© 2014 IJEDR | Volume 2, Issue 4 | ISSN: 2321-9939

Mining Data Using Various Sequential Patterns Mining Algorithm in Semantic Web Environment

[1]Janki M. Padaliya, [2]Arjun V. Bala,
[1]PG Student, [2]Assistant Professor,
[1] Department of computer engineering,
[2]Darshan institute of Engineering and Technology, Rajkot, Gujarat, India.

Abstract: **The semantic Web is a second generation of the current World Wide Web (WWW). Where information is given well-defined meaning, and also changing Web contents into machine understandable form, it would promote quality and intelligence of the Web. To facilitate this process, RDF (Resource Description Frameworks) and OWL (Web Ontology Language) have been developed as standard formats for the sharing and integration of data and knowledge the latter in the form of rich conceptual schemas called ontologies. Data Mining is the nontrivial process of identifying valid, previously unknown, potentially useful patterns in data. Semantic Web Mining refers to the application of data mining techniques to extract knowledge from WWW or the area of data mining that refers to the use of algorithms for extracting patterns from resources distributed over in the web. In this paper we propose a new method for mining sets of sequential patterns for classification, where patterns are represented as SPARQL queries over RDFS. From the various Data mining algorithms we have apply here sequential pattern mining algorithms in semantic web to fetch sequential pattern from the semantic web data efficiently.**

Keyword**: Semantic Web, Data Mining, Sequential pattern Discovery, RDF, SPARQL, Ontology.**

I. Introduction

Semantic Web Mining is an integration of two important research areas: Semantic Web and Data Mining [1]. The existing Web (WWW) has a huge amount of information that is often unstructured and only human understandable. Web is rich with data; gathering and making sense of the information in the web is more difficult because the document of the Web is largely unorganized and unstructured. On the unstructured human readable web data, semantic web is used to effectively and efficiently creating a machine-understandable. In Semantic Web Mining, it refers to the application of data mining techniques to mine knowledge from World Wide Web [2] or the part of data mining that refers to the use of algorithms for extracting patterns from resources scattered over in the web.

II. PRELIMINARIES

This section gives some background about the two integrated research areas, the semantic web and sequential pattern mining.

A. Semantic Web

The **Semantic Web** is an evolving extension of the **World Wide Web** in which the semantics of information and services on the web is defined, making it possible for the web to understand and satisfy the requests of people and machines to use the Web content. It derives from W3C director Tim Berners-Lee vision of the Web as a universal medium for data, information and knowledge exchange. Semantic web information can support data integration, data discovery, navigation, and automation of tasks.

The Semantic Web has a layer structure that defines the levels of abstraction applied to the Web. At the lowest level is the familiar World Wide Web, then progressing to XML, RDF, Ontology, Logic, Proof and Trust [3, 4].In semantic web main tools that are currently being used in the Semantic Web are ontologies based on OWL (Web Ontology Language) and its associated reasoners. The Semantic Web Layered Architecture will describe in Figure 1.

Techniques of Semantic Web Representation

There are many techniques and models available used to represent and express the semantic of data such as the standard techniques recommended by W3C named Extensible Markup Language(XML), Resource Description Framework(RDF), and Web Ontology Language(OWL)[5]. They are in brief explained below.

1. **Extensible Markup Language (XML)**
   The XML technique has been established as a generic technique to store, organize, and retrieve data on or from the web. By enabling users to create their own tags and it allows them to define their content easily .So, the data and its semantic relationships can be represented [7,8]. XML Namespace in semantic web is used to avoid conflict data or names. XML layer aim to provide the basic syntax and structure of the data on the web.

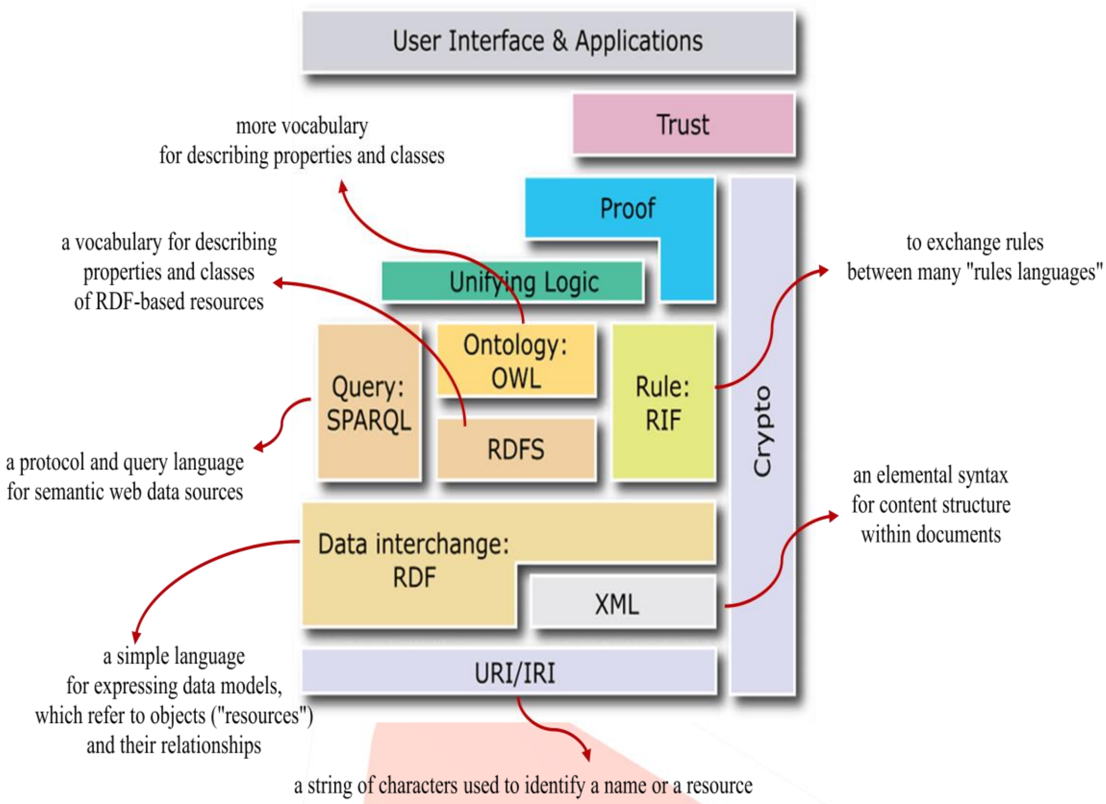

Fig.1 Layers of the Semantic web

2. **Resource Description Framework (RDF)**
   RDF is a framework for semantic web based on XML it's a common language that enables the facility to store resource's information that is available in World Wide Web using their own domain vocabularies [5, 6]. There are three types of elements contented in the RDF: resources (entities identified by Uniform Resource Identifiers URIs), literals (atomics values like strings and numbers), and properties (binary relationships identified by URIs) [7]. This is a very effective method to represent any kind of information that could be defined on the web [5].
3. **Web Ontology Language (OWL)**
   The Web Ontology Language OWL is a semantic markup language for publishing and sharing ontologies on the World Wide Web. OWL is developed as a vocabulary extension of RDF. The Web Ontology Language (OWL) is used a more difficult language with better machine-interpret ability than RDF. OWL has main two type properties i.e. Object properties and Data type properties.
4. **Query Language: SPARQL**
   Finally, a crucial puzzle piece which pushed the recent wide uptake of Semantic Web technologies at large was the availability of a standard query language for RDF, namely SPARQL [16], which plays the same role for the Semantic Web as SQL does for relational data. SPARQL's syntax is roughly inspired by Turtle and SQL, providing basic means to query RDF such as unions of conjunctive queries, value filtering, optional query parts, as well as slicing and sorting results. The recently rechartered SPARQL 1.1 W3C working group1 aims at extending the original SPARQL language by commonly requested features such as aggregates, sub-queries, negation, and path expression.

B. Sequential Pattern Mining

Sequential pattern mining is a popular data mining task with wide applications, which extract frequent subsequences from a sequence database. A subsequence is called sequential pattern or frequent sequence if it frequently appears in a sequence database, and its frequency is no less than a user-specified minimum support threshold minsup [9].. Sequential pattern mining is essential to a wide range of applications such as the analysis of web click-streams, program executions, medical data, biological data and e-learning data [10].

Definition 1 (**sequential pattern**). A **sequential pattern** is a sequence. A **sequence SA** = X1, X2, ... Xk, where X1, X2... Xk are itemsets **is said to occur in another sequence SB** = Y1, Y2, ... Ym, where Y1, Y2... Ym are itemsets, if and only if there exists integers 1 <= i1 < i2... < ik <= m such that X1 ⊆ Yi1, X2 ⊆ Yi2, ... Xk ⊆ Yik.

Definition 2 (**support of a sequential pattern**). The **support of a sequential pattern** is the number of sequences where the pattern occurs divided by the total number of sequences in the database.

Definition 3 (**frequent sequential pattern**). A **frequent sequential pattern** is a sequential pattern having a support no less than the minsup parameter provided by the user.

Definition 4 (**closed sequential pattern**). A **closed sequential pattern** is a sequential pattern such that it is not strictly included in another pattern having the same support.

Definition 5 (**maximal sequential pattern**). A **maximal sequential pattern** is a sequential pattern such that it is not strictly included in another closed sequential pattern.

There are main two Methods for sequential pattern mining 1) Apriori-based Approaches 2).Pattern-Growth-based Approaches. Several algorithms have been proposed for sequential pattern mining such as GSP[11], SPADE[12] are Apriori based and PrefixSpan[13], MaxSP[14] ,VMSP[15] are pattern growth based algorithm. For example how the VMSP work is describe below.

- ***VMSP (Vertical mining of Maximal Sequential Patterns) Algorithm***

VMSP is an algorithm for discovering maximal sequential patterns in sequence databases. VMSP is developed for the general case of a sequence database rather than strings and it can capture the complete set of maximal sequential patterns with a single database scan. We performed an experimental study with have real-life datasets to compare the performance of VMSP with MaxSP, the state-of-the-art algorithm for maximal sequential pattern mining. Results show that VMSP is up to two orders of magnitude faster than MaxSP, and perform well on dense datasets.

Pesudocode:

Input:

1. Takes as input a sequence database SDB, a user-specified threshold named minsup (a value in [0, 1] representing a percentage), Maximum pattern length.

2. First scans the input database SDB once to construct the vertical representation of the database V (SDB) and the set of frequent items F1

3. Each frequent item s ϵ F1, the procedure calls the SEARCH procedure [15].

Output:

Discovering maximal sequential patterns

III. PROPOSED APPROACH

In this approach sequential pattern mining algorithm is used. This process work in two step.

1. First OWL file is convert in specific spmf text file format
2. Then respective sequential pattern mining algorithm (i.e. VMSP) is applied in spmf framework to fetch efficient sequential pattern.

Pseudocode:

Input: OWL file & other parameters which required by respective algorithm (i.e. VMSP), SPARQL query

Require: To generate an instance transaction T, the ontology O, the instance store IS, and the composition triples ISQ from given OWL file.

Output: Sequential pattern according to respective algorithm (i.e. VMSP)

Algorithm:

1) Load OWL file from specified location(web/local)
2) Apply SPARQL query on OWL
3) Extract Instances from result of SPARQL query
4) POST process the Instance Store(IS)(i.e. convert to sequential transaction items)
5) Save IS to Text file
6) Apply Respective algorithm (i.e. VMSP) to file
7) Display result

IV. METHODOLOGY

In this section we present a method which depicts a schematic overview of the whole process. The user specifies a mining pattern following an extended SPARQL syntax. Then, the transaction extractor is able to identify and construct transactions according to the mining pattern previously specified. Finally, the set of transactions obtained and user input parameters (SP) are processed by a traditional pattern mining algorithms, which finds sequential pattern of the form specified in the mining pattern with support greater than user's specified ones.

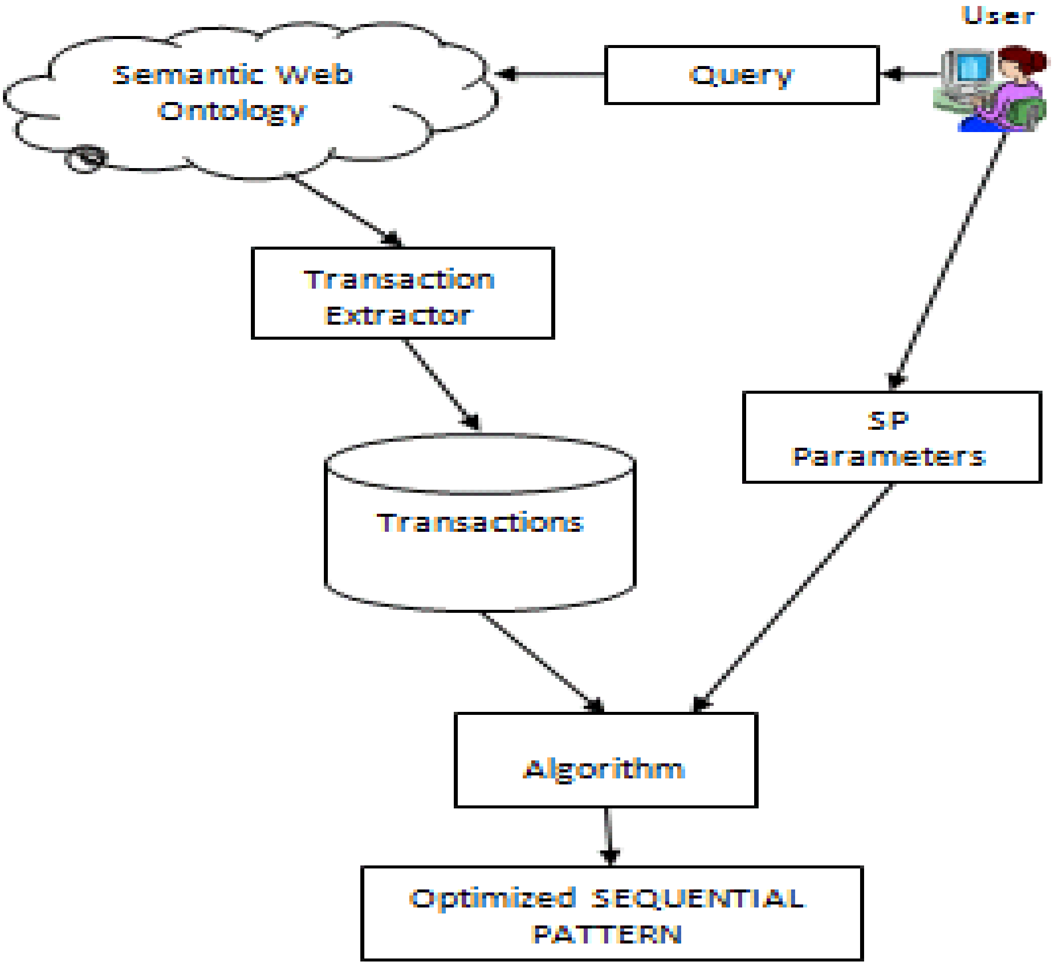

Fig.2 Architecture of Proposed System

V. CONCLUSION

In this paper, we have presented a novel method for mining sequential pattern from heterogeneous semantic data repositories expressed in RDF and OWL. We apply sequential pattern mining algorithms in semantic web to fetch user specific sequential patterns efficiently.

VI. FUTURE SCOPE

The method proposed in this paper could be applied in different areas, like, where the mining tasks are transaction oriented. Also, in future work, the knowledge encoded in the ontology would be used to filter and prune discovered pattern. It would also be helpful to express the user goals.

VII. ACKNOWLEDGMENT

Special thanks to Darshan institute of engineering and technology to support this research. I am grateful to prof. Arjun V. Bala for helpful discussion and comments.

REFERENCES

[1] O. Mustapaşa, A. Karahoca, D. Karahoca and H. Uzun- boylu, "Hello World, Web Mining for E-Learning," *Pro- cedia Computer Science*, Vol. 3, No. 2, 2011, pp. 1381- 1387. doi:10.1016/j.procs.2011.01.019

[2] Bellaachia, A., Vommina, E., and Berrada, B. (2006). Minel: A Framework for Mining E-learning Logs, Proceedings of the 5th IASTED International Conference on Web-based Education, Puerto Vallarta, Mexico, 259-263.

[3] 2. Berendt, B., Hotho, A., Mladenic, D., van Someren, M., Spiliopoulou, M., Stumme, G.: A Roadmap for Web Mining: From Web to Semantic Web. Web Mining: From Web to Semantic Web Volume 3209/2004 (2004) 1–22

[4] 1. Stumme, G., Hotho, A., Berendt, B.: Semantic Web Mining: State of the art and future directions. Web Semantics: Science, Services and Agents on the World Wide Web 4(2) (2006) 124 – 143 Semantic Grid –The Convergence of Technologies.

[5] D. Jeon and W. Kim, "Development of Semantic Deci- sion Tree," *Proceedings of the 3rd International Confer- ence on Data Mining and Intelligent Information Tech- nology Applications*, Macau, 24-26 October 2011, pp. 28- 34.

[6] V. Sugumaran and J. A. Gulla, "Applied Semantic Web Technologies," Taylor & Francis Group, Boca Raton, 2012.

[7] V. Nebot and R. Berlanga, "Finding Association Rules in Semantic Web Data," *Knowledge-Based Systems*, Vol. 25, No. 1, 2012, pp. 51-62. doi:10.1016/j.knosys.2011.05.009

[8] A. Jain, I. Khan and B. Verma, "Secure and Intelligent Decision Making in Semantic Web Mining," *Interna-tional Journal of Computer Applications*, Vol. 15, No. 7, 2011, pp. 14-18. doi:10.5120/1962-2625

[9] 1. Agrawal, R., Ramakrishnan, S.: Mining sequential patterns. In: Proc. 11th Intern. Conf. Data Engineering, pp. 3{14. IEEE (1995)

[10] Mabroukeh, N.R., Ezeife, C.I.: A taxonomy of sequential pattern mining algorithms, ACM Computing Surveys 43(1), 1{41 (2010)

[11] Shrikant, , R., Agrawal, R.: Mining Sequential Patterns: Generalizations and Performance Improvements. In: Apers, P.M.G., Bouzeghoub, M., Gardarin, G. (eds.) EDBT 1996. LNCS, vol. 1057, pp. 3–17. Springer, Heidelberg (1996)

[12] Zaki, M.J.: SPADE: An efficient algorithm for mining frequent sequences. Machine Learning 42(1), 31{60 (2001)

[13] Pei, J., Han, J., Mortazavi-Asl, B., Wang, J., Pinto, H., Chen, Q., Dayal, U., Hsu, M.: Mining sequential patterns by pattern-growth: the Pre_xSpan approach. IEEE Trans. Known. Data Engin. 16(11), 1424{1440 (2004)

[14] Philippe Fournier-Viger, Cheng-Wei Wu, Antonio Gomariz, Vincent S : VMSP: Efficient Vertical Mining of Maximal Sequential Patterns(2014)

[15] Philippe Fournier-Viger, Cheng-Wei Wu and Vincent S. Tseng : Mining Maximal Sequential Patterns without Candidate Maintenance(2013)

[16] C. Kiefer, A. Bernstein, A. Locher, Adding data mining support to SPARQL via statistical relational learning methods, in: S. Bechhofer, M. Hauswirth, J.Hoffmann, M. Koubarakis (Eds.), ESWC, Lecture Notes in Computer Science,vol. 5021, Springer, 2008, pp. 478–492.