that are relevant to user needs. The second challenge is to re-evaluate the retrieved articles in order to get quality studies and filter-out all the low quality articles. In this paper, we approach to solve these challenges by proposing two methods to construct knowledge-intensive query for relevant evidence acquisition and statistical model for quality evidence appraisal. The construction of knowledge-intensive query is based on the term expansion using domain model, name variants, and terminological variants. The statistical model is learnt on a corpus prepared through automatic construction of feature vectors from data and metadata features. We evaluate the results at two levels; 1) pre-appraisal stage, 2) post-appraisal stage. We compared the results based on the retrieved result sets with knowledge-intensive query approach and simple query approach. The proposed knowledge-intensive query approach successfully retrieves the potential evidences with average 12.33% improved accuracy in contrast to simple query approach. Furthermore, we performed human evaluation to identify the overall satisfaction of the proposed approach. From the user input, we learned that the proposed approach contributes to maximizing the clinical throughput of clinicians by minimizing the unnecessary intermediary manual steps in evidence retrieval and the appraisal process.**

***Keywords*—Content-based retrieval, Information filtering, Keyword search, Machine learning, Evidence Appraisal.**

I. INTRODUCTION

Evidence-based systems have long been used in clinical domain for clinical practice as a support to the clinicians in order to make better clinical decisions. Individual clinical expertise and best available external evidences complement each other and good doctors use them in combination, and neither alone is enough [1,2]. A number of obstacles need to be handled in accessing to correct information including query formulation, evidence acquisition, and evidence appraisal.

A significant body of literature exist on query generations for efficient information retrievals [3-5], where, queries may be generated manually, automatically, or semi-automatically. In manual case, users regardless of their expertise level, create query of their choice and retrieve the results. In semi-automatic case, users are supported with tools to create query by adding some of the query ingredients automatically [6]. In automatic case, system creates query without involving users. For semi- and automatic cases, the input source plays pivotal role to automatically create ingredients of query [7,8]. Similarly, for query expansion, a number of approaches are employed such as; query expansion with UMLS [3,4,9], WordNet [10], MeSH [11], and combination of them [7]. Similarly, acquiring right evidences and appraising their relevancy to the domain has been investigated using clinical query filters [12] and statistical techniques [13,14]. In broader terms, an integrated approach for evidence acquisition and appraisal is largely missing.

In the area of contextual information retrieval, the most prominent approach is Infobuttons [15-19], firstly defined by Cimino that automatically generate queries to e-resources using contextual information and patient data from EMR. The main focus of Infobuttons approach is to establish context-specific links to health information e-resources. The query topics are pre-specified, clicking on specific Infobutton generates a query from the terms associated with the clicked topic and returns the potential external e-resources.

The main difference of the proposed approach with existing approaches is the domain specific knowledge utilization in query reformulation and appraisal of the retrieved set of articles. Some of the existing systems utilizes documents in training set for query construction and others uses patient records. We utilized the domain specific knowledge to reformulate the user query into concept-based knowledge-intensive query with SNOMED CT concepts in addition to name variants, and domain variants. This work as a continuation of our previous work [20,21], essentially contributes to bring effectiveness to the clinical systems through relevant evidence retrieval methods. We use PubMed search service by implementing Entrez Utility API [22] to automatically retrieve the potential evidences from MEDLINE databases [23]. The retrieved evidences are evaluated through validated model with text processing machine learning algorithms such as bootstrap, k-nearest neighbor, and others [24]. The training data is prepared and preprocessed to learn the model offline and utilized for query results evaluation. Finally, the classified set of evidences are ranked based on similarity score checked with gold standard articles prepared with the help of domain experts. The ranked set of evidences is presented to the domain expert for final appraisal and approval.

focuses on synonym terms for a concept, so descriptions file is used to acquire the synonym terms for query expansion. For instance, if we are looking to expand a concept "oral cavity", the synonym terms acquired from the descriptions include "entire oral cavity (body structure)", "entire oral cavity", "mouth", and "oral cavity" concepts (Table 1 last column). In our previous work [25] related to domain ontology extraction, we provided details about identifying a concept in domain ontology extracted from overall SNOMED CT ontology.

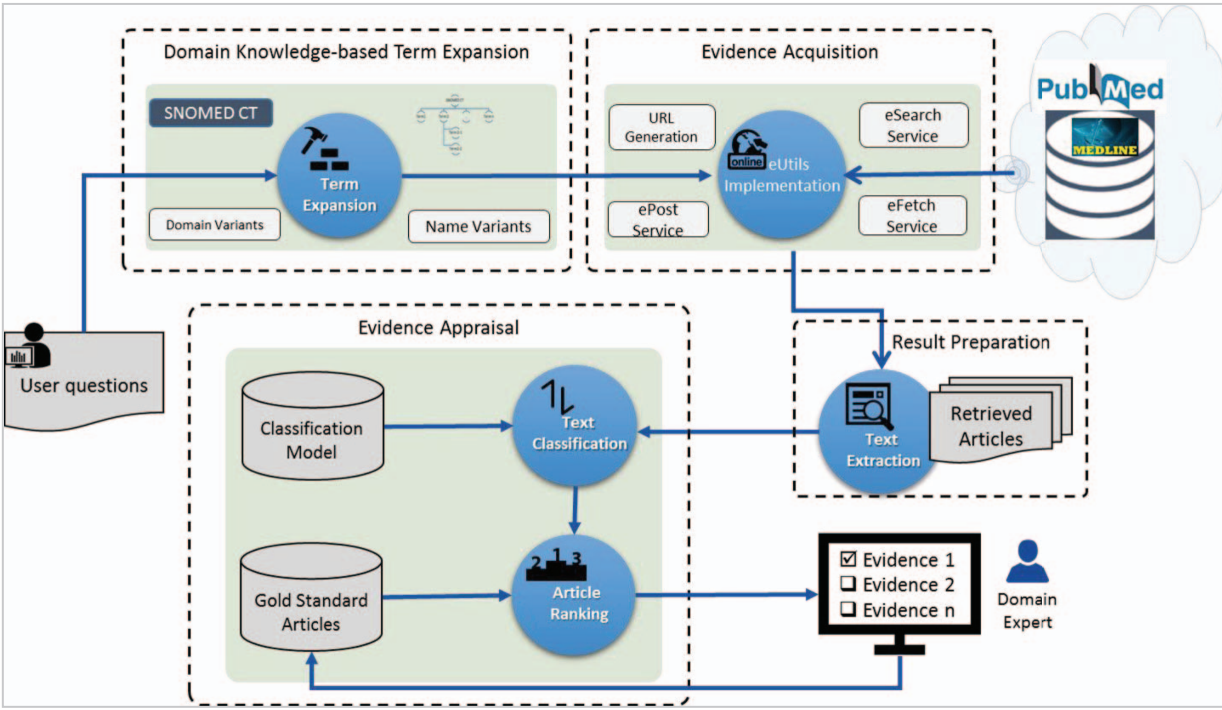


Figure 1: Evidence Retrieval and Appraisal Model based on Knowledge-instensive Queries

II. METHODOLOGY

We approach the problem by first building knowledge-intensive queries for retrieving the relevant MEDLINE articles, then extracting the title and abstracts from the retrieved articles, then appraising the evidence by text classification and ranking, and finally applying the evidence to be part of the clinical practice. In this process, we employ a series of methods described in this section and are shown in Figure 1.

A. QUERY REFORMULATION

Taking user query as input, firstly, base concepts of SNOMED CT are identified using mappings, partially described in Table 1, followed by Synonym variations acquired from IHTSDO (International Health Terminology Standards Development Organization) SNOMED CT Browser[1].

A.1. SNOMED CT Variants

IHTSDO manages SNOMED CT terminologies in three types of files; concepts, relationships, and descriptions. Concepts file hold the set of actual concepts; relationships file contains the parent, child, and sibling relationships; and descriptions file consists of synonym concepts. This work only

A.2. Domain Variants

Another level of term expansion is made with alternative terms by designing variant lexicon created with the support of domain experts involved in this work. For "oral cavity" as an example, the list of alternative terms includes "oral", "lip", and "oral tongue" concepts. This type of expansion minimizes the possibility of missing local variations used for a standard concept. The limitation of this strategy is the dependency on domain experts, however, it can be resolved by knowledge of domain.

A.3. Name Variants

For name variants that are commonly used in clinical practice, we employed the strategy of automatically generating the variants using manually crafted heuristics similar to [5]. For instance, clinical staging value "Clinical Stage 1", a variant "Clinical Stage I" is generated according to the heuristic of converting Arabic numerals to Roman numerals. The same can appear without space "Clinical StageI" or with hyphen "Clinical Stage-I" where we employ the rules of space removal and hyphen insertion. The numeral variations are generated for numbers with simple replacements. The process initially works

[1] IHTSDO (International Health Terminology Standards Development Organization)

SNOMED CT Browser, http://browser.ihtsdotools.org/

to generate the name variants with space and hyphen and add the numeral variants accordingly. For example, clinical stage I is first expanded to Clinical StageI and ClinicalStage-I followed by expansion set of Clinical Stage1 and ClinicalStage-1. Name variants for two important topics used in head and neck cancer are shown in Table 2.

A. Evidence Acquisition

The expanded query is executed to acquire the relevant evidence from PubMed databases through the implementation of PubMed searching service called Entrez Programming Utilities (eUtils) [22]. The eUtils provide a stable interface into the Entrez query and database system including 23 databases covering a variety of biomedical data.

To access these data, a piece of software posts eUtils URL to database in order to retrieve the results. Using eUtils, we build PubMed URL consisting of "Base URL" and the query terms. We employ the automatic term mapping (ATM) process provided by PubMed [11]. ATM uses translation via MeSH for indexing and searching of the MEDLINE database of journal citations. A neglected term in query is added to the MeSH term of original query to access MeSH field of MEDLINE documents. We implemented two important services eSearch and eFetch. The output of eSearch is passed to eFetch to get the final instances of MEDLINE articles.

Table 1: Source Concepts Expanded with SNOMED CT Variants (Description Values)

Source Concept	SNOMED CT Base Concept	SNOMED Variants (Description Values)
Oral Cavity	Oral Cavity	Entire oral cavity, Entire oral cavity, Mouth, Oral cavity
C, Chemotherapy	Chemotherapy	Chemotherapy (procedure), CT – Chemotherapy, Chemotherapy
R, Radiotherapy	Radiotherapy	Radiation oncology AND/OR radiotherapy, RT – Radiotherapy, Radiation oncology AND/OR radiotherapy, Radiothera py
S, Surgery	Surgery	Surgical procedure, Operation, Op erative procedure
IC, Induction Chemotherapy	Induction Chemotherapy	Induction chemotherapy, (induction chemotherapy
T1	T1 Stage	T1 category, T1 stage, Tumor stage T1,Tumour stage T1
N0	N0 Stage	N0 category, N0 stage, Node stage N0
Stage I, Clinical stage I	Clinical Stage I	Clinical stage finding, Tumor stage finding, Clinical stage I
Squamous cell carcinoma	Squamous cell carcinoma	Squamous cell carcinoma

A. Data Preprocessing

Table 2: Example of Name Variants for Clinical Staging and Lymph Nodes

Clinical Staging		Lymph Nodes	
Clinical Stage I	Clinical-StageI, ClinicalStage-I	Lymph Node I	Lymph NodeI, LymphNode-I
Clinical Stage II	Clinical StageII, ClinicalStage-II	Lymph Node II	Lymph NodeII, LymphNode-II
Clinical Stage III	Clinical StageIII, ClinicalStage-III	Lymph Node III	Lymph NodeIII, LymphNode-III
Clinical Stage IV	Clinical StageIV, ClinicalStage-IV	Lymph Node IV	Lymph NodeIV, LymphNode-IV

As a pre-step for text classification, the text in title and abstracts is extracted from the retrieved articles using algorithm described in *Algorithm 1*:

Algorithm 1: Data Preparation

```
-----------------------------------------
For each article in the set
    Extract tile and abstract
    For title and abstract words
        Create word vector with schema of TF-IDF (Term
        Frequency- Inverse Document Frequency) and absolute
        Pruning method with pruning all the words having below 3
        absolute value or above 100 absolute value.
        Tokenize the words using linguistic tokens of English
        Transform all the words into lower case
        Filter the stop words
        Porter-stem all words
    End
End
----------------------------------------------
```

B. Evidence Appraisal

The prepared data is passed to classify the text using the already learned classification model on training data in order to get the classified articles. The training model is built on training data collected from NCCN (National Cancer Comprehensive Network) guidelines. Detailed description of the data is described in data collection part here.

B.1. Data Collection

We prepared the training data set from MEDLINE databases based on the references used in NCCN guideline preparation for head & neck cancer. This data set consists of 452 MEDLINE documents, classified across two dimensions: diagnosis sites (oral cavity, salivary glands, larynx, and head neck) and treatment modality (surgery, radiotherapy, and induction chemotherapy). For this study, we focused on treatment modality by utilizing a subset of collection consisting of 104 documents in training set for cross-validation.

B.2. Corpus Preparation

For the collected data set, unique PubMed identifiers called PMID are acquired from the references. The PMIDs are posted to PubMed database using ePost service of eUtils API. The MEDLINE articles of the posted PMIDs are retrieved through the implementation of eSearch and eFetch services of eUtils. After retrieving the articles, title and abstract text is retrieved using *Algorithm 1*.

B.3. Machine Learning Methods

We experimented with four supervised machine learning methods: Naïve Bayes (Kernel), kNN, Poly SVM, and Bootstrap (Bagging). Among these three classifiers we select only Naïve Bayes (Kernel) and kNN with k=5, based on their better performance. Brief specifications of the selected two algorithms are described in Table 3.

III. Experimental Results

We applied these methods in RapidMiner [26]: a well-known tool used for text processing and classification applications. In training, we split the data set into ten sets of equal size and performed 10-fold cross-validation. We evaluated the results on two grounds; first, is to prove concept-based knowledge-intensive queries produce improved results than simple term-based queries; second, is to prove that automatic query generation and expansion reduce considerable amount of time over manual queries. The retrieval performance was measured for recall (the proportion of retrieved relevant documents by the query in the overall relevant document set), precision (the proportion of retrieved relevant documents in the overall retrieved document set), and overall accuracy. We executed 9 queries, 3 from each categories of surgery, induction chemotherapy, and radiotherapy and evaluated the results on top 100 articles retrieved in each category.

We here present both the classification results obtained in the ten-fold cross-validation performed on training set and the esults obtained on test data set containing documents retrieved by running the simple term-based as well as concept-based knowledge-intensive queries. The model accuracy of selected algorithms in validation stage. As shown in Table 4, kNN with k = 5 performed with better accuracy of 82.64% as compared to Naïve Bayes (Kernel) which showed accuracy of 79.55%. Individually, Naïve Bayes (Kernel) showed better performance in precision (90.00%) for induction chemotherapy and kNN with recall (91.43%) for induction chemotherapy.

A. Relevant Evidence Appraisal

We check the exact similarity of retrieved MEDLINE articles by executing example queries in each category of simple term-based and concept-based knowledge-intensive queries. Our hypothesis was that knowledge-intensive queries results in improved results as compared to simple term-based queries and the results in Table 5, proves the hypothesis of efficient retrieval by showing improved accuracy for queries with domain knowledge in contrast to simple term-based queries. The overall accuracy score of Naïve Bayes (Kernel) for knowledge-intensive queries noticed is 38.18% which is 20.93% improved than the average recall score of 31.69% for simple term-based queries. Similarly, kNN accuracy for knowledge-intensive is noticed as 50.39% which is 3.74% better than accuracy of 48.57% for simple term-based queries as shown in Table 5.

Table 3: Machine Learning Methods Specifications used in study

Method	Specifications
Naïve Bayes (Kernel)	As a probabilistic classifier, Naïve Bayes estimates the probability of class C using training data by applying the maximum a posteriori (MAP) rule. In contrast to Naïve Bayes, Naïve Bayes (Kernel) is applied with default setting of Kernel value = 10.
kNN	The k_Nearest Neighbor algorithm is based on learning lazy algorithm by analogy. It performs very well when the data set belongs to one domain sharing similar characteristics. We select k value = 5, measuretypes = NumericalMeasures, and numerical measure = CosineSimilarity.

Table 4: Classifier Performance at cross-validation stage measured in Recall, Precision, and Accuracy

Classifier	Class Value	Recall (%)	Precision (%)	Overall Accuracy (%)
Naïve Bayes (Kernel)	Surgery	**82.86**	76.32	79.55
	Induction Chemotherapy	77.14	**90.00**	
	Radiotherapy	64.41	75.00	
kNN	Surgery	77.14	79.41	82.64
	Induction Chemotherapy	**91.43**	82.05	
	Radiotherapy	79.41	**87.10**	

Table 5: Performance Measurement of Retrieved Test Data Set for simple term-based queries and knowledge-intensive queries.

Query Type	Classifier	Class Value	Recall (%)	Precision (%)	Overall Accuracy (%)
Simple Term-based Queries	Naïve Bayes (Kernel)	Surgery	**48.81**	26.28	31.69
		Induction Chemotherapy	37.14	34.90	
		Radiotherapy	18.01	**36.25**	
	kNN	Surgery	48.81	33.88	48.57
		Induction Chemotherapy	**61.43**	**56.95**	
		Radiotherapy	37.27	53.10	
Concept-based Knowledge-Intensive Queries	Naïve Bayes (Kernel)	Surgery	**54.10**	34.20	38.18
		Induction Chemotherapy	31.93	34.55	
		Radiotherapy	29.86	**52.44**	
	kNN	Surgery	44.26	44.63	50.39
		Induction Chemotherapy	**65.55**	**56.52**	
		Radiotherapy	43.06	49.21	

B. MAXIMIZATION OF THROUGHPUT

Automatic query construction followed by expansion saved considerable time in terms of manual query generation. Manually, if these queries could be written by users (clinicians), it might not be practical or extremely hard to control. Even an expert clinician can write an average query, it takes around 2 minutes on average query construction (provided that all the domain data is available which seems impractical). We perform experiments to write manually three type of queries: simple (consisting < 5 concepts), average (consisting between 6 and 20 concepts), and complex (consisting > 20 concepts) queries. These queries are written by two kind of users; expert and average. Expert has good domain knowledge as well as fair in typing, while average user possess average domain knowledge and fairly good in typing. We set the experiment by printing the corresponding queries on paper, open the PubMed browser to write the query, and record the time spent on each query as shown in Figure 3. On average the expert user spent 4.7 minutes while average user spent 8.5 minutes on writing different queries. During experiment, we neglected the mistakes made during writing. A reasonable amount of time is saved through automation of the query process which ultimately added to maximize the throughput of the clinicians at clinical practice

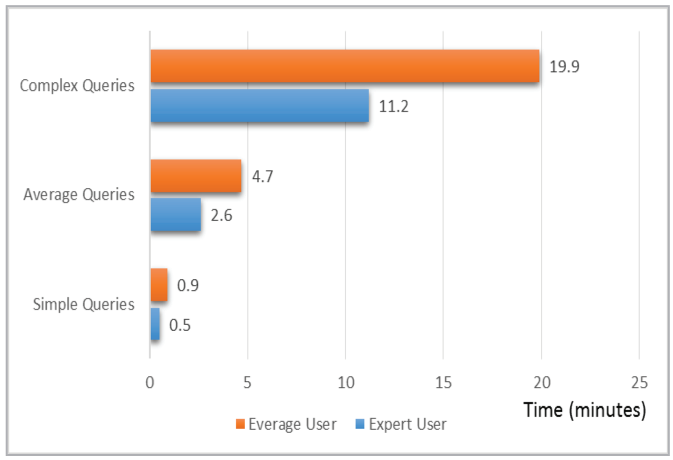


Figure 3: Query-by-query writing time in minutes by expert and average user

IV. CONCLUSION

Automatic reformulation of concept-based knowledge-intensive queries is an important task to acquire relevant evidences from the literature. The proposed work present a novel model employing methods of concept-based knowledge-intensive query reformulation, evidence acquisition, and evidence appraisal. From the results, we infer that adding knowledge query automatically to the query, on one hand, saved the tedious work of manual query writing, while on the other hand, retrieved potentially relevant evidences. Our next plan is to improve the retrieval performance by including semantic similarity methods and concrete appraisal of retrieved documents.

ACKNOWLEDGEMENT

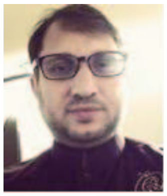
This work was supported by the Industrial Core Technology Development Program (10049079 , Development of mining core technology exploiting personal big data) funded by the Ministry of Trade, Industry and Energy (MOTIE, Korea)".

REFERENCES

1. Sackett, D.L.; Rosenberg, W.; Gray, J.A.; Haynes, R.B.; Richardson, W.S. Evidence based medicine: What it is and what it isn't. *Bmj* **1996**, *312*, 71-72.
2. Sackett, S.; Richardson, R. Evidence-based practice. *Found Evidence-Based Social Work Pract* **2006**, *35*.

3. Aronson, A.R.; Rindflesch, T.C. In *Query expansion using the umls metathesaurus*, Proceedings of the AMIA Annual Fall Symposium, 1997; American Medical Informatics Association: p 485.
4. Hersh, W.; Price, S.; Donohoe, L. In *Assessing thesaurus-based query expansion using the umls metathesaurus*, Proceedings of the AMIA Symposium, 2000; American Medical Informatics Association: p 344.
5. Zhong, M.; Huang, X. In *Concept-based biomedical text retrieval*, Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval, 2006; ACM: pp 723-724.
6. Perez-Rey, D.; Jimenez-Castellanos, A.; Garcia-Remesal, M.; Crespo, J.; Maojo, V. Cdapubmed: A browser extension to retrieve ehr-based biomedical literature. *BMC medical informatics and decision making 12*, 29.
7. Zhou, W.; Yu, C.; Smalheiser, N.; Torvik, V.; Hong, J. In *Knowledge-intensive conceptual retrieval and passage extraction of biomedical literature*, Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, 2007; ACM: pp 655-662.
8. Kim, Y.S.; Kang, B.H. Search query generation with mcrdr document classification knowledge. In *Knowledge engineering: Practice and patterns*, Springer: 2008; pp 292-301.
9. Kakade, V.; Sharangpani, M. Improving the precision of web search for medical domain using automatic query expansion. *On-line]. http://www.stanford.edu/class/archive/cs/cs276a/cs276a* **2004**, *1032*.
10. Liu, S.; Liu, F.; Yu, C.; Meng, W. In *An effective approach to document retrieval via utilizing wordnet and recognizing phrases*, Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval, 2004; ACM: pp 266-272.
11. Lu, Z.; Kim, W.; Wilbur, W.J. Evaluation of query expansion using mesh in pubmed. *Information retrieval* **2009**, *12*, 69-80.
12. Wilczynski, N.L.; Morgan, D.; Haynes, R.B. An overview of the design and methods for retrieving high-quality studies for clinical care. *BMC medical informatics and decision making* **2005**, *5*, 20.
13. Kilicoglu, H.; Demner-Fushman, D.; Rindflesch, T.C.; Wilczynski, N.L.; Haynes, R.B. Towards automatic recognition of scientifically rigorous clinical research evidence. *Journal of the American Medical Informatics Association* **2009**, *16*, 25-31.
14. Aphinyanaphongs, Y.; Tsamardinos, I.; Statnikov, A.; Hardin, D.; Aliferis, C.F. Text categorization models for high-quality article retrieval in internal medicine. *Journal of the American Medical Informatics Association* **2005**, *12*, 207-216.
15. Cimino, J.J.; Elhanan, G.; Zeng, Q. In *Supporting infobuttons with terminological knowledge*, Proceedings of the AMIA Annual Fall Symposium, 1997; American Medical Informatics Association: p 528.
16. Del Fiol, G.; Rocha, R.A.; Clayton, P.D. In *Infobuttons at intermountain healthcare: Utilization and infrastructure*, AMIA Annual Symposium Proceedings, 2006; American Medical Informatics Association: p 180.
17. Currie, L.M.; Graham, M.; Allen, M.; Bakken, S.; Patel, V.; Cimino, J.J. In *Clinical information needs in context: An observational study of clinicians while using a clinical information system*, AMIA Annual Symposium proceedings, 2003; American Medical Informatics Association: p 190.
18. Cimino, J.; Del Fiol, G. Infobuttons and point of care access to knowledge. *Clinical decision support-the road ahead* **2007**, 345-372.
19. Cimino, J.J. In *Use, usability, usefulness, and impact of an infobutton manager*, AMIA Annual Symposium Proceedings, 2006; American Medical Informatics Association: p 151.
20. Afzal, M.; Hussain, M.; Khan, W.A.; Ali, T.; Lee, S.; Kang, B.H. In *Knowledgebutton: An evidence adaptive tool for cdss and clinical research*, Innovations in Intelligent Systems and Applications (INISTA) Proceedings, 2014 IEEE International Symposium on, 2014; IEEE: pp 273-280.
21. Afzal, M.; Hussain, M.; Ali, T.; Khan, W.A.; Lee, S.; Kang, B.H. Mlm-based automated query generation for cdss evidence support. In *Ubiquitous computing and ambient intelligence. Personalisation and user adapted services*, Springer: 2014; pp 296-299.
22. Sayers, E.; Wheeler, D. Building customized data pipelines using the entrez programming utilities (eutils). **2004**.
23. Citations added to medline by fiscal year, http://www.Nlm.Nih.Gov/bsd/stats/cit_added.Html. **2014**.
24. Kotsiantis, S.B.; Zaharakis, I.; Pintelas, P. Supervised machine learning: A review of classification techniques. 2007.
25. Ali, T.; Hussain, M.; Khan, W.A.; Afzal, M.; Lee, S. In *Customized clinical domain ontology extraction for knowledge authoring tool*, Proceedings of the 8th International Conference on Ubiquitous Information Management and Communication, ACM: p 23.
26. Rapidminer, https://rapidminer.Com/. **Accessed: May 31 2015**.

AUTHORS BIOGRAPH



Muhammad Afzal is a Ph.D. student in Ubiquitous Computing Laboratory (UCLab), Department of Computer Engineering, Kyung Hee University, South Korea. He got his M.S Degree in Information Technology from School of Electrical Engineering and Computer Science, National University of Sciences and Technology (NUST), Pakistan in 2009. He secured his B.S degree in Computer Science from Kohat University of Science and Technology (KUST). He has been serving as proctor of HL7 International for conducting HL7 Standard Certifi- cation Exams in Pakistan since 2010. His current research includes Evidence-based Knowledge Acquisition and Systems, Applications of Machine Learning, Text Processing, and Information Extraction.
Mr. Muhammad Afzal is HL7 International Member since 2010.

Sungyoung Lee received his B.S. from Korea University, Seoul, Korea. He got his M.S. and Ph.D. degrees in Computer Science from Illinois Institute of Technology (IIT), Chicago, USA in 1987 and 1991 respectively. He has been a professor in the department of Computer Engineering, Kyung Hee University, Korea since 1993. He is a founding director of the Ubiquitous Computing Laboratory, and has been affiliated with a director of Neo Medical ubiquitous-Life Care Information Technology Research Center, Kyung Hee University since 2006. Before joining Kyung Hee University, he was an assistant professor in the Department of Computer Science, Governors State University, llinois, USA from 1992 to 1993. His current research focuses on Ubiquitous Computing and Applications, Wireless Ad-hoc and Sensor Networks, Context-aware Middleware, Sensor Operating Systems, Real-Time Systems and Embedded Systems, Activity and Emotion Recognition.
Prof. Sungyoung Lee is a member of ACM.