# SEMANTIC WEB PERSONALIZATION FRAMEWORK NON-SEQUENTIAL PATTERN MINING

GOPAL SINGHAL, Dr. M. Deepamalar
Research Scholar, SSSUTMS, SEHORE, MP
Research Guide, SSSUTMS, SEHORE, MP

**ABSTRACT**

Semantic Web Mining can be utilized to find fascinating client route designs, which at that point can be connected to genuine issues, for example, website change, extra theme/item suggestions, client conduct's examination and so forth. This work centers the utilization of semantic web mining in web personalization. Web personalization framework not just gives client an arrangement of customized pages yet in addition gives client a rundown of spaces the client might be keen on. In this manner client can change to various interests when surfing on the web for data. Other than this, web personalization has expanded the precision of proposals essentially of the non-sequential pattern mining.

**Keyword:** SPM, IAP, NSPM

## INTRODUCTION

The Sequential Pattern Mining (SPM) algorithm SPMining is depicted in Algorithm 1. In this algorithm, we apply the pruning scheme for the purpose of eliminating non-closed patterns during the process of sequential patterns discovery. The key feature behind this recursive algorithm is represented in the first line of the algorithm, which describes this pruning procedure. In this algorithm, all (n-1)Terms of length patterns are diagnosed to determine whether or not they are closed patterns after all nTerms of length patterns are generated from the previous recursion. The algorithm repeats itself recursively until there is no more pattern discovered. As a result, the output of algorithm SPMining is a set of closed sequential patterns with relative supports greater than or equal to a specified minimum support.

The algorithm SPMining is developed for the purpose of mining all frequent sequential patterns from documents. In addition to sequential patterns, non-sequential patterns mining (NSPM) from a set of textual documents is another application of the data mining mechanism. From the data mining point of view, non-sequential patterns can be treated as frequent itemsets extracted from a transactional database. Frequent

itemset mining is one of the most essential issues in many data mining applications. The pseudo code of NSPMining is listed in Algorithm 2.

**Method:**
1: $SP \leftarrow SP - \{Pa \in SP \mid \exists Pb \in PL$ Such That $len(Pa) = len(Pb) - 1$ And $Pa < Pb$ And $supp_a(Pa) = supp_a(Pb)\}$
//Pattern Pruning
2: $SP \leftarrow SP \cup$ PL //*nTerms pattern set*
3: $PL' \leftarrow \phi$
4: **For Each** *pattern p* In *PL* **Do Begin**
5: generating *p-projected* database *PD*
6: **For Each** frequent *term t* In *PD* **Do Begin**
7: $P' = p \oplus t$ //sequence extension
8: **If** $supp_r(P') \geq min\_sup$ **Then**
9: $PL' \leftarrow PL' \cup P'$
10: **End If**
11: **End For**
12: **End For**
13: **If** $|PL'| = 0$ **Then**
14: Return //no more pattern
15: **Else**
16: Call SPMining($PL'$, *min_sup*)
17: **End If**
18: Output *SP*

**Algorithm 2.** NSPMining(*NP*, *FT*, *min_sup*)
**Input:** a list of *nTerms* frequent non-sequential patterns, *NP*; a list of *1Term* frequent patterns, *FT*; minimum support, *min sup*.
**Output:** a set of frequent non-sequential patterns, *FP*.
**Method:**
1: $FP \leftarrow FP \cup$ NP //*nTerms* non-sequential patterns
2: $NP' \leftarrow \phi$
3: **For Each** pattern *p* in *NP* **Do Begin**
4: **For Each** frequent term *t* in *FT* **Do Begin**
5: $P' = p \cup \{t\}$ //pattern growing
6: **If** $supp_r(P') \geq min\_sup$ **Then**
7: $NP' \leftarrow NP' \cup P'$
8: **End If**
9: **End For**
10: **End For**
11: **If** $|NP'| = 0$ **Then**
12: Return //no more pattern
13: **Else**
14: Call NSPMining($NP'$, *FT*, *min_sup*)
15: **End If**
16: Output *FP*

The precision is the fraction of retrieved documents that are relevant to the topic, and the recall is the fraction of relevant documents that have been retrieved. For a binary classification problem the judgment can be defined within a contingency table as depicted in Table I. According to the definition in this table, the measures of Precision and Recall are denoted as TP/(TP+FP) and TP/(TP+FN) respectively, where TP (True

Positives) is the number of documents the system correctly identifies as positives; FP (False Positives) is the number of documents the system falsely identifies as positives; FN (False Negatives) is the number of relevant documents the system fails to identify. The precision of top-K returned documents refers to the relative value of relevant documents in the first K returned documents. The value of K we use in the experiments is 20, denoted as "t20". Breakeven point (b/e) is used to provide another measurement for performance evaluation. It indicates the point where the value of precision equals to the value of recall for a topic. Both the b/e and F1-measure are the single-valued measures in that they only use a figure to reflect the performance over all the documents. However, we need more figures to evaluate the system as a whole.

Therefore, another measure, Interpolated Average Precision (IAP) is introduced. This measure is used to compare the performance of different systems by averaging precisions at 11 standard recall levels (i.e., recall=0.0, 0.1, ..., 1.0). The 11-points measure is used in our comparison tables indicating the first value of 11 points where recall equals to 116 Experiments and Results zero. Moreover, Mean Average Precision (MAP) is used in our evaluation which is calculated by measuring precision at each relevance document first, and averaging precisions over all topics.

The Fig.1 shows the comparison of all data mining methods in precision at standard recall points on the first 50 RCV1 topics. It reveals that all data mining methods have the similar performance. The SCPM and NSCPM, which adopt closed patterns, perform a little better than others around the low recall situation. PTM outperforms data mining methods due to the use of pattern pruning and deploying techniques. The comparison of number of patterns and runtime using data mining methods is indicated in Fig. 2. The NSPM method needs a lot of runtime since the huge amount of patterns generated during the projected stage. The closed pattern methods, such as SCPM and NSCPM, produce less number of patterns due to the adoption of pruning process during the pattern discovery stage. PTM also outperforms data mining methods since the less number of patterns is used.

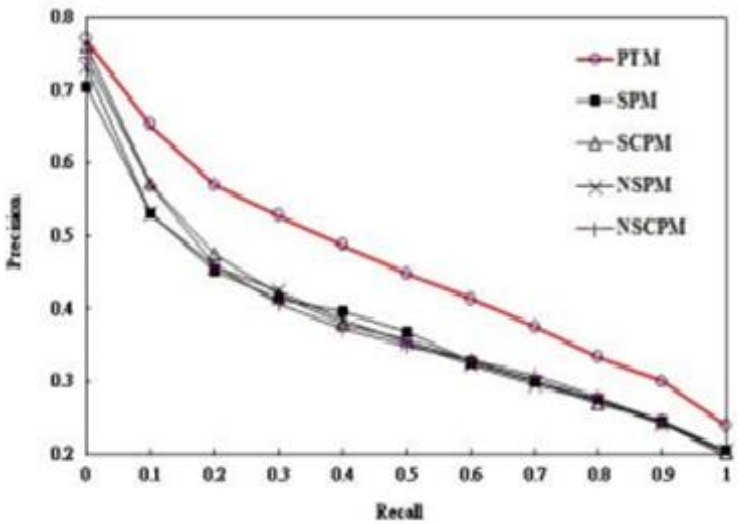

**Figure 1: comparison of all the methods in precision at standard recall points on the first 50 RCV1 topics**

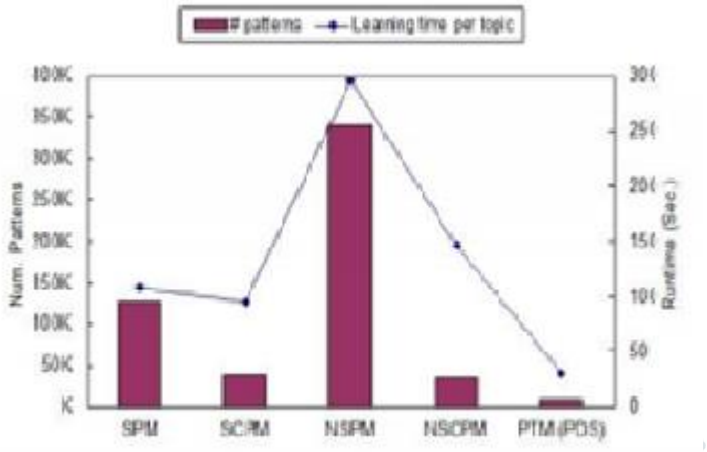


**Figure 2: comparison of number of patterns and runtime using data mining methods**

In general, a significant amount of patterns can be retrieved by using the data mining techniques to extract information from Web data. Many data mining techniques have been proposed in the last decade. These techniques include association rule mining, frequent itemset mining, sequential pattern mining, maximum pattern mining, and closed pattern mining. However, how to effectively use these discovered patterns is still an unsolved problem. Another typical issue is that only the statistic properties (such as support and confidence) are used while evaluating the effectiveness of patterns.

A comprehensive comparison of data mining methods applied for Web mining task is performed in this study. The experimental results show that closed pattern methods, such as SCPM and NSCPM, have better performance due to the use of pruning mechanism in the pattern discovery stage. It also proves that the less number of patterns used by these two methods does not affect their effectiveness.

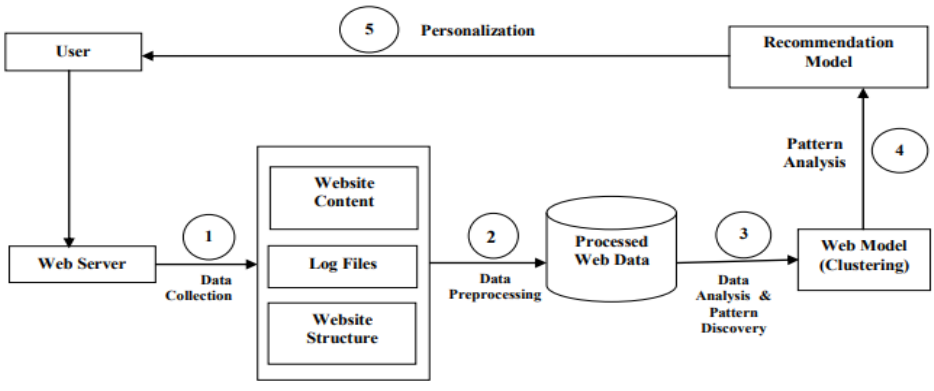


Figure 5. Web Personalization System Architecture

**Figure 3: web personalization system framework**

The response of a cell or reference to ionizing reference has been shown to be mediated by direct damage to cellular components like semantic web, lipids, proteins and small molecules as well as indirect damage mediated by weblysis of water . It is now apparent that the target for the web effects of ionizing reference is not solely the irradiated cells but also includes the surrounding cells and references. Reference-induced bystander effect is defined by the observance of web effects of reference in cells that were not themselves in the field of reference. Estimates of the various damage types at sub-cellular, cellular and supra-cellular level induced by low doses, where no new data are available, are generally based on linear back-extrapolations from data at higher doses. However, a large amount of data produced during the last decade seem to indicate higher damage yields than expected, possibly due to the so-called “bystander effect”.

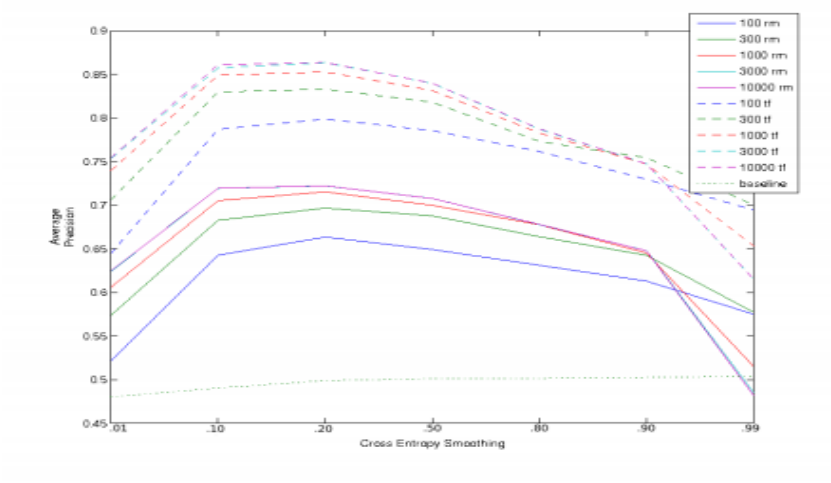


**Figure 4: Average Precision Scores for Language Model Parameters: Relevance Feedback from Hypertext to Semantic Web**

One of the most intriguing aspects of these experiments lies in that bystander damage is observed even at very low doses, down to the mGy level, and it does not significantly increase with dose. No clear-cut conclusions can be drawn on the mechanisms underlying bystander effect explanation. Different pathways may be involved, depending on the specific conditions. In any case the various forms of cellular letter seem to play a key role, since damage in non-directly hit cells may represent a response to signals coming from cells that were directly hit by reference. If these findings are confirmed, webbiology experiments may need to be re-interpreted and the models of low-dose reference action may need to be revised.

Dynamic nature of web ad hoc networks poses a number of problems in designing proper routing protocols to forward data packets from source to destination (Royer and Toh 1999). Proactive protocols suffer from large consumption of bandwidth by exchange of routing information and reactive protocols suffer from excessive flooding and route discovery latency. Mobile agents similar to ants can be used to collect all

topology-related information from each node in the network and distribute them periodically as updates to other nodes (Correia et al 2009). Mobile agents or messengers or ants that hop around in the network, collect information from these nodes, meet other agents in their journey, interact with both to collect updates of parts of the network that they have not visited or have visited a long time back, and handover these collected data sets to newly visited nodes and agents. Chaudhary et al (2004) proposed ant based routing algorithms for mobile ad hoc networks. They divide the work into two parts. First part is an agent-based framework in which agents move from one node to another in the network and collect topology information and distribute them into the information cache of all other nodes. Second part makes use of this information to establish and maintain communication link between two nodes.

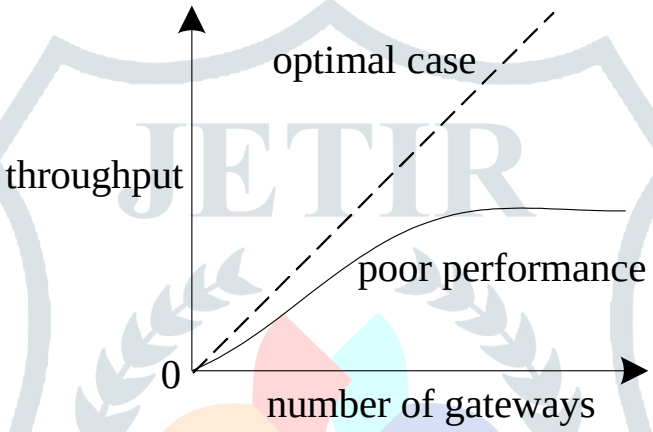


**Figure 5: Throughput should increase with number of gateways.**

Furthermore, a significant role of bystander effects in vivo would imply the second look of models of damage at reference and organ level. A typical example is provided by low-dose semantic risk, which up to now has generally been estimated by assuming a linear, no-threshold dose-response.

Routing consists of establishing a so - called virtual channel between source and destination at the beginning of the connection and guiding packets (called cells) over it. There are two main routing protocols in ATM networks, Private Network - Node Interface (PNNI) and Interim Inter switch Signaling Protocol (IISP). IISP provides a static routing solution and is based on manually configured routing tables. Smaller ATM based systems are simple to deploy and they can also use proprietary implementations of PNNI. However, for large networks it is prone to errors and time - consuming to configure. In contrast to IISP, PNNI supports QoS and crank back.

It provides two significant services: network topology call establishment and discovery. It is a hierarchical, dynamic link - state, source routing protocol. So upon receiving the call request a source router references the PNNI routing table to determine a path to the intended destination that is capable to support the QoS requirements specified by the caller. The connection message is then forwarded to the dest ination along the

potential path and if sufficient resources are available on every intermediate node, the transmission can start. Otherwise the crank back or back off occurs and a new path has to be computed. If it does not satisfy the request requirements, the connection is refused. Numerous telecom companies have deployed ATM networks. However, it has failed to gain wide use as a LAN technology and its great complexity has hampered its full deployment as the single integrating network technology in the w ay that its inventors originally intended.

The DV strategy was implemented in many routing protocols. One of the most common examples is the Routing Informatio n Protocol (RIP) . It was initially deployed with Xerox Network Services by Xerox. RIP is most commonly used as a routing protocol on intranets.

The main advantage of the DV protocol is its simplicity and small overhead. This permits distributed and asynchronous operation and requires only locally available costs to compute and select routes. However, the Bellman - Ford algorithm causes problems when the network architecture is unstable. It suffers then from a count - to - infinity problem and does not prevent routing loops. Thus another class of protocols (called link - state) for data networks was developed. The LS algorithm was also adapted by the Internet Engineering Task Force in the Open Shortest Path First (OSPF) protocol for use in the Internet. Protocols, IS - IS and OSPF, are broadly utilized for intra autonomous sy stem routing. In a LS protocol a router periodically broadcasts its local view of the network as link states, in terms of the properties of the links connecting it to neighboring nodes.

The main finding of these studies, lie in the fact that such treatment can reduce survival of unexposed cells. Therefore the onus lies on non gap junction mediated cell message. Evidence for some forms of bystander effect has been suggested also by conformist reference with low doses. One of the first studies in this field has showed a significant, unexpected increase in the frequency of sister chromatid exchanges (SCEs) after exposure of Chinese hamster ovary (CHO) cells to very low doses of alpha particles, from 0.03 to 0.25 cGy . While about 1% or less of the cells were actually traversed by a particle track, 30–45% of the individual cells showed increased levels of SCE.

## CONCLUSION

WI has been perceived as a standout amongst the most imperative and additionally the quickest developing IT look into fields in the time of the World Wide Web, information Web, matrix figuring, intelligent specialist innovation, and universal social processing. WI advancements will keep on producing the new apparatuses and the foundation segments important for making intelligent venture entrances that can serve users carefully. To meet the solid requests for support and the developing interests in WI, the Web Intelligence Consortium (WIC) was shaped in spring 2002. The WIC (http://wiconsortium.org/) is a worldwide non-benefit association devoted to propelling world-wide logical research and modern advancement in the field of WI. It advances

coordinated efforts among worldwide WI explore focuses and authoritative individuals, innovation features at WI related meetings and workshops, WIC official book and diary productions, WIC pamphlets, and WIC official arrivals of new mechanical arrangements and measures. Notwithstanding significant WI related meetings/workshops, for example, IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology, and various extraordinary issues in global diaries/magazines, for example, IEEE Computer, a WI-centered logical diary, Web Intelligence and Agent Systems: An International Journal (allude to the WIC landing page), has been giving a standard universal discussion to spreading aftereffects of cutting edge innovative work in the field of WI. The enthusiasm for WI is becoming quick. We might want to welcome everybody, who are occupied with the WI related innovative work exercises, to join the WI people group. Your info and support will decide the fate of WI.

## REFERENCES

1. Hayles, N. K. (2014). My Mother was a Computer: Digital Subjects and Literary Texts. University of Chicago Press, Chicago, Illinois.

2. Israel, D. and Perry, J. (2010). What is information? In Hanson, P., editor, Information, Language, and Cognition, pages 1–19. University of British Columbia Press, Vancouver, Canada. Jacobs, I. (2010). W3C Mission Statement. Technical report, W3C. http://www.w3.org/Consortium/.

3. Jacobs, I. and Walsh, N. (2014). Architecture of the World Wide Web. Technical report, W3C. http://www.w3.org/TR/webarch/ (Last accessed Oct 12th 2011).

4. Kahn, R. and Wilensky, R. (2013). A framework for distributed digital object services. International Journal on Digital Libraries, 6(2).

5. Levy, P. (2014). Collective Intelligence: Mankind's Emerging World in Cyberspace. Plenum Press, New York City, New York, USA.

6. Masterman, M. (2011). Semantic message detection for machine translation, using an interlingua. In Proceedings of International Conference on Machine Translation of Languages and Applied Language Analysis, London, United Kingdom.

7. May, R., Levin, S., and Sugihara, G. (2011). Ecology for bankers. Nature, 451:893– 895.

8. Newman, M. (2014). Power laws, Pareto distributions and Zipf's law. Contemporary Physics, 46:323–351.

9. Oren, E., Delbru, R., Catasta, M., Cyganiak, R., Stenzhorn, H., and Tummarello, G. (2011). Sindice.com: A document-oriented lookup index for open linked data. International Journal of Metadata, Semantics, and Ontologies, 3(1):37–52.

10. Pas¸ca, M. (2012). Weakly-supervised discovery of named entities using web search queries. In Proceedings of the sixteenth ACM conference on Conference on information and knowledge management (CIKM), pages 683–690, Lisbon, Portugal.

11. Parsia, B. (2013). Message to www-rdf-comments@w3.org. http://lists.w3.org/Archives/Public/www-rdf-comments/2013JanMar/0366.html (Last accessed on May 20th 2011).

12. Quillian, M. R. (2011). Semantic memory. In Minsky, M., editor, Semantic Information Processing, pages 216–270. MIT Press, Cambridge, Massachusetts, USA.

13. Raggett, D., LeHors, A., and Jacobs, I. (2010). HTML 4.01 Specification. Recommendation, W3C. http://www.w3.org/TR/REC-html40/ (Last accessed March 7th 2011).

14. Rees, J. (2011). URI Documentation Protocol. Technical report, Science Commons. http://neurocommons.org/page/URI documentation protocol (Last accessed on Nov 20th 2011).

15. Salton, G., Wong, A., and Yang, C. S. (2014). A vector space model for automatic indexing. Communications of the ACM, 18(11):613–620.

16. Searle, J. (2014). The Construction of Social Reality. The Free Press, New York City, New York, USA.

17. Winograd, T. (2013). Towards a procedural understanding of semantics. Stanford Artificial Intelligence Laboratory Memo AIM-292.

18. Xu, J. and Croft, W. B. (2013). Query expansion using local and global document analysis. In Proceedings of the Nineteenth Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval, pages 4–11, Zurich, Switzerland.

19. Yarowsky, D. (2014). Unsupervised word sense disambiguation rivaling supervised methods. In Proceedings of the Meeting of the Association for Computational Linguistics (ACL), pages 200–200, Cambridge, Massachusetts, USA.

20. Zhai, C. and Lafferty, J. (2011). A study of smoothing methods for language models applied to ad hoc information retrieval. In Proceedings of the Twenty-Fourth Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval, pages 334–342, New Orleans, Louisiana, USA.