International Journal of Engineering & Technology, 7 (2.14) (2018) 80-84

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research Paper

A Java Based Web Service Model for Library Information Discovery

M Nordin A Rahman* and Usman Hafizu Uba

Faculty of Informatics & Computing, University of Sultan Zainal Abidin, Besut Campus, 22000 Besut, Terengganu
**Email: mohdnabd@unisza.edu.my*

Abstract

Library information discovery is a vital aspect of education development and management. Nowadays, many libraries offer the information regarding to facilities, resources and services to users through their website. The existing library information systems discover only their information and do not integrate the related information from other libraries. Therefore, web service technology has been developed, which is uses a standard interface to enable the consumer-to-business (C2B) or business-to-business (B2B) easily interact over the website. Currently, there is no standard model is developed to facilitate the users to discover and integrate the information of facilities, resources and services that published in the libraries website. This article discussed the development of a web service model that can discover and integrate the information in the various libraries website to the users by using Java Technology application program interface and tools provided by the integrated Web Services Stack. In the modelling process, the library websites are converted into Extensive Markup Language format (XML) for examining the Simple Object Access Protocol (SOAP) of each website. The converted XML files are stored as the application data source, which lies as the core of the web service model. By using an XML parser, the XML files are parsed into XMLDOM format for manipulating its contents. XML request object hypertext protocol is used to receive or reload the data from the client to the server. Finally, HTML and cascading style sheet (CSS) are used to reload the information to the client and display it. The results show that the proposed model is able to achieve over 90% of information discovery average precision performance regarding to user's query on facilities, resources and services provided in the library websites. The developed model can benefit library's users to extract the resources, facilities and services provided by the library intelligently.

Keywords: *Information discovery, web service, XML, trees representation, Java*

1. Introduction

Generally library provides physical and digital access to knowledge content including books, magazines, journals, online database and multimedia contents [1]. Some libraries also provide space for rental such as auditorium, seminar hall and meeting room. In the digital era, library presents their services through respective websites or web portals. Through this website, users can obtain information about the services, resources, physical space, infrastructure and knowledge content offered by the library. The number of libraries developed by the government or the private sector is increasing every year [1][2], and therefore public users have a variety of information to be discovered.

The general concept of library website information organization is based on several categories such as library profile, facilities, services, catalogues, content resources, library-guides and events. Basically, user who wants to get information about the services offered by the library needs to browse directly to the library's website. They need to manually analyze for each service displayed
on the website. This process is quite difficult and time consuming if many websites need to be visited and analyzed of their content. The complexity of analysis may increase if each website has different framework or architecture in representing the information. Therefore, a special system that can gather and analyze the contents of website is essential for searching process that involve various types of website frameworks. This kind of system also could encourage the librarians to increase and update the contents of library website constantly [3].

Usually web technology is used as a platform for storing and sharing data or document via the Internet. Nowadays, web is used to provide a service for data processing, analytics, organizing and visualization. In this case, URL is used as a mechanism to invoke services from in specific computer location via Internet [4][5]. This technology is called as web services and applied in many domains such weather forecasting, online ticketing, currency conversion, online dictionaries, bank account managing etc. Web services are application interface which allow application-to-application to communicate over the Internet [4][5]. This interface provides a level of abstraction and loose coupling between the cooperating software components. Web services can be perceived as a way to interpretation the functionalities of information system and make it available through a standard web technology. Based on the concept of Service-Oriented Architecture (SOA), web services enable interoperability and reusability of data from different application services that available on the Internet [4][6].

Currently, there is no standard web service model that facilitates the user to discover information or data of facilities, resources and

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

services that are published in the libraries website. With the features of Java web services technology such as Extensive Markup Language format (XML) and Simple Object Access Protocol (SOAP), this article will discuss the modelling of library website information discovery. The Java web service is self-describing modular application that can be published and invoked across the Internet [7].
The rest of the paper is organized as follows: Section 2 describes related work on web services technology and information discovery across website. Section 3 discusses the detailed of the proposed library information discovery using Java web service technology. Experimental results of the proposed model are presented in Section 4. Finally, conclusion is placed in Section 5.

2. Related Work

Within the context of information system, information services can be defined as an application that can help the users to support operation management and analyze the data for decision support making purposes [8][9]. It is also involved of data mining, extracting, discovery and ontology algorithms from web contents. The emerging of data management technology such as XML, WSDL, ontology web language, and open data initiative is contributed significantly in application that applying the concept of service oriented architecture.

2.1. Web Services Technology

Web service technology is self-contained modules that can be described, published and located over Internet [4][6]. A particular service can be described and used through an application interface which is developed based on open service architecture. The provided services can be in different types of functionalities, from simple requests such as converting a file format to another format, to very complicated business processes. Web services exposed the functionality of an application system, make it available through standard web technologies which could benefit the process of application integration, information gathering, and business-to-business enterprise system development. The principles of web services can be defined as follows [10]:

- Provides a higher level of abstraction that allows developers to integrate required functionalities to their applications easily and rapidly.
- Provides loosely coupled, independent functionalities.
- Supports both synchronous and asynchronous modes of communication.

Web services technology has four (4) fundamental components that provide universal descriptive, discovery and integration mechanism for communication protocol purposes [9]. These components are including XML, SOAP, WSDL and UDDI:

- XML – a uniform data representation and exchange mechanism from client to web server or from web server to client. XML's documents are made up of storage units called entities which contain either parsed or unparsed data and can be read by human and machine.
- SOAP – is the standard interface to enable applications on distributed computers across Internet, with different operating systems, platform etc. to interact each other and describe all message transactions.
- WSDL – a document with Meta language that written in XML format for machine understanding. It describes a web service, service availability, protocol types, invoke services and specify the location of the service and the operations (or methods) functioning.
- UDDI – a mechanism to register and locate web service based application. It is a directory service where transactions can be registered and search for services. UDDI also can be functioned as directory for storing information about a web services and a directory of web service interface described by WSDL.

2.2. Web Services Technology

Currently, a number of standard tools, search engine and hierarchical indices have been developed to help people in finding information [11]. These type of applications can help particular users to locate information in the hidden contents in a website, such as favourable information sources content, publish the queries, extract the results or data from website and information integrations from various websites. Generally, there are four (4) issues can be summarized in developing hidden information discovery in website content [11][12]:

- To define and locate structured data of user's interest.
- To persuade wrapper algorithms to extract relevant data object from websites.
- To annotate and label the fields of extracted data.
- To integrate various data object from multi-sources into user's format and schemas.

Information discovery from websites contents involves a process of navigating through a website collection, extracting, ranking and summarizing in order to create an analytics abstraction of the desire query. The common architecture of information discovery from websites is depicted in Fig. 1 [11][12][13]. The efficiency of discovery is mainly based on the data representation in the websites. Structured data with tag or label allows for more effective discovery, where the scraping engine can easily differentiate between information that is more or less significant.

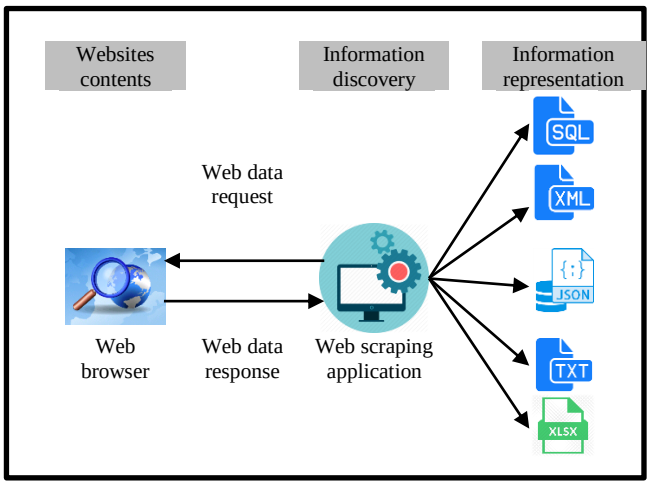


Fig. 1 : Basic website information discovery architecture

3. The Information Discovery Model

In the context of library websites information discovery, this model aims to discover information that categorize into three (3) different types: *resources (online database and e-book)*, *facilities* and *services*. These types of information are the common data that search by public users. The clustering process used in information discovery of library website content is divided into two basic steps:

- Feature extraction
- Similarity evaluation and clustering.

Before the contents of library websites can be categorized or compared, its need to be transformed into a representation that allow for effective processing. From various sources of library websites, the discovery engine has to extract essential information from all

websites pages and represent it in a form that will facilitate the matching process. Here, the proposed model used XML format to achieve this objective with trees structure of presentation.
The formal trees representation of XML contents can be defined as a 4-tuple of R = <N_R, E_R, L_R, α>, where:

- N_R = {n_1, n_2, n_3, …, n_n) is a set of tree nodes where n_1 represents the root node;
- E_R = {(n_i, n_j): n_i, $n_j \in N_R$} is a set of edges in trees structure, where n_i is the parent node of n_j;
- L_R is a set of labels referring to element and attribute names of XML document;
- α: $N_R \rightarrow L_R$ is a function for mapping each node into a label.

The example for general scheme of trees structure that can be constructed from a particular library website is illustrated in Fig 2.

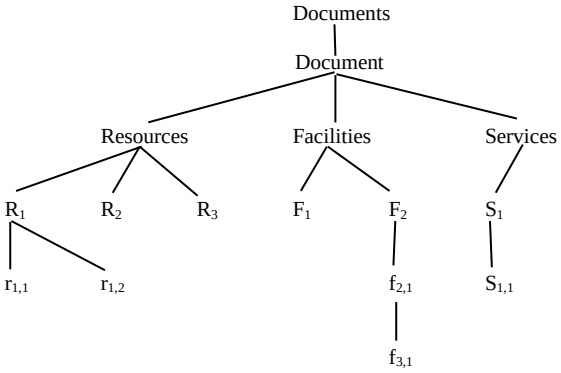


Fig. 2 : General scheme of trees representation

The next step consists of two basic tasks including find the most similar cluster to user query and perform local search within the most similar cluster. The most similar clusters are calculated based on the distance between user query and the centers of all clusters. Upon completed, the calculated clusters distances are sorted in ascending order. Sequentially, a search process is executed in the cluster elements that has the least distance between its center. Fig. 3 shows the search algorithm of the trees representation of XML scheme.

Input: Website dataset, D = {d_1, d_2, d_3, …,d_k},
Tree clusters, R = {r_1, r_2, r_3, …, r_L},
XQuery
Output: A set of matched documents

- Start
- clusters_distance[2][L] ←{∅}
- results ← {∅}
- for i=1 to L
  - clusters_distance[0][i] = c_i
  - clusters_distance[1][i] = distance between the requested document and the center of c_i
- sort ascending, clusters_distance
- for i=1 to L
  - for j=1 to CN_i (number of documents ∈ C_i)
    - compare between the requested document and D_j, where $D_j \in C_i$
    - if matched
      - results ← results ∪ D_i
  - if size of results > 0
    - list, results
  - else
    - break

Fig. 3 : Search algorithm for cluster similarity

4. Experimental Result

The prototype of model has been developed using Java programming language under JCreator environment. In the experiment, 21 library websites are selected for the purpose of discovering the contents of the relevant information. The performance of the proposed model is measured based on *precision*. Here, precision is the capability of the model in discovering only relevant information. Precision considers all discovered information where it is the part of the relevant discovered answers.
If A is relevant information discovered and B is number of irrelevant information retrieved, then the formula of precision is as follows:

$$\text{Precision} = \frac{A}{(A+B)} \quad (1)$$

As depicted in Table 2, the precision performance of *resources* information discovery in each website is above 0.80 and the average precision score is 0.92. Table 2 shows the precision performance of *facilities* information discovery of each website. As tabled, the model could discover 100 percent information regarding to facilities provided in library 1, 4, 5, 6, 7, 9, 11, 12, 13, 15, 17 and 19. Relatively, a low precision performance for library 21 could be caused by the complex feature extraction process of the website.
There are some terms used in the website that cannot be interpreted during the establishment of a tree representation scheme. The average precision score for this discovery context is 0.95. For *services* information discovery, the model could discover all related information in nine websites which are library 3, 4, 8, 12, 13, 16, 18, 19 and 20. The average precision for this type of discovery is 0.90.
Figure 4, Figure 5 and Figure 6 show the sample of reports for resources, facilities and services discovery respectively.

Table 1 : Resources information discovered

No.	Precision	No.	Precision
Lib 1	0.91	Lib 12	0.95
Lib 2	0.92	Lib 13	0.83
Lib 3	0.93	Lib 14	1.00
Lib 4	0.95	Lib 15	0.96
Lib 5	1.00	Lib 16	0.86
Lib 6	0.87	Lib 17	0.93
Lib 7	1.00	Lib 18	0.94
Lib 8	0.96	Lib 19	1.00
Lib 9	0.81	Lib 20	0.97
Lib 10	0.96	Lib 21	0.83
Lib 11	0.95		

Table 2 : Facilities information discovered

No.	Precision	No.	Precision
Lib 1	1.00	Lib 12	1.00
Lib 2	0.83	Lib 13	1.00
Lib 3	0.93	Lib 14	0.93
Lib 4	1.00	Lib 15	1.00
Lib 5	1.00	Lib 16	0.90
Lib 6	1.00	Lib 17	1.00
Lib 7	1.00	Lib 18	0.89
Lib 8	0.91	Lib 19	1.00
Lib 9	1.00	Lib 20	0.94
Lib 10	0.93	Lib 21	0.77
Lib 11	1.00		

Table 3 : Services information discovered

No.	Precision	No.	Precision
Lib 1	0.85	Lib 12	1.00
Lib 2	0.83	Lib 13	1.00
Lib 3	1.00	Lib 14	0.83
Lib 4	1.00	Lib 15	0.83
Lib 5	0.86	Lib 16	1.00
Lib 6	0.67	Lib 17	0.83
Lib 7	0.83	Lib 18	1.00
Lib 8	1.00	Lib 19	1.00
Lib 9	0.92	Lib 20	1.00
Lib 10	0.89	Lib 21	0.75
Lib 11	0.80		

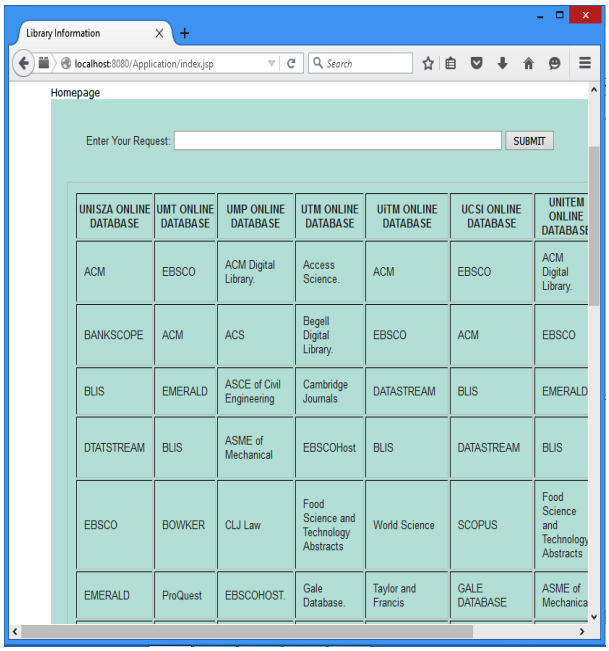


Fig. 4 : Resources information retrieved



Fig. 5. : Facilities information retrieved



Fig. 6 : Services information retrieved

5. Conclusions

This article presents the modelling of library website information discovery using Java web service technology. Based on Java web service standard feature such as XML and SOAP, the model is able to gather and organize the discovered information from a number of library websites. The type of information discovered are include knowledge contents (resources), facilities and services. The experiment results show that the proposed model could provide over 90% average precision for each discovery aspect in the library websites. For future work, the developed model will be enhanced by embedding the concept semantic and ontology based discovery.

Acknowledgments

Special thanks to the Ministry of Higher Education Malaysia for funding this research work under the scheme of FRGS 2015 (Code: RR143). Special thanks to Faculty of Informatics & Computing, University of Sultan Zainal Abidin for allowing the use of computer facilities.

References

[1] David, W. L. From stacks to web: the transformation of Academic library collections. Journal of Colleges and Research Libraries, (2013), 74: 159 – 176.

[2] Michael, M., Rozaklis, L., Hall, C. A survey of the use of ethnographic methods in the study of libraries and library users. Journal of Library and Information Science Research, (2012), 34: 82 – 91.

[3] Wang, Z. G. (2015). Analysis of digital library resource discovery service system. Proc. of the 5th International Conference on Education Management Information and Medicine, (2015), 719 – 723.

[4] Huang, C-Y. and Wu, C-H. A Web Service Protocol Realizing Interoperable Internet of Things Tasking Capability, Sensors, (2016), 16(9), 1 – 23.

[5] Vathsala, A. V. and Mohanty, H. A Survey on Checkpointing Web Services, Proc. of the 6th International Workshop on Principles of Engineering Service-Oriented and Cloud Systems, (2014), pp. 11 – 17.

[6] Lin, M. and Cheung, W. Automatic Tagging Web Services using Machine Learning Techniques, Proc. of the IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technologies, (2014), pp. 258 – 265.

[7] Taboada, G. L., Ramos, S., Expósito, R. R., Touriño, J., and Doallo, R. Java in the High Performance Computing arena: Research, practice and experience, Science of Computer

Programming, (2013), 78(5), 425 – 444.
[8] Tejeda-Lorente, Á., Porcel, C., Peis, E., Sanz, R. and Herrera-Viedma, E. A quality based recommender system to disseminate information in a university digital library. Information Sciences, (2014), 261: 52 – 69.
[9] Alvaro T-L., Porcel C., Peis E., Sanz, R. and Enrique, H-V. A quality based recommender system to disseminate information in a university. *Information Sciences.* (2014), 261: 52 – 69.
[10] Hamish, M. W., Weizhong, L., Mahmut, U., Silvano, S., Young, M. P., Nicola, B., Andrew, P. C. andRodrigo, L. Analysis Tools Web Services from the EMBL-EBI. Journal of Nucleic Acids Research, (2013), 41: 597 – 600.
[11] Higgins, S. Hilton, C. and Dafis, L. Archives Context and Discovery: rethinking arrangement and description for digital age. Proc. of the 2nd Annual Conference of the International Council on Archives, (2014), 1 – 18.
[12] Sadeh, T. From Search to Discovery. Proc. of the IFLA World Library and Information Congress, (2013), 1 – 17.
[13] Lagoze, C. and Singhal, A. Information Discovery: Needless and Haystack. IEEE Internet Computing, (2005), 16 – 18.