

# Semantic Information storage and retrieval in a Peer-to-Peer corporate memory

Ana B. Rios-Alvarado, Ricardo Marcelin-Jiménez, R. Carolina Medina-Ramírez

Universidad Autónoma Metropolitana - Iztapalapa  
México, DF

**Abstract.** The paper presents a semantic approach for storing and retrieving documents from a Corporate Semantic Web (CSW). We illustrate the approach through the embedding of two graphs  $G_1$  into  $G_2$ .  $G_1$  represents the CSW and whose nodes represent a collection of documents having a common range of semantic indices.  $G_2$  represents a P2P storage network. We use "Ant Colony Optimization" metaheuristic, to solve the corresponding instances of graph embedding.

## 1 Introduction

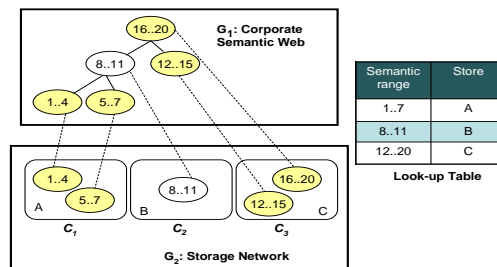
The semantic Web approach [1] relies on ontologies, annotations and formal knowledge representation languages. A "Corporate Semantic Web (CSW) is built up from ontologies, resources (documents or humans) and annotations on these resources, where these annotations rely on the ontologies[2]. There is a meetpoint between Web and corporate memories: both gather heterogeneous and distributed information and share the same concern about the relevance of information retrieval. Nevertheless, corporate memory has a context, an infrastructure and a scope limited to the organisation where they are applied.

IP routing task, at the Internet, is supported by two complementary procedures: table maintenance and table querying. In this work, we propose the organization of document storage and retrieval in a Corporate Semantic Web (CSW), based on two procedures: First, we solve content location and built a table, whose entries shows the places in charge of a given set of documents. Second, we perform look-up on this table in order to consult the corresponding contents. An ontology can be regarded as a hierarchy of concepts. Each of them corresponds to a semantic index. Besides, each semantic index has associated a collection of documents belonging to the CSW. Therefore, we can model a CSW as a graph  $G_1$  (Fig. 1), where each node is featured by two parameters: a range of semantic indices and a weight. The first one represents the concepts it gathers according to its place in the hierarchy. The second one, represents the amount of information given by the collection of documents in the given range. We model the storage network using a second graph  $G_2$ . Each of its nodes (from now on stores) has an associated capacity  $c_j$  that features the maximal amount of information it is able to contain.

## 2 Methodology and assumptions

Content placement implies the embedding of  $G_1$  into  $G_2$ . We decided to tackle our instances of graph embedding using the ant colony optimization algorithm (ACO)[3]. Our method consists of creating  $z$  scout ants. Every ant is charged to perform a random depth first search on  $G_1$ . As each ant travels across the graph, it associates the nodes that visits to a given store  $j$  of  $G_2$ . When the aggregated nodes weight exceeds the capacity of the current store, it reassigns the last node to successor store  $j+1$  and starts this filling process over again, as long as there are still nodes to visit.

When our particular instance of graph embedding is successfully solved, each store receives a copy of the look-up table. Each row in this table has two parts, the left entry indicates a range of semantic indices, while the right entry indicates the store in charge of the documents in this range. Figure 1 shows how  $G_1$  has been embedded into  $G_2$  and the Look-up Table. We have used a discrete event simulator [4], for implementing our algorithm.



**Fig. 1.** Embedding the corporate semantic Web ( $G_1$ ) into a distributed storage network( $G_2$ )

We have run our simulation using a variable number  $z$  of ants, nodes in  $G_1$  have weights following an uniform random distribution, and stores in  $G_2$  have a constant capacity.

## 3 Conclusion

We have presented a semantic approach for storing and retrieving documents from a Corporate Semantic Web (CSW). We illustrate the approach through the embedding of two graphs  $G_1$  into  $G_2$ . We have used "Ant Colony Optimization", to solve the corresponding graph embedding.

From preliminar results, we can say that there is an optimal number of initial ants producing the highest variance. This optimal depends on the size of  $G1$ , and is roughly  $O(\sqrt{n})$ , where  $n$  is the total number of nodes in  $G1$ .

## References

1. Shadbolt, N. Berners-Lee, T., Hall W.: The semantic Web Revisited, IEEE intelligent Systems, V21, N3, pp 96-101. 2006
2. Dieng-Kuntz, R.:Corporate Semantic Webs, ERCIM News No. 51, pp19-21. 2002
3. Dorigo, M.: Optimization, Learning and Natural Algorithms. Ph.D. Thesis, Dept. of Electronics, Politecnico di Milano. Italy (1992)
4. Marcellín-Jiménez, R.: A Flexible Simulator for Distributed Algorithms. Proceedings of the ENC03 by the IEEE Computer Society Press, (2003) pp. 176-181.