

Towards a Semantically Driven E-learning Framework

Suhare M. Solaiman

Department of Computer Science,
King Abdulaziz University, Jeddah
Saudi Arabia

Imtiaz Hussain Khan

Department of Computer Science,
King Abdulaziz University, Jeddah
Saudi Arabia

Muazzam Ahmed Siddiqui

Department of Information
System, King Abdulaziz University,
Jeddah Saudi Arabia

ABSTRACT

E-learning offers great benefits over the conventional learning process. However, the huge unstructured information, which is freely available on the Web poses significant challenges in accessing the desired information in a timely manner. To tackle this problem different information retrieval (IR) approaches have been proposed in literature. These approaches are predominantly influenced by classical keyword-based IR techniques. However, with recent technological advances and a flood of information on the Web, the performance of keyword-based IR techniques has greatly suffered. Therefore, recently some more intelligent IR techniques have been proposed to enhance the utility of e-learning systems. In this study, a semantically oriented ontology-based personalized framework is proposed for effective e-learning. The proposed framework is implemented and its effectiveness is thoroughly assessed as a case study to learn Java programming language. The proposed system is evaluated on an indigenous medium-sized corpus (≈ 2600 documents) in terms of standard accuracy measures for IR. The findings in this paper reveal that semantic based IR for e-learning is a robust methodology and it can advance the field of e-learning in an elegant manner.

General Terms

Information retrieval for e-learning

Keywords

Information retrieval, Semantic similarity, Semantic annotation, Keywords-based retrieval, Ontology, Query expansion

1. INTRODUCTION

Currently, the Web is undoubtedly the largest repository of information on diverse topics. The Web offers a vast medium for sharing plethora of information and services, which in the absence of the Web could have been a dream. However, most of this information on the Web is not well structured or organized in a systematic manner. Therefore, users generally face significant difficulty in getting their desired information in a timely manner because mostly the information is heterogeneous and stored on distributed locations. The most common challenge faced by users is when they are looking for domain specific information because on the Web there is a lot of information available on different domains. The existing information retrieval (IR) systems are predominantly based on classical keyword-based search methods. In such methods, the content of documents collections is described by using different lists of keywords. However, the problem arises with this approach as the bare keywords do not capture the semantic content of documents and hence generally fail in retrieving the desired documents. Despite the fact that the performance of keyword-based IR has been deteriorating with

the increase in the information volume, these techniques are still popular among the researchers [1, 2].

Recently, there has been a boom in using e-learning for educational purposes. The e-learning systems generally attempt to organize learning content in such a manner that they should be easily accessible to the users on demand. However, it has been observed that these systems lack in usability, especially when information is required from different sources in a personalized manner. More specifically, in recent years different learning management systems (LMS) have been developed, which offer various features and support for distance learning. However, all these systems have some fundamental underlying limitations. Among these limitations, the most obvious one is that the learning content of an LMS is primarily specified by instructors and this content generally remains fixed. This static nature of learning content of the LMS does not allow the users to add information from other sources which is an important limitation [3, 4]. That is, even though the users of the LMS are main stakeholders, the system tends to meet the needs of organization and instructor disregarding the actual needs of the users; hence these models are not user centric. The existing LMSs also have some usability issues, for example, users normally get confused with different standards to be followed in teaching and learning [5]. Such systems generally lack in adapting to the personalized needs of students because they disregard their knowledge background and individual interests [6]. Recent advances in using ontology-based semantically driven information presentation to the users is paving the way to enhance the quality of IR systems for e-learning thereby increasing their acceptability and utility.

In this study, we proposed a semantically driven ontology-based IR model to build a personalized framework for effective e-learning. This is a user-centric model where a user can get structured information on demand from various different unstructured sources of information. The proposed framework is implemented to learn a fundamental Java programming language course. The usability and acceptability of the implemented system is thoroughly assessed. The proposed system is evaluated on an indigenous medium-sized corpus (≈ 2600 documents) in terms of precision and recall, which are well-known measures for IR. Our findings reveal that semantic based IR for e-learning is a robust methodology and it can advance the field of e-learning in an elegant manner.

2. RELATED WORK

In this section, we discuss the state of the art in e-learning and IR techniques used for e-learning. At present, the e-learning field is getting considerable attention and it has already made tremendous contribution in education sector, especially in distance learning [7]. This revolution in the education industry has been only possible with the development of modern

technology, especially the network equipment for instant information transmission and the different Web standards for providing learning solutions. As compared to classical education systems, e-learning offers numerous benefits including flexibility, and it has generally been proven as more cost effective [8, 9]. With continuous advancement in the e-learning field, organizations are investing a lot in this sector and a lot of active research is going on how to make e-learning scalable so that large pool of participants can benefit from the same program with minimum effort [10]. This way, it could be possible for the users to get the same content in the same fashion, thereby imitating the conventional class-room based education system. Even though e-learning has been among the circles for a while, the journey of e-learning systems can be classified into three major generations [11] for better understanding how the field has emerged. The first generation is called LMS that primarily focuses on making learning available online in a faster accessible manner. The most commonly used LMS software are Moodle, Blackboard and WebCT. These sophisticated software tools not only manage and present educational material in a systematic manner, they also offer appropriate tracking and record keeping of students' information that gives lot of confidence and acceptability in using such systems [12]. These systems are still growing and a lot of research is going on to enhance their quality. The next generation is broadly influenced by a revolution in the field of artificial intelligence, particularly how to present information to the users according to their needs in an adaptive and effective manner. During this period, major advancement in the field is the development of adaptive e-learning systems, which are more users centric. Finally, the third generation in e-learning is influenced by dramatic enhancement in the field of Semantic Web at the core of which is the idea of using domain-specific ontologies. Our current research is also an attempt to enhance e-learning systems by incorporating semantic-based ontology-driven content presentation to the users.

E-learning is only possible if sophisticated IR techniques are available to search for unstructured documents on very large repositories like the Web [13]. An IR system must be capable of four main processes, information representation, storage, organization and access, which are the core of an IR system [14]. The basic IR model starts by a query submission, goes through an intensive process of documents search in the repository, e.g. the Web, and finally returns the best matched results to the user [15]. In recent years, lot of efforts has been made to enhance the performance of IR systems, but the basic model still remains the same. In literature, a lot of work is reported on IR models, their performance evaluation techniques, and their strengths and limitations. Majority of earlier work on IR is dominated by keywords-based IR techniques [16, 17, 18]. In this model, a list of the most representative keywords is extracted from the document to form the document signature in terms of a query. This query is then submitted to search for the source documents in the desired repository. However, the main limitation of this plain query based on bare keywords is that it does not capture the important semantic information and hence generally retrieves many unwanted documents, which are semantically very different than the one, which are intended.

One important study in keywords-based IR is due to [19], in which the author proposed to use vector space model (VSM). The VSM model paved the way for other researchers to build upon this model for more effective IR systems. Building on the VSM model, in [20], the authors proposed a term-weighting scheme, in which higher weight values are given to

those terms in a document, which offer more discriminative power. One such approach is term frequency (TF), where TF assigns weights to the terms based on their occurrences in the given document.

In recent years, there has been a departure from the keywords-based IR towards semantic-based IR, which is predominantly based on ontology-based methods. The underlying concepts in ontology-based techniques for IR are query expansion, semantic annotation, and visualization. In an earlier work on ontology expansion [21], the authors proposed a methodology whereby first the query terms were disambiguated. Subsequently, the related concepts to the disambiguated query term are added in the query to construct an extended query for search purposes. In a similar approach [22], a word-sense disambiguation technique is used for query expansion, in which the correct sense of a given term was determined from a set of available senses. In another study [23], the researchers proposed to use spatial terms in their ontology-based query expansion approach. They developed their spatial ontology-based query expansion framework for IR in tourism domain and found that it is an effective approach for better IR.

In another work [24], the authors built a prototype semantic search engine, which exploits semantic information to retrieve documents. They also attempted to resolve semantic ambiguity in a domain-specific manner where the domain specific knowledge was encoded in the search engine. They also combined their ontology with local context analysis technique [25, 26] and found that it is more effective than the classic keywords-based IR. In another study [27], a set of ontological relations were used for query expansion purposes. They evaluated the performance of their approach on gene ontology and found that their approach offers better accuracy.

In a work [28], an ontology-based method is proposed for the semi-automatic annotation of documents to enhance the search performance of IR system. They used the traditional VSM model and applied an annotation weighting algorithm in a newspapers domain in Spanish language. They compared the performance of their approach with keywords-based approach and found that the former outperformed the latter in attaining better accuracy results. Building on the work of [28], the authors in [29] built an ontology-based IR system for banking domain in Tamil documents. They shown that the annotation weighting scheme works better and is language independent. In another study [16], a hybrid approach is pursued in which the IR model combines keywords-based search with semantic-based search and found that the hybrid approach offers better performance than the individual approaches. In a similar work [30], another hybrid approach is adopted by combining ontology-based IR with keywords-based search. The authors shown empirically that their approach works better than simple keywords and query expansion approaches. In another work [31], a semantic indexing approach is combined with an ontology-based IR model. Their evaluation revealed impressive results. In [32], the authors built an annotation model by interviewing scientists to develop a discourse-level annotated document. They showed that theirs is an effective approach to search for queries based on scientists needs. In [33], an ontology-based scheme is used for e-learning; evaluation results showed better accuracy for the IR system.

Query expansion and annotation schemes are important for IR, but nevertheless browsing could prove useful when users are uncertain about the information they are looking for. Browsing could be used to grasp the general view of a domain and hence can aid users in locating their desired information

[34, 35]. Moreover, it has been argued that effective visualization can improve the performance of document retrieval task [36]. In [37], an empirical approach was carried out to investigate the effectiveness of ontology visualization; the authors found that ontology visualization is an effective way to enhance the performance of an IR system. The work discussed so far in this section mainly lacks in personalized document retrieval, which is primary aim of our research.

3. SIMILARITY AND EVALUATION MEASURES FOR IR SYSTEMS

Literature on IR suggests that a lot of research has been conducted to develop effective document similarity measuring techniques. Many similarity measures have been proposed ranging from character-level similarity to word/lexical level similarity, and even more advanced semantic-based similarity techniques. The performance of an IR system largely depends on the similarity measure used to compare the similarity or dissimilarity between the query and the source document. One of the most popular semantic similarity approaches is edge-based similarity, which measures the similarity between two concepts in ontology. In this approach, ontology is realized as an undirected graph and the distance between nodes is a measure of similarity; shorter the distance, more similar the concepts are and vice-versa [38, 39]. In contrast to edge-based similarity, the node-based similarity measure computes the similarity between two nodes based on the content they share, which is commonly known as information content [40]. Building on earlier work by [40], Lyn [41] proposed a information-theoretic similarity measure which is based on the core idea, which assumes that the most the two concepts have in common, more similar they are and vice-versa.

The earlier similarity approaches are primarily based on character-level similarity measures, including edit distance [42] and Levenshtein distance [43]. Even though the character-level measures have been on the centerstage for a longer period of time, some term-based similarity measures were also proposed, for example cosine similarity. Another interesting term-based similarity measure is dice coefficient [44]. Whereas character-level similarity measures use various n-grams models to compare the documents character by character, such techniques generally fail in locating the paraphrased documents. On the other hand, semantic-based similarity measures are generally more effective but at the same time, they are computationally expensive.

The performance of an IR model should be measured in an objective manner to enhance the acceptability of the system. Ideally, an IR system should be able to retrieve all relevant documents and disregard all irrelevant documents. However, generally it has been observed that many state of the art IR systems, even modern search engines like Google, retrieve many irrelevant documents. Therefore, IR community has proposed different evaluation measures to gauge the effectiveness of an IR system and most commonly used evaluation precision and recall [45]. Precision is the ratio of related retrieved documents to the related documents as shown in Equation (1). On the other hand, the recall measures the ratio of related retrieved documents to the relevant documents as shown in Equation (2).

$$Precision = \frac{|{\{Relevant\ documents\}} \cap {\{Retrieved\ documents\}}|}{|{\{Retrieved\ documents\}}|} \quad (1)$$

$$Recall = \frac{|{\{Relevant\ documents\}} \cap {\{Retrieved\ documents\}}|}{|{\{Relevant\ documents\}}|} \quad (2)$$

Often there exists a tradeoff between precision and recall. Generally, higher the precision, lower the recall; for instance when the recall approaches to 1, precision tends to drop to 0 and vice-versa [46]. Therefore, which of the two measures should be maximized is dependent on the task and also the choice of the researcher, that is why generally, an f-measure, which is calculated, based on precision and recall as shown in Equation (3), is also reported along with precision and recall.

$$F - score = \frac{(1+\beta)*Precision*Recall}{Precision+\beta^2*Recall} \quad (3)$$

4. PROPOSED FRAMEWOK

Our proposed framework is an extension to the basic IR model [15] and it exploits semantic information in a systematic manner to enhance the performance of IR. The proposed semantic-based framework is shown in Figure 1, which consists the following main components: semantic indexing, query processing, query expansion, searching, and ranking. Among these components, the most salient one, which distinguishes our framework from the traditional basic model, is semantic indexing. This component starts by creating a semantic index, which uses weighted annotated scheme to link the Web documents with the ontological concepts for further processing. The main steps of our semantic-based IR framework are outlined below.

- The document retrieval process starts by a user query, which is a sequence of keywords. This query is then submitted to the query processing component wherein the query is mapped to ontological concepts in Java programming language (henceforth JPL).
- The query expansion component exploits the JPL concepts available in the domain ontology in a systematic manner and finds the most similar concepts from the domain and appends those with the concepts in the given user query. This way an expanded query is constructed, which is subsequently used by the searching and ranking components.
- The searching component attempts to find the most closely matched documents from the repository, the Web in our cases, and the ranking component ranks those documents based on their relevance to ontological concepts. Finally, the top ranked documents are returned to the user; in our case, we decided to retrieve only top 15 documents as most successful match.
- All the components have access to the ontological knowledge base (KB) which is implemented as various classes. Our KB is based on Java learning object ontology [47] that covers basic Java programming concepts. The ontology is structured as a taxonomy of concepts in a hierarchical fashion where concepts denote the topics, properties are description of those topics and the connections among the topics are realized as relations.

Unlike the traditional keywords-based IR model, which uses inverted index scheme where keywords point to the documents in which they appear, in our framework domain-specific ontological concepts point to the document in which those concepts appear.

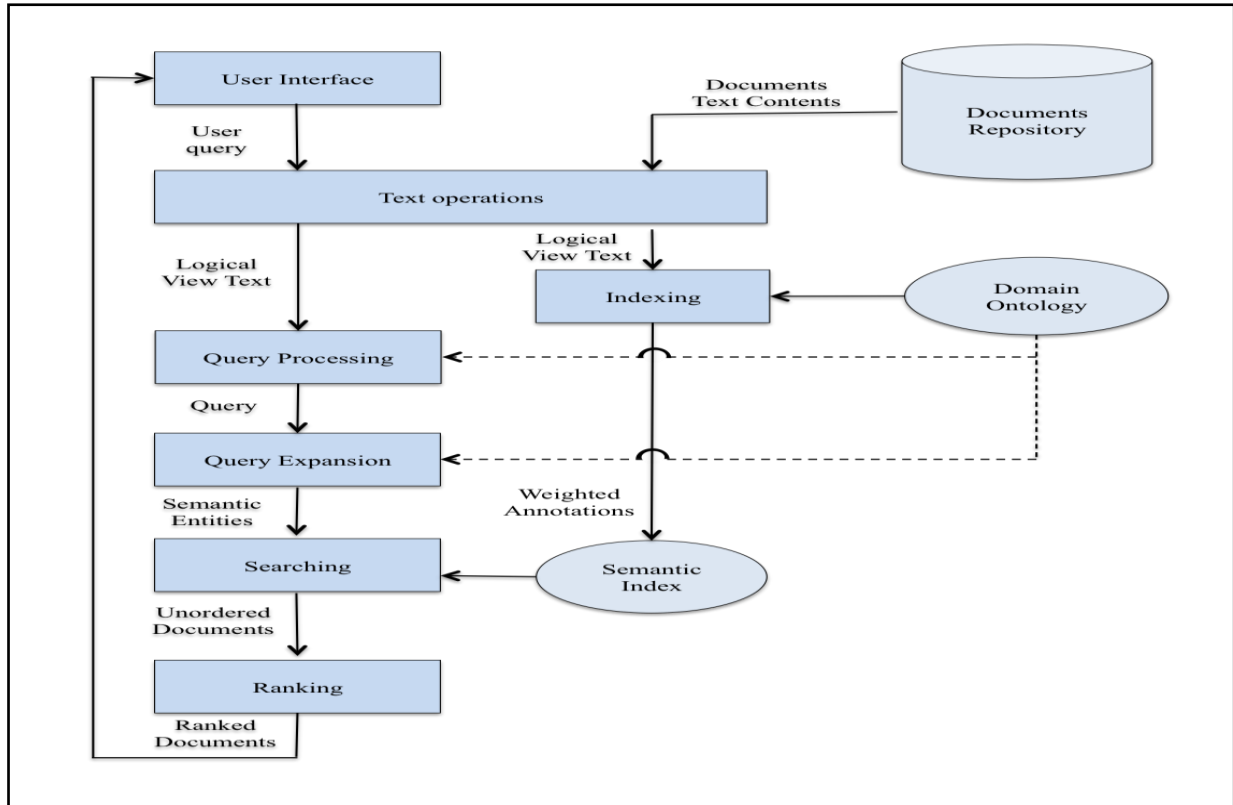


Figure 1: The proposed semantic-based IR framework

5. EVALUATION

We evaluated the performance of our semantic-based IR framework against the widely used keywords-based IR. The accuracy of the two systems is measured in terms of precision and recall. We used interpolated precision over standard recall

levels, and top 15 retrieved documents were used for these computation purposes. The results are shown in Figure 2, which elucidate that our approach outperforms the standard keywords-based IR in terms of attaining better precision and recall.

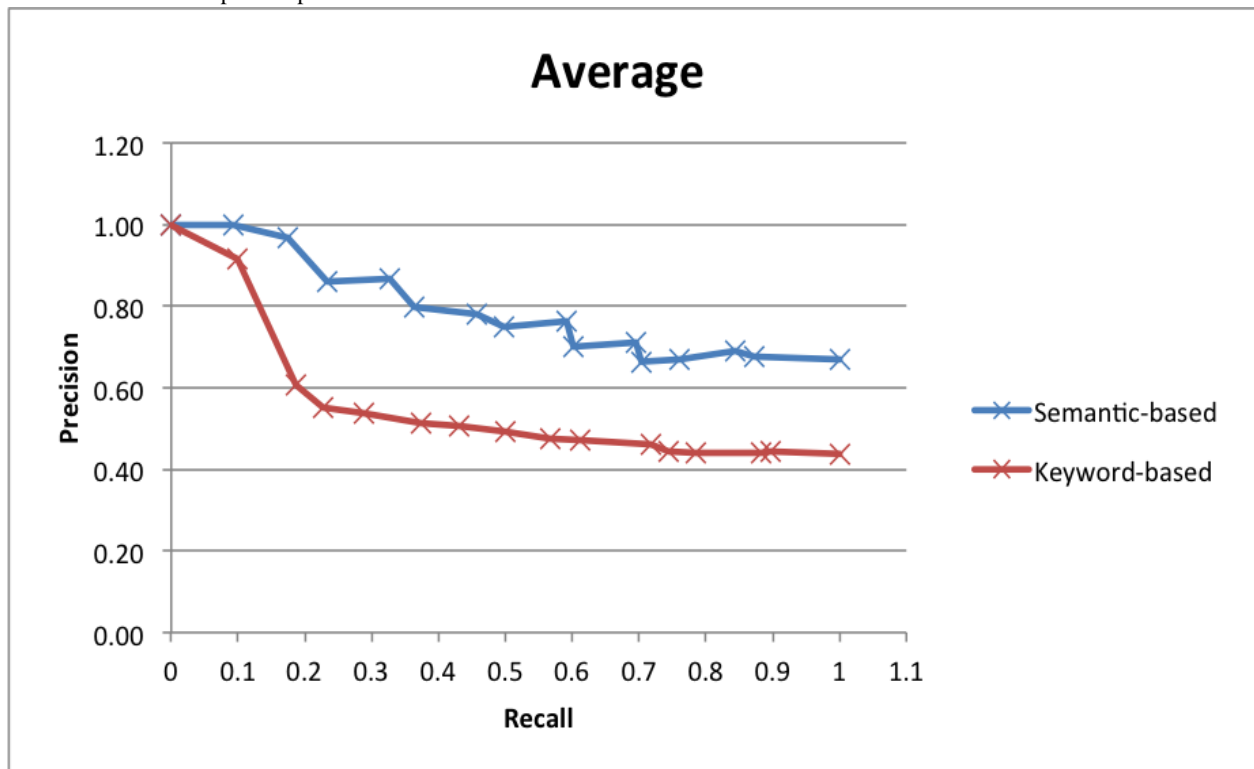


Figure 2: Evaluation of semantic-based IR vs. keywords-based IR

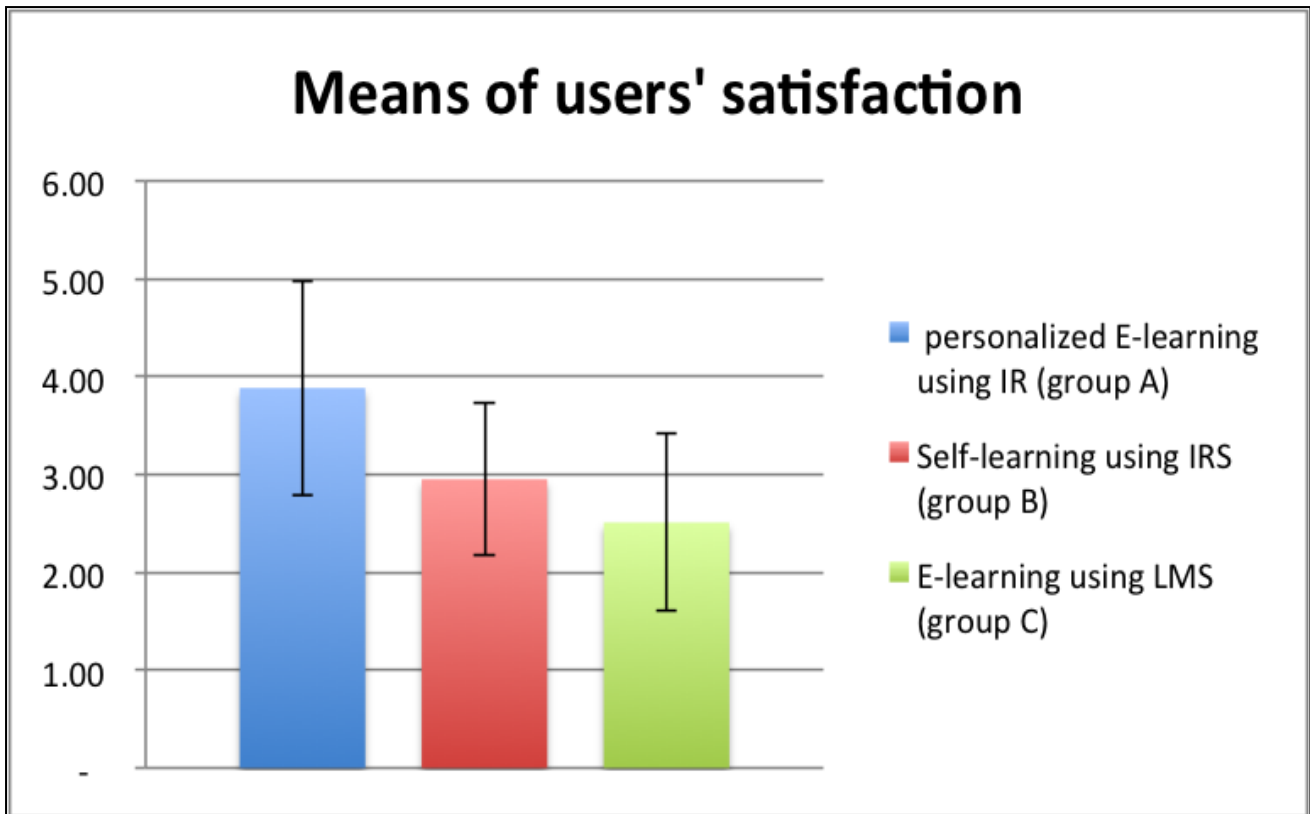


Figure 3: User satisfaction on the acceptability of our proposed e-learning system

We also evaluated the acceptability of our adaptive e-learning framework by conducting an experiment with human participants who were students of an introductory Java course. We used a well-defined questionnaire developed by [48], which is intended to measure the usability of Web-based learning in a systematic manner. The participants were divided into three groups each comprising 24 participants. We compared the performance of our system against LMS and IRS and found that users acceptability on a 5-point scale of our system (Mean = 3.8, STD = 1.2) is better than that of LMS (Mean = 3.05, STD = 1.1) as shown in Figure 3. These results suggest that adaptive e-learning framework backed by semantic-based IR is more effective than the existing state of the art e-learning systems.

The results in this paper generally indicate that in order to build an effective e-learning system, it must be adaptive to user's needs. Moreover, the performance of such a system is highly dependent on the underlying IR model. The traditional keyword-based IR models could offer better time efficiency as compared to semantic-based models as the latter require more computational resources. Moreover, the domain-specific ontology development also poses significant challenges as it is generally difficult to port ontologies from one domain to other domain. Therefore, if an ontology is created for programming domain, it would not work for example in information security domain as there is little overlap between the two domains. Moreover, there is an important challenge of presenting specific information to users in a customized way, which makes it hard to develop adaptive e-learning systems. This is an important limitation in the existing e-learning systems, which are primarily based on keywords-based IR models. Recently, interests have been developed in the e-learning research community to build adaptive e-learning

systems by incorporating ideas from artificial intelligence and other related discipline.

6. CONCLUSION

In this study, a novel method is presented, which uses semantic-based ontology driven IR approach for e-learning. The proposed framework offers personalized information in a timely manner and hence enhances the effectiveness of e-learning. It has generally been observed that e-learning provides an effective enhancement over the traditional learning process and many people worldwide are attracting themselves towards e-learning. Despite huge investments in e-learning projects, however, the field is still in infancy and faces significant challenges mostly because of the large amount of unstructured information on the Web. The problem becomes even cumbersome when a user tries to search in a specific domain, for example programming, to learn. To tackle this problem traditional keyword-based approaches generally degrade whereas modern semantically-oriented approaches offer better performance in providing accurate information in a timely manner. The proposed framework is implemented and its effectiveness is thoroughly assessed as a case study to learn Java programming language. The proposed system is evaluated thoroughly in terms of precision and recall. The findings in this paper elucidated that semantic based IR for e-learning is a robust methodology and it can advance the field of e-learning in an elegant manner.

There is still room for improvement in IR systems for e-learning, especially developing more sophisticated and rich ontologies, which should offer intelligent and optimized search results in a more personalized manner. The effective use of artificial intelligence techniques is already underway to develop more sophisticated adaptive e-learning systems and it seems to progress with pace in future.

7. ACKNOWLEDGMENTS

The authors thank reviewers for their invaluable feedback to improve the paper.

8. REFERENCES

- [1] Baazaoui, H., Aufaure, M. A., Soussi, R., Laboratoy, R. G., & Manouba, E. C. U. (2008). Towards an on-line semantic information retrieval system based on fuzzy ontologies. *Journal of Digital Information Management*, 6(5), 375.
- [2] Singh, P., Dhawan, S., Agarwal, S., & Thakur, N. (2015). Implementation of an efficient fuzzy logic based information retrieval system. *arXiv preprint arXiv:1503.03957*.
- [3] Linta, S. R., Khan, R., & Ahmed, F. (2011). Towards e-learning management system using semantic web technologies.
- [4] Garrido, A., & Morales, L. (2014). E-Learning and intelligent planning: improving content personalization. *Tecnologias del Aprendizaje, IEEE Revista Iberoamericana de*, 9(1), 1-7.
- [5] Hatakka, M., Avdic, A., & Andersson, A. (2007). SCORM: from the perspective of the course designer: a critical review. In *ECEL 2007: 6th European Conference on E-Learning: Copenhagen Business School, Denmark, 4-5 October 2007* (pp. 307). Academic Conferences Limited.
- [6] Raspopović, M., & Cvetanović, S. (2011). Implementation of adaptive e-Learning through Workflow Technology.
- [7] Rosenberg, M. J. (2001). *E-learning: strategies for delivering knowledge in the digital age*, (3). New York: McGraw-Hill.
- [8] Acampora, G., Loia, V., & Gaeta, M. (2010). Exploring e-learning knowledge through ontological memetic agents. *Computational Intelligence Magazine, IEEE*, 5(2), 66-77.
- [9] Tolea, E. E. (2012). A prototype for intelligent computer based training systems in business. *Global Journal on Technology*, 2.
- [10] Alsultanny, Y. A. (2006). E-learning system overview based on semantic web. *The Electronic Journal of E-Learning*, 4(2), 111-118.
- [11] Dicheva, D. (2008). Ontologies and semantic web for e-learning. In *Handbook on information technologies for education and training* (pp. 47-65). Springer Berlin Heidelberg.
- [12] Oneto, L., Abel, F., Herder, E., & Smits, D. (2009). Making today's learning management systems adaptive. In *Learning Management Systems meet Adaptive Learning Environments, Workshop at European Conference on Technology Enhanced Learning (ECTEL)*.
- [13] Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. New York, NY, Cambridge university press.
- [14] Baeza-Yates, R., & Ribeiro-Neto, B. (1999). *Modern information retrieval*. Harlow, England: Addison-Wesley.
- [15] Khan, L., McLeod, D., & Hovy, E. (2004). Retrieval effectiveness of an ontology-based model for information selection. *The International Journal on Very Large Data Bases*, 13(1), 71-85.
- [16] Fernández, M., Cantador, I., López, V., Vallet, D., Castells, P., & Motta, E. (2011). Semantically enhanced information retrieval: An ontology-based approach. *Web Semantics: Science, Services and Agents on the World Wide Web*, 9(4), 434-452.
- [17] Ishii, N., Suzuki, Y., Ito, M., & Hara, T. (2015). A new approach to web mining: A search engine offering result of no assumption. *Information Engineering Express*, 1(2), 51-59.
- [18] Thenmalar, S., & Geetha, T. V. (2014). Enhanced ontology-based indexing and searching. *Journal of Information Management*, 66(6), 678-696.
- [19] Salton, G. (1971). *The SMART retrieval system: Experiments in automatic document processing*. Upper Saddle River, NJ, Prentice-Hall, Inc.
- [20] Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing and Management*, 24(5), 513-523.
- [21] Voorhees, E. M. (1994). Query expansion using lexical-semantic relations. In *SIGIR'94* (pp. 61-69). Springer London.
- [22] Andreou, A. (2005). *Ontologies and query expansion*. (Master's thesis). School of Informatics, University of Edinburgh.
- [23] Fu, G., Jones, C. B., & Abdelmoty, A. I. (2005). Ontology-based spatial query expansion in information retrieval. In *On the move to meaningful internet systems 2005: CoopIS, DOA, and ODBASE* (pp. 1466-1482). Springer Berlin Heidelberg.
- [24] Gao, G., Liu, Y. S., Wang, M., Gu, M., & Yong, J. H. (2015). A query expansion method for retrieving online BIM resources based on Industry Foundation Classes. *Automation in Construction*, 56, 14-25.
- [25] Xu, J., & Croft, W. B. (1996). Query expansion using local and global document analysis. In *Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 4-11). ACM.
- [26] Xu, J., & Croft, W. B. (2000). Improving the effectiveness of information retrieval with local context analysis. *ACM Transactions on Information Systems (TOIS)*, 18(1), 79-112.
- [27] Segura, N. A., García-Barriocanal, E., & Prieto, M. (2011). An empirical analysis of ontology-based query expansion for learning resource searches using MERLOT and the Gene ontology. *Knowledge-Based Systems*, 24(1), 119-133.
- [28] Vallet, D., Fernández, M., & Castells, P. (2005). An ontology-based information retrieval model. In *The Semantic Web: Research and Applications* (pp. 455-470). Springer Berlin Heidelberg.
- [29] Karthikeyan, A. (2010). An novel approach using semantic information retrieval for Tamil documents. *International Journal of Engineering Science and Technology*, 2(9), 4424-4433.

- [30] Kara, S., Alan, Ö., Sabuncu, O., Akpınar, S., Cicekli, N. K., & Alpaslan, F. N. (2012). An ontology-based retrieval system using semantic indexing. *Information Systems*, 37(4), 294-305.
- [31] Zidi, A., & Abed, M. (2013). A generalized framework for ontology-based information retrieval: Application to a public-transportation system. In *Advanced Logistics and Transport (ICALT)*, 2013 International Conference on (pp. 165-169). IEEE.
- [32] De Ribaupierre, H., & Falquet, G. (2014). User-centric design and evaluation of a semantic annotation model for scientific documents. In *Proceedings of the 14th International Conference on Knowledge Technologies and Data-driven Business*. ACM.
- [33] Tulasi, R. L., Rao, S. M. & Gouda, R. G. (2013). Study of e-learning information retrieval model based on ontology. *International Journal of Computer Applications* 61(17):9-13.
- [34] Barreau, D., & Nardi, B. A. (1995). Finding and reminding: file organization from the desktop. *ACM SigChi Bulletin*, 27(3), 39-43.
- [35] Guangzuo, C., Fei, C., Hu, C., & Shufang, L. (2004). OntoEdu: a case study of ontology-based education grid system for e-learning. In *GCCCE2004 International conference*, Hong Kong.
- [36] Katifori, A., Vassilakis, C., Lepouras, G., & Torou, E. (2015). Effectiveness of visualization for information retrieval through ontologies with entity evolution: The impact of ontology modeling. *International Journal of Information Retrieval Research (IJIRR)*, 5(2), 66-91.
- [37] Katifori, A., Vassilakis, C., Lepouras, G., Torou, E., & Halatsis, C. (2014). Visualization method effectiveness in ontology-based information retrieval tasks involving entity evolution. In *Semantic and Social Media Adaptation and Personalization (SMAP)*, 9th International Workshop on (pp. 14-19). IEEE.
- [38] Wu, Z., & Palmer, M. (1994). Verbs semantics and lexical selection. In *Proceedings of the 32nd annual meeting on Association for Computational Linguistics* (pp. 133-138). Association for Computational Linguistics.
- [39] Pedersen, T. (2010). Information content measures of semantic similarity perform better without sense-tagged text. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics* (pp. 329-332). Association for Computational Linguistics.
- [40] Resnik, P. (1995). Using information content to evaluate semantic similarity in a taxonomy. *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 1*, 448--453. Morgan Kaufmann Publishers.
- [41] Lin, D. (1998). An information-theoretic definition of similarity. In *ICML 98*, 296-304.
- [42] Ukkonen, E. (1985). Algorithms for approximate string matching. *Information and Control*, 64(1), 100-118.
- [43] Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet Physics Doklady*, 10(8), 707-710.
- [44] Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology*, 26(3), 297-302.
- [45] Salton, G. and M. McGill (1983). *Introduction to modern information retrieval*. New York, NY: McGraw-Hill.
- [46] Lopez, V., Uren, V., Sabou, M., & Motta, E. (2011). Is question answering fit for the semantic web? A survey. *Semantic Web*, 2(2), 125-155.
- [47] Gao, G., Liu, Y. S., Wang, M., Gu, M., & Yong, J. H. (2015). A query expansion method for retrieving online BIM resources based on Industry Foundation Classes. *Automation in Construction*, 56, 14-25.
- [48] Lund, A.M. (2001). Measuring usability with the use questionnaire. *STC Usability SIG Newsletter*, 8(2).