DESIDOC Journal of Library & Information Technology, Vol. 36, No. 4, July 2016, pp. 199-204 DOI: 10.14429/djlit.36.4.9600

Personalised Information Recommender Using Framework for Ontology Alignment Among Digital Libraries

Anirban Chakrabarty* and Sudipta Roy**

*Assam University, Silchar, India
E-mail: chakrabarty.anirban@rediffmail.com
**Computer Science and Engineering Department, Triguna Sen School of Technology,
Assam Central University, Silchar, Assam
E-mail: sudipta.it@gmail.com

ABSTRACT

In the digital erantology is considered as one of the powerful tools for knowledge representation and efficient information retrieval. Ontology alignment is a process that discovers mapping between source and target ontologies, where each mapping is a relationship based on some similarity measure. This paper, has presented a new context aware alignment approach that needs little human intervention and it can map multiple ontologies to generate user interest dynamically. The objective is to design and develop an ontology alignment model that provides more benefits to its stakeholders in sharing resources and searching across digital libraries based on priorities of users. The experimental results evidently indicate significant improvement in search results when user profile and navigational pattern ontologies are aligned with digital library ontology.

Keywords: Digital library, ontology alignment, user interest, context similarity, information retrieval

1. INTRODUCTION

Digital library is a structured collection of digital content like text, video, or audio developed and made available on the web to meet the information need of end users. Digital libraries have great prospective as they can offer a wide range of benefits to researchers, academicians, institutions and learners worldwide[1]. IEEE Xplore Digital Library is an example of research library where journals, research articles and conference proceedings are organised for people to access remotely[2]. A personalised digital environment is need of the hour where users can organise the digital space, exchange information with each other, build communities and get recommendations, particularly when digital preservation of electronic resources is becoming more challenging[3].

Ontologies have come to the forefront of information science as a way to model a domain using a set of concepts and the relationships among them[4]. Ontology can be considered as a description of the elements it contains and when combined with other ontologies it can help to form a community of such elements[5]. Ontology alignment or ontology mapping is a useful tool for organising information semantically which can integrate disparate databases so that they can be accessed concurrently in a uniform manner. Several communities have taken mapping initiatives to address the issue of searching across multiple domains[6]. However, the effectiveness of search is hampered by the fact that individual library resources are usually not interconnected to the web and lacks the context.

The following motivations led to this research: Firstly, an ontology hierarchy within a digital library can enhance the access process by enabling ontological search using the context. Secondly, ontology alignment can be used to map multiple digital libraries simultaneously. This type of framework for a personalised digital library can be helpful in the following scenarios: Faculty members can navigate among the content of digital libraries in providing interactive e-learning sessions based on learner's preference or for researchers who want to search for publications under a specific domain[7]. Thirdly with ontology mapping, these personalised libraries can be built from previous resources and new information added after observing the usage pattern of library.

The challenging task is that one can rarely find two different ontologies communicating on the same hierarchy and for heterogeneous ontologies interpretation of elements becomes problematic. The major issue is to determine the user's profile, including navigational history and preferences. This information should help users to improve his searches on multiple digital libraries.[8]

2. RELATED WORK

Over the past years digital libraries (DL) have steadily enhanced their services to conveniently fulfill information need of diversified users, by offering them

Received 12 January 2016, revised on 30 May 2016, online published 28 July 2016

broader spectrum of services[9]. However, digital libraries are still experiencing difficulties in frequently updating cross-referenced areas, absence of shared approach on how to synchronise different DL profiles and structures. A digital library should not only adjust to the specific characteristics of each user profile, but also to the particular necessities and preferences of each user combining both library archives and profile level personalisation[10]. Luo[11], *et al*. introduced concepts related to user browsing history and proposed a hybrid user profile model and a personalised recommender system to utilise the semantic information between the items and user profile model to make recommendations.

Kruk[12], *et al*. highlighted that a semantic personalised digital library should enhance information extraction, facilitate query refinement and also provide recommendation services using community-aware ontologies.

Related study shows that ontologies share certain components that are important to their alignment like concepts, attributes, instances and relations. Typically, concepts represent the objects in a domain that have a variety of attributes. The relations between different concepts can be expressed using a concept hierarchy or through the properties that connects the concepts[13]. Khoo,[14] *et al.* developed a disease treatment ontology model that divides disease treatment information into five classes—disease, treatment, condition, effect and evidence. The usage of sub-classes, properties and instances of these main classes were illustrated in this paper. The widely used language to model ontologies is the Web Ontology Language (OWL), which is a form of RDF and is written using a subset of XML. The elements in OWL ontology are defined as RDF resources and are recognised by URI[15]. The use of ontology alignment in a digital library brings the possibility of predicting user requirements in advance and to offer personalised services ahead of actual need. Several ontology alignment algorithms have been developed over the past years were attempts have been made to develop matchers that perform large scale ontology alignments in the minimum amount of time[16]. Patkar[17] in his work indicates that Ontology is one of the latest tools for information retrieval from libraries in digital age. The paper discusses advances of information managing tools terminating in ontology and highlights the applications of ontology among the different fields.

The related study reveals that ontologies have been developed in different areas and to obtain good results, we need to find the relationships between terms in the different ontologies. Presently, there exist a number of systems that support users in aligning ontologies, but there is a need for mapping heterogeneous ontology that can provide personal recommendations to users based on interest areas. Most digital library retrieval services are oriented towards a generic user and provide minimum support for mapping between libraries. The work attempts to develop an ontology alignment technique to map different ontologies, perform advanced search, and provide recommendations and share information amongst heterogeneous library archives.

3. AIM

The major contributions in this paper are to:

- Propose a methodology for ontology mapping between concepts across different digital libraries by using this mapping to develop user's profile ontology. For this purpose web application is used to study the navigation and search patterns of selected users on digital libraries maintained by some Universities in India and which is publicly accessible.
- Suggest architecture for personalised digital library that allows individuals, institutions and associations to share, search digital content amongst libraries.
- Develop of a personalised information recommender for searching in a digital library environment based on user profiles and search patterns.

4. METHODOLOGY

Personalisation of services in digital libraries is need of the hour but there are some critical issues which need to be addressed - users differ in their personal likings and work modes, and libraries usually differ in the ontology as they are built for different purposes. The use of ontologies for describing the possible instances of use in a digital library provides the prospect of forecasting user requirements and to offer personalised services ahead of actual need[18]. The paper suggests a new architecture that addresses many of these problems, mentioned as personalised digital environment (PDE) is shown in Fig. 1. It is a system that provides recommendations using the knowledge extracted from searching and browsing profiles of users and also from knowledge integrated in ontology content of digital libraries. The solution has been evaluated on a web application where the goal was not to build an ontology for describing contents of a digital library but to map ontologies based on users search patterns.

5. ONTOLOGY ALIGNMENT

The objective of this section is to discuss the use of an application specific ontology to search for library resources annotated with other ontology. To overcome such diversities an additional level of functionality among ontologies have been introduced to deploy ontology alignment to define how concepts from different ontologies relate each other[19] is shown in Fig. 2. This mapping is modeled on the type of match used to interpret user preferences by ranking the similarity scores.

In this paper ontology is defined as a 5-tuple: $O=\{C,P,Rc,Rp,I\}$ where C and P are the set of concepts and properties respectively. R_c defines the hierarchical relationships (c_i,c_j) were c_i is the sub concept of c_j. Similarly, Rp defines the hierarchical relationships

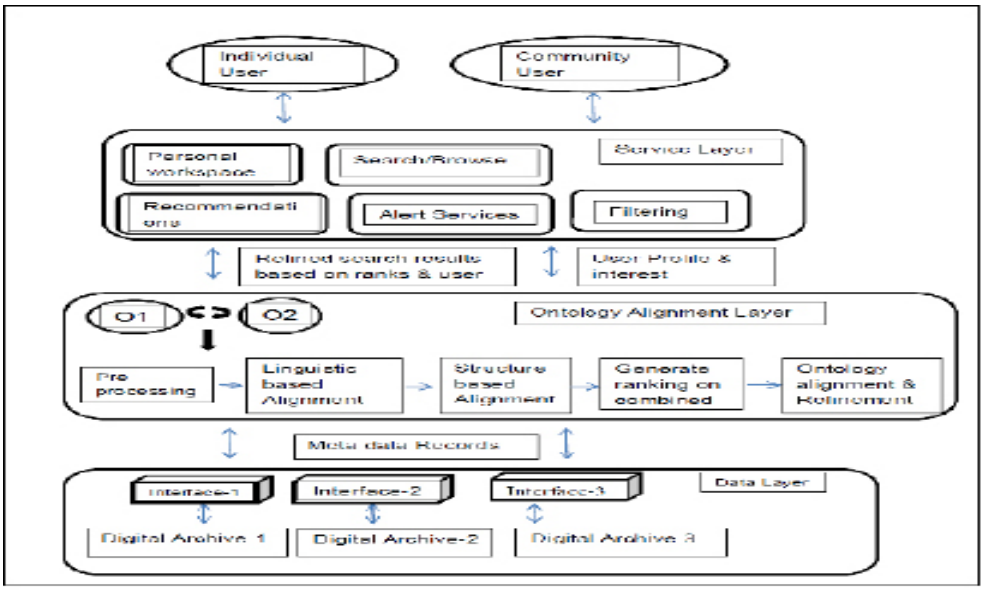

Figure 1. A layered architecture of a personalized digital library.

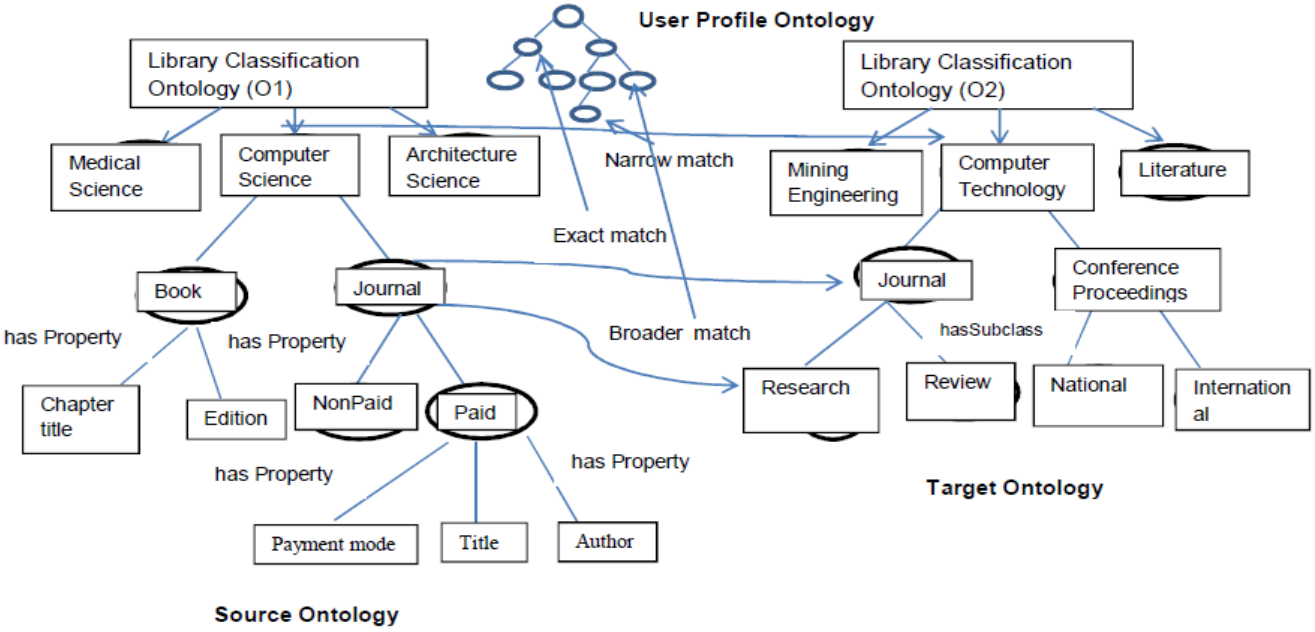

Figure 2. Ontology alignment for user profile generation.

between each property and its sub properties, I is a set of instances of concepts and properties. OWL uses owl: Class and rdfs: subClassOf to define the concepts and sub concepts, rdfs: Property and rdfs: subPropertyOf to define property and sub properties.

Ontology alignment with respect to the concepts accessed by user is defined as:

Align(O1,O2) = { userid, c_{i1}, rank c_{i1}, c_{i2}, rank c_{i2}, $relation_i$, $score_i$ } were userid is a unique system generated id for each user accessing the library ;c_{i1} € O1, c_{i2} € O2 ; rank of concept is defined as the ratio of count of access made by a user to the total access for the concept; relationi €{exactmatch, broad, narrow, conflict}; $score_i$ €[0..1] here score provides the alignment score of two concepts in the ontology. Each tuple in Align(O1,O2) represents that concept ci1 in O1 is mapped to concept c_{i2} in O2 with scorei and the alignment type relationi while the rank is used to provide recommendations to user.

Each Concept(c) in the ontology is represented as 5-tuple {Meta(c), Hier(c), Co-related(c), ,Inst(c),Count(c)} were meta(c) gives the set of words describing metadata of concept like labels, comments; Hier(c) is a vector that provides the sub and super concepts of c; Co-related(c) is a vector storing words that indicate synonym, hyponym of concept c; Count(c) gives no. of access made for the concept in the digital library (Table 1).

Table 1. Example of user profile generated from ontology alignment

User id	Concepts accessed in 01	Rank of concept in 01	Concepts accessed in 02	Rank of concept in 02	Relation type	Score
U1	Journal	0.6	Research article	0.4	broader	0.75
U2	Computer	0.7	Computer	0.6	Exact	1
U3	Medical	0.2	Literature	0.3	Conflict	0
U4	Payment mode	0.8	Payment process	0.7	Exact	1

6. PROPOSED ALGORITHMS

6.1 Ontology Alignment

The objective of aligning among different ontology-based digital libraries is to enable the use of ontology A to search resources annotated with concepts from another library with ontology B. The user ontology hierarchy generated consists of matched concepts from two ontologies along with their ranks.

Step-1: Contextual Match—Perform matching between concepts from source and target ontologies using Eqn.1 below. Initially the pairs are sorted into three buckets: above the upper threshold - similar, between upper and lower threshold – uncertain bucket and below the lower threshold-conflicts. The threshold is selected based on experimentation.

Similarity_Score(c1,c2)= |p(c1)∩p(c2)| / |p(c1) U p(c2) | ..... (1)

Here, p(c) denotes the property set of concept c and |p(c)| gives the no. of properties belonging to concept c. It can be interpreted that more the properties match in two concepts, higher the similarity score of two concepts.

Step-2: Subtree Matching—For concepts falling in the uncertain bucket do the following:

(a) For each pair of concept from two ontologies, compare their parent nodes to determine the parental similarity and use it to increment or decrement the original similarity score for concepts.

(b) For each pair of concept compare their child nodes to determine the child similarity and use it to update the original similarity score.

Step-3: Match refinement—Continue matching concepts to identify type of match and compute the rank of concepts accessed by user. Take average of the results from similarity_score calculated above from Eqn. 4 and rank of concepts to generate the user profile score.

6.2 Personal Recommendation based on Ontology Alignment

When a user navigates from one digital library to another, the library ontologies are aligned first to see the matches and based on ranking of concepts, users are provided with smart recommendations while inputting any character for search as shown in Fig. 3.

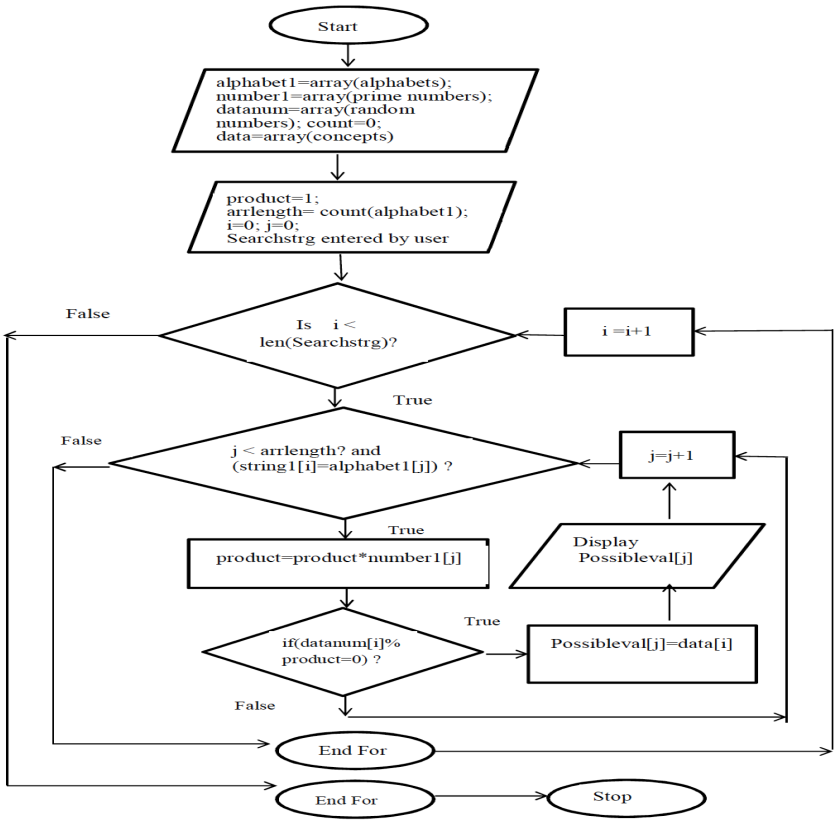

Figure 3. Flow chart for personalised recommendation.

7. DATA COLLECTION

7.1 Data Set

The experimental data is collected from publicly available digital libraries of Central Universities from North east India, namely, Assam Central University, Silchar and North Eastern Hill University, Shillong, India for study purpose. The experiment was done selecting five users randomly who had different interest areas. The software used for this work is JDK 1.7, HTML, Tomcat 8, and Protégé 4.1.

7.2 Experimental Metrics

The experimental evaluation was designed to address the following:

(a) To measure the effectiveness of personalised recommendation in ontology alignment F-measure values have been calculated and their variation with threshold changes has been studied. The F-measure[20] is a balanced mean of information retrieval between precision and recall metrics and has been defined as F= 2*P*R/(P+R

(b) Comparison of ontology alignment based search has been studied using one, two and three keywords for top-n documents (Figs 4-6).

7.3 Discussion of Experimental Results

The comparison of F-measure for search results of five users in different digital libraries is depicted in Fig 3. For all such users there has been a significant improvement in F-measure values in case of ontology Alignment model when compared with individual libraries. Since the value of similarity depends to some extent on the threshold considered in the algorithm, a variation of threshold with F-measure for different libraries and our alignment model has been studied for same set of five users. The highest F-measure value was reached for a threshold of 0.6 in the ontology model. Figures 5 and 6 provide the percentage of improvement in Precision and recall values for top-n search results using one, two and three keywords. Both results show that there is a

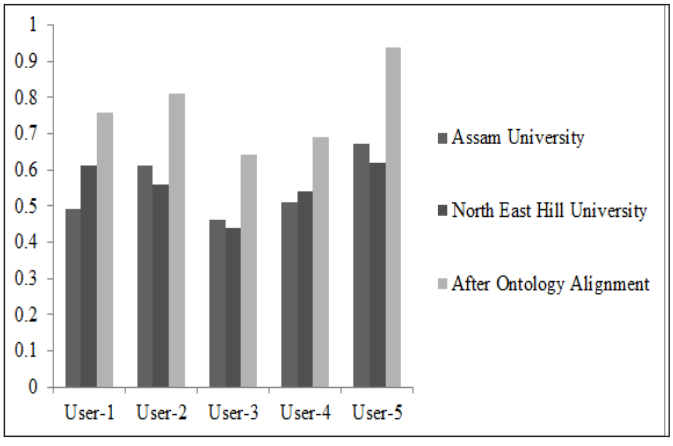

Figure 4. Comparison of F-measure from search results of five users.

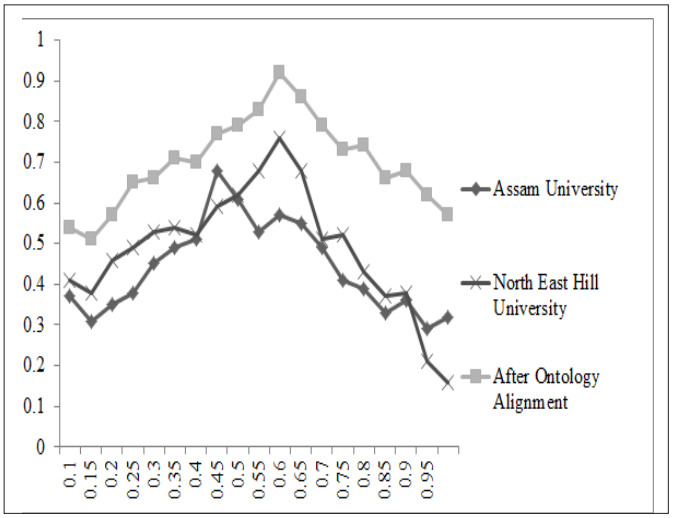

Figure 5. Comparison of F-measure with varying threshold.

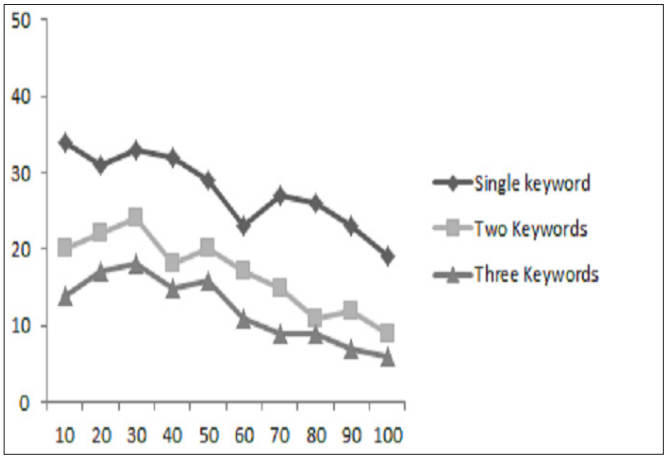

Figure 6. Percentage of improvement in precision using ontology alignment for top-n search results.

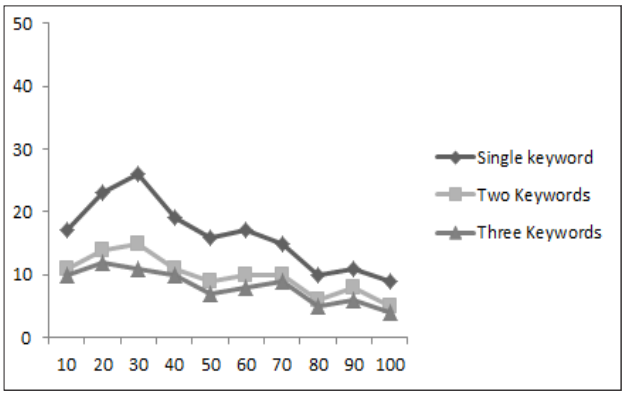

Figure 7. Percentage of improvement in recall using ontology alignment for top-n search results.

significant rise in precision and recall values using single keyword for ontology alignment approach when compared with individual library results. The improvement, however, decreases as the user provides more no. of keywords for search.

8. CONCLUSIONS

The work demonstrates the use of ontology alignment

to implement and transfer the concept of user profile and user navigational behaviour to other digital libraries and databases, so when a digital library user switches from one service to another, the user profile including preferences can be transferred from one database to another effectively. The work uses contextual matching, and sub tree matching to accurately identify matches and supplements them with rank of concepts accessed by user to generate profile dynamically. In future, the work will be extended to identify threats from malicious attacker on defence websites by aligning the ontology access patterns in weblogs.

REFERENCES

1. Witten, I.H.; Bainbridge,D. & Nichols, D.M. How to build a digital library. Morgan Kaufmann Publishers, Burlington, 2010.
2. http://portal.acm.org/dl.cfm (accessed on 19 December 2015)
3. Gaur, Ramesh C. & Tripathi, Manorama. Digital preservation of electronic resources. *DESIDOC J. of Lib. & Inf. Tech*., 2012, **32**(4), 293-301. doi: Http://dx.doi.org/10.14429/djlit.32.4.2522 (accessed on 15 December 2015).
4. Deshpande, Neela J. & Kumbhar, Rajendra. Construction and applications of ontology: Recent trends. *DESIDOC J. of Lib. & Inf. Tech.*, 2011, **31**(2). http://dx.doi.org/10.14429/djlit.31.2.859 (accessed on 12 December 2015).
5. Riano, David, *et al*. An ontology-based personalization of healthcare knowledge to support clinical decisions for chronically ill patients. *J. of Bioinf.*, 2012, **45**, 429-46, doi:10.1016/j.jbi.2011.12.008.
6. Juanzi, Li; Jie, Tang; Yi, Li & Qiong, LuoRi MOM: A dynamic multistrategy ontology alignment framework. *IEEE Trans. on Knowl. and Data Engg.*, 2009, **21**(8), 1218-32, doi: 10.1109/TKDE.2008.202.
7. Joshi, Nikita. Semantic web-driven e-learning system. *DESIDOC J. of Lib. & Inf. Tech*., 2011, **31**(4), 213-16. doi:http://publications.drdo.gov.in/ojs/index.php/djlit/article/view/1098. (accessed on 25 November 2015).
8. Towards personalization in digital libraries through ontologies. *Library Management*, 2005, **26**(4), 206-17, doi:10.1108/01435120510596062. (accessed on 17 November 2015).
9. Brahaj, A.; Razum, M. & Hoxha, J. Defining digital library research and advanced technology for digital libraries. *Lecture Notes in Comp. Sci.,* 2013, 8092, 23-8.
10. Southwick, S.B. A guide for transforming digital collections metadata into linked data using open source technologies. *J. of Lib. Metadata*, 2015, **15**(1), 1-35. http://dx.doi.org/10.1080/19386389.2015.1007009. (accessed on 21 December 2015).
11. Y. Luo, B.; Xu, H. Cai & F. Bu. A hybrid user profile model for personalized recommender system with linked open data. Enterprise Systems Conference (ES), 2014, Shanghai, 2014, pp. 243-48, doi: 10.1109/ES.2014.16.
12. Kruk, S.R. & Mc Daniel, B. Semantic digital libraries. Springer-Verlag, Berlin Heidelberg, 2009.
13. Volker, J.; Haase, P. & Hitzler, P. Learning expressive ontologies. *In* Ontology learning and population: Bridging the gap between text and knowledge. IOS Press, 2008, pp. 45-69.
14. Christopher,S.G, *et al*. Developing an ontology for encoding disease treatment information in medical abstracts. *DESIDOC J. of Lib. & Inf. Tech.*, 2011, **31**(2),103-115. http://publications.drdo.gov.in/ojs/index.php/djlit/article/view/862. (accessed on 21 October. 2015).
15. Ontology web language, 2008. http://www.w3.org/TR/owl-features/ (accessed on 18 December 2015).
16. Ashraf, J.; Khadeer Hussain, O. & Khadeer Hussain, F. A framework for measuring ontology usage on the web. Oxford University Press, 2012.
17. Patkar, Vivek. A passage to ontology tool for information organisation in the digital age. *DESIDOC J. of Lib. & Inf. Tech.,* 2011, **31**(2). http://publications.drdo.gov.in/ojs/index.php/djlit/article/view/861 (accessed on 25 November 2015).
18. Paulo, Cremonesi; Yehuda, Koren & Roberto, Turrin. Performance of recommender algorithms on top-n recommendation tasks. *In* Proceedings of Fourth ACM Conference on Recommender Systems, New York, USA, 2010, pp. 39-46.
19. Ehrig, M.; Staab, S. & Sure, Y. Bootstrapping ontology alignment methods with APFEL. *In* Proceedings of International Conference Semantic Web (ISWC '05), 2005, pp. 186-200.
20. Liu, Yuan-chao; Wu, Chong & Liu, Ming. Research of fast SOM clustering for text information. *J. of Expert Sys. with Appli.*, 2011, **38**, 9325-333.

Contributors

Mr Anirban Chakrabarty is a Research Scholar in Computer Science and Engineering Department, Assam Central University, Silchar, India. He is MTech (IT), and MCA and is pursuing his doctoral research in the area of Information Technology. He has contributed several research papers in National and International journals, conferences and has experience of teaching, industry and research. His research areas include: Text mining, natural language processing, data analysis, and algorithms.

Dr Sudipta Roy is Professor and Dean of Triguna Sen School of Technology, Assam Central University, India. He has recently been recognised as a Senior member, IEEE. He is PhD, MTech and has extensively contributed to reputed journals, seminars and conferences both at national and international levels. His areas of interests include: Image processing, networking, and data mining.