ONTOSPREAD: Activation of Concepts in Ontologies through the Spreading Activation algorithm

Jose María Álvarez[1], Diego Berrueta[2], Luis Polo[2], and José Emilio Labra[1]

[1] WESO RG, Universidad de Oviedo, Oviedo, Asturias, Spain,
{josem.alvarez,jelabra}@weso.es,
WWW home page: http://www.weso.es
[2] Fundación CTIC, Gijón, Asturias, Spain,
{diego.berrueta,luis.polo}@fundacionctic.org,
WWW home page: http://www.fundacionctic.org

Abstract. The present article introduces the ONTOSPREAD API for the development, configuration, customization and execution of the Spreading Activation techniques over semantic networks and more specifically over RDF graphs and ontologies arising from the Semantic Web area. These techniques have been used to the efficient exploration and querying of large and heterogeneous knowledge bases based on semantic networks in the Information or Document Retrieval domains. ONTOSPREAD implements the double process of activation and spread of concepts in ontologies, implicit graph structures, applying different restrictions of the original model like weight degradation according to the distance or others coming from the extension of these techniques like the converging paths reward. The main application of Spreading Activation lies in two different areas of interest to digital libraries: 1) construction of hybrid semantic search engines 2) ranking of information resources according to an input set of weighted resources. These techniques provide a whole framework to ease the information access, a common required features in the exploitation of new and existing digital libraries. Finally, in this work an evaluation methodology and an example using the Galen ontology are provided to validate the goodness, the improvement and the capabilities of this framework applied to digital libraries.

1 Introduction

The improvements in digitization lead us to a new environment in which digital libraries and archives are designed and used in a new way. This situation implies new challenges in the digital formats, storage (information is continuously growing) and information retrieval models. Following the recommendations of the European Commision [9] the digital libraries are a key factor to bring out the full economic and cultural potential of Europe's cultural and scientific heritage through the Internet. The online presence of material from different cultures and in different languages will make it easier for citizens to appreciate their own

cultural heritage as well as the heritage of other European countries. Besides its fundamental cultural value, cultural material is an important resource for new added value services. That is why more sophisticated software tools and methods are needed to meet the expectations of users easing the information retrieval of these large datasets and overcoming the classical problems of information overloading.

Initiatives like Semantic Web and Linked Data [19] tries to define vocabularies and ontologies enabling the data interoperability and sharing that enable by means of the Web infrastructure the access to the contents across different platforms and applications. The development of tools using these common data formats and models is largely implemented and representative to digital libraries but some algorithms and methods are not yet promoted to work with them in a standard way preventing the improvement and effectiveness of information access.

In this sense Spreading Activation technique (hereafter SA) introduced by [8], in the field of psycho linguistics and semantic priming, purpose a model in which all relevant information is mapped on a graph as nodes with a certain "activation value". Relations between two concepts are represented by a weighted edge. If a node is activated their activation value is spread to their neighbor nodes. This technique was adopted by the computer science community and applied to the resolution of different problems, see Sect. 3. In the field of digital libraries this technique can ease the information access providing a connectionist method to retrieve data like brain can do. Although SA is widely used, more specifically in recent years have been successfully applied to ontologies, a common and standard API is missing and each third party interested in their application must to implement its own version [29] of SA.

Taking into account this new information realm and the leading features of putting together the SA and the Semantic Web technologies, new enriched services of searching, matchmaking, recommendation or contextualization can be implemented to fulfill the requirements of access information in digital libraries of different trending scopes like e-procurement, legal document databases [4], e-tourism or e-health.

1.1 Main Contributions

The proposed work aims to afford a framework for SA to ease the configuration, customization and execution of them over semantic networks and more specifically over RDF graphs and ontologies. It is relevant to digital libraries access and interoperability due to the fact that this technique provides a set of proven algorithms for retrieving and recommending information resources in large knowledge bases. Following the specific contributions of this work are listed:

- Study and revision of the classical constrained SA.
- Study and definition of new restrictions for SA applied to ontologies and RDF graphs.

- Implementation of a whole and extensible framework (called ONTOSPREAD API). to customize and perform the SA based on good practices in software programming.
- Outlining of a methodology to configure and refine the execution of SA.
- An example of configuration and refinement applying SA over a well-known ontology, the Galen ontology.

2 Background

In this section, the theoretical model of *SA* [8,2] is reviewed to illustrate the basic components and the operations performed by SA during their execution, specially the spread of the activation from a node to their adjacent, see Fig. 1. This model is made up of a conceptual network of nodes connected through relations (conceptual graph). Taking into account that nodes represent domain objects or classes and edges relations among them, it is possible to establish a semantic network in which SA can be applied. The process performed by the algorithm is based on a thorough method to go down the graph using an iterative model. Each iteration is comprised of a set of beats, a stepwise method, and the checking of a stop condition. Following the different stages of SA are presented and defined:

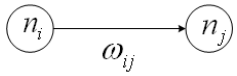

Fig. 1. Graphical model of *Spreading Activation*

***Preadjustement*:** This is the initial and optional stage. It is usually in charge of performing some control strategy over the target semantic network.

***Spreading*:** This is the spread stage of the algorithm. Concepts are activated in activation waves. The spreading node activates its neighbor nodes, see Fig. 2.

The calculation of the activation rank I_i of a node n_i is defined as follows:

$$I_i = \sum_j O_j \omega_{ji} \tag{1}$$

I_i is the total inputs of the node n_i, O_j is the output of the node n_j connected to n_i and ω_{ji} is the weight of the relation between n_j and n_i. If there is not relation between n_j and n_i then $\omega_{ji} = 0$.
The activation function f is used to evaluate the "weight" of a node and decide if the concept is active.

$$N_i = f(I_i) = \begin{cases} 0 & \text{if } I_i < \jmath_i \\ 1 & \text{if } I_i > \jmath_i \end{cases} \tag{2}$$

N_i is 1 if the node has been activated or 0 otherwise. $\jmath_i$, the threshold activation value for node i, depends on the application and it can change from a node to others. The activation rank I_i of a node n_i will change while algorithm iterates.

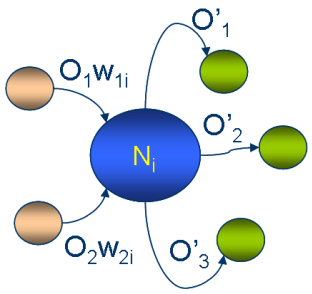

Fig. 2. Activation of concepts in *Spreading Activation*

***Postadjustment*:** This is the final and optional stage. As well as *Preadjustment* stage, it is used to perform some control strategy in the set of activated concepts.

3 Related Work

Since SA was introduced by [8] in the field of psycho linguistics and semantic priming it has been applied to the resolution of problems trying to simulate the behavior of the brain using a connectionist method to provide an "intelligent" way to retrieve information and data.

The use of SA was motivated due to the research on graph exploration [24,2,7]. Nevertheless the success of this technique is specially relevant to the fields of Document [20] and Information Retrieval [10,3,1,18]. It has been also demonstrated its application to extract correlations between query terms and documents analyzing user logs [11] and to retrieve resources amongst multiple systems [27] in which ontologies are used to link and annotate resources.

In recent years and regarding the emerging use of ontologies in the Semantic Web area new applications of SA have appeared to explore concepts [25,6] addressing the two important issues: 1) the selection and 2) the weighting of additional search terms and to measure conceptual similarity [17]. On the other hand, there are works [21,12] exploring the application of the SA on ontologies in order to create context inference models.The semi-automatically extension and

refinement of ontologies [22] is other trending topic to apply SA in combination with other techniques based on natural language processing. Data mining, more specifically mining socio-semantic networks[30], and applications to collaborative filtering (community detection based on tag recommendations, expertise location, etc.) are other potential scenarios to apply the SA theory due to the high performance and high scalability of the technique.

In particular, annotation and tagging [14,15] services to gather meta-data [16] from the Web or to predict social annotation [5] and recommending systems based on the combination of ontologies and SA [13] are taken advantage of using SA technique.

Finally the semantic search [28,31] is a highlight area to apply SA following hybrid approaches [4,26] or user query expansion [23] combining metadata and user information.

Although this technique is widely accepted and applied to different fields open implementations, Texai[3] company offers a proprietary implementation of SA, are missing. Moreover the Apache Mahout [4] project, a recent scalable machine learning library that supports large data sets, does not include an implementation of SA instead of providing algorithms for the classification, clustering, pattern mining, recommendation and collaborative filtering of resources in which SA should be representative.

4 ONTOSPREAD API

5 Evaluation of ONTOSPREAD API

6 Conclussions and Future Work

References

1. Maristella Agosti and Fabio Crestani. A Methodology for the Automatic Construction of a Hypertext for Information Retrieval. In *SAC '93: Proceedings of the 1993 ACM/SIGAPP symposium on Applied computing*, pages 745–753, New York, NY, USA, 1993. ACM Press.
2. John R. Anderson. A Spreading Activation Theory of Memory. *Verbal Learning and Verbal Behavior*, (1):261–295, 1983.
3. Helmut Berger, Michael Dittenbach, and Dieter Merkl. An adaptive information retrieval system based on associative networks. In *CRPIT '04: Proceedings of the first Asian-Pacific conference on Conceptual modelling*, pages 27–36, Darlinghurst, Australia, Australia, 2004. Australian Computer Society, Inc.
4. Diego Berrueta, Jose Emilio Labra, and Luis Polo. Searching over Public Administration Legal Documents Using Ontologies. In *Proceedings of Joint Conferente On Knowledge-Based Software Engineering (JCKBSE 2006)*, pages 167 – 175, July–August 2006.

[3] http://texai.org/

[4] http://mahout.apache.org/

5. Abon Chen, Hsin-Hsi Chen, and Polly Huang. Predicting social annotation by spreading activation. In *Proceedings of the 10th international conference on Asian digital libraries: looking back 10 years and forging new frontiers*, ICADL'07, pages 277–286, Berlin, Heidelberg, 2007. Springer-Verlag.
6. H. Chen and T. Ng. An Algorithmic Approach to Concept Exploration in a Large Knowledge Network (automatic thesaurus consultation): Symbolic Branch-and-Bound search vs. connectionist Hopfield net activation. *J. Am. Soc. Inf. Sci.*, 46(5):348–369, 1995.
7. Paul R. Cohen and Rick Kjeldsen. Information Retrieval by Constrained Spreading Activation in Semantic Networks. *Inf. Process. Manage.*, 23(4):255–268, 1987.
8. Allen M Collins and Elizabeth F Loftus. A spreading activation theory of semantic processing. *Psychological Review*, 82(6):407–428, 1975.
9. European Commission Communications. On the Digitisation and Online Accesibility of Cultural Material and Digital Preservation. Commission recommendation, European Commission, 2006.
10. F. Crestani. Application of Spreading Activation Techniques in Information Retrieval. *Artificial Intelligence Review*, (11):453–482, 1997.
11. H. Cui, J. Wen, J. Nie, and W. Ma. Query Expansion by Mining User Logs. *IEEE Transaction on Knowledge and Data Engineering*, 15(4):829–839, July 2003.
12. Alan J. Dix, Akrivi Katifori, Giorgos Lepouras, Costas Vassilakis, and Nadeem Shabir. Spreading Activation over Ontology-Based Resources: from Personal Context to Web Scale Reasoning. *Int. J. Semantic Computing*, 4(1):59–102, 2010.
13. Qi Gao, Junwei Yan, and Min Liu. A Semantic Approach to Recommendation System Based on User Ontology and Spreading Activation Model. In *NPC '08: Proceedings of the 2008 IFIP International Conference on Network and Parallel Computing*, pages 488–492, Washington, DC, USA, 2008. IEEE Computer Society.
14. Jose Emilio Labra Gayo, Patricia Ordoñez de Pablos, and Juan M. Cueva Lovelle. Combining Collaborative Tagging and Ontologies in Image Retrieval Systems. 2007.
15. Jose Emilio Labra Gayo, Patricia Ordoñez de Pablos, and Juan Manuel Cueva Lovelle. WESONet: Applying semantic web technologies and collaborative tagging to multimedia web information systems. *Comput. Hum. Behav.*, 26:205–209, March 2010.
16. Fatih Gelgi, Srinivas Vadrevu, and Hasan Davulcu. Improving Web Data Annotations with Spreading Activation. In *WISE*, pages 95–106, 2005.
17. Stephan Gouws, G-J van Rooyen, and Herman A. Engelbrecht. Measuring Conceptual Similarity by Spreading Activation over Wikipedia's Hyperlink Structure. In *Proceedings of the 2nd Workshop on The People's Web Meets NLP: Collaboratively Constructed Semantic Resources*, pages 46–54, Beijing, China, August 2010. Coling 2010 Organizing Committee.
18. Maurice Grinberg, Vladimir Haltakov, and Hristo Stefanov. Approximate Spreading Activation for Efficient Knowledge Retrieval from Large Datasets. In *Proceeding of the 2011 conference on Neural Nets WIRN10: Proceedings of the 20th Italian Workshop on Neural Nets*, pages 326–333, Amsterdam, The Netherlands, The Netherlands, 2011. IOS Press.
19. Tom Heath and Christian Bizer. *Linked Data: Evolving the Web into a Global Data Space*, volume 1. Synthesis Lectures on the Semantic Web: Theory and Technology, Morgan & Claypool, 2011.
20. H.R.Turtle. *Inference Networks for Document Retrieval.* PhD thesis, University Illinois, Urbana, IL, USA., 1991.

21. Akrivi Katifori, Costas Vassilakis, and Alan J. Dix. Ontologies and the brain: Using spreading activation through ontologies to support personal interaction. *Cognitive Systems Research*, 11(1):25–41, 2010.
22. W. Liu, A. Weichselbraun, A. Scharl, and E. Chang. Semi-Automatic Ontology Extension Using Spreading Activation. *Universal Knowledge Management*, 0(1):50–58, 2005.
23. Jian-Yun Nie. Query Expansion and Query Translation as Logical Inference. *J. Am. Soc. Inf. Sci. Technol.*, 54(4):335–346, 2003.
24. Scott Everett Preece. *A Spreading Activation Network Model for Information Retrieval*. PhD thesis, University Illinois, Urbana, IL, USA., 1981.
25. Yonggang Qiu and Hans-Peter Frei. Concept-based query expansion. In *Proceedings of SIGIR-93, 16th ACM International Conference on Research and Development in Information Retrieval*, pages 160–169, Pittsburgh, US, 1993.
26. Cristiano Rocha, Daniel Schwabe, and Marcus Poggi de Aragão. A Hybrid Approach for Searching in the Semantic Web. In *WWW*, pages 374–383, 2004.
27. Kinga Schumacher, Michael Sintek, and Leo Sauermann. Combining Metadata and Document Search with Spreading Activation for Semantic Desktop Search. In Sean Bechhofer, Manfred Hauswirth, Jörg Hoffmann, and Manolis Koubarakis, editors, *Proc. of ESWC*, pages 569–583. Springer, June 2008.
28. Ján Suchal. On finding power method in spreading activation search. In Viliam Geffert, Juhani Karhumäki, Alberto Bertoni, Bart Preneel, Pavol Návrat, and Mária Bieliková, editors, *SOFSEM (2)*, pages 124–130. Safarik University, Kosice, Slovakia, 2008.
29. Proletina Todorova, Atanas Kiryakov, Damyan Ognyanoff, Ivan Peikov, Ruslan Velkov, and Zdravko Tashev. D2.4.1 Spreading Activation Components (v1). Deliverable or research project, LarKC FP7 project–215535, 2009.
30. Alexander Troussov, Mikhail Sogrin, John Judge, and Dmitri Botvich. Mining Socio-Semantic Networks Using Spreading Activation Technique. 2008.
31. By Michael Wolverton, Michael Wolverton, Michael Wolverton, Barbara Hayes-roth, and Barbara Hayes-roth. Retrieving Semantically Distant Analogies with Knowledge-Directed Spreading Activation. In *In Proceedings AAAI-94*, pages 56–61. AAAI Press, 1994.