

Corporate Brains - A Case Study of a real world topic map application

Elmar Seestädt <elmar.seestaedt@emprise.de>

Abstract

In every day business a lot of time is spent on finding a contact person for a specific problem or subject. This presentation will outline the ontology, application architecture and experiences gained during a real world topic map project.

Table of Contents

1. Introduction	1
2. Problem Domain	1
2.1. Finding the right contact	1
2.2. Buzzwords, Technical Terms and Acronyms	2
2.3. Technical Requirements	2
3. Ontology	2
3.1. Persons	2
3.2. Topics	2
3.3. Projects	2
3.4. Departments	2
3.5. Terms	3
3.6. Categories	3
4. Possible solution paths	3
4.1. Topic Map approach	3
4.2. Classical O/R bases solution	4
5. Conclusion	4
Glossary	4

1. Introduction

EMPRISE Consulting Düsseldorf was asked by a leading German and international telecommunications provider to help with a knowledge management problem. In cooperation with the customer EMPRISE worked out the core problems of daily business and developed a prototype application to solve the identified problems quickly . The prototype was widely accepted by the customers employees and management. After a short production time demand for a more sophisticated solution arose which was successfully satisfied by a second release.

2. Problem Domain

2.1. Finding the right contact

A lot of effort has been taken into making implicit knowledge publicly available. The know how accumulated at experienced employees is essential for successfull companies. It can hardly be written down and archived in a document management system. We believe that communication between the companies experts is inevitable. In daily business a lot of time is spent on finding the right contact. Sometimes it takes hours or even days and weeks with several unsuccessfull tries to find the right contact. Usually existing intranets or yellow pages are not sufficient. In an intranet the semantics of the association between topics, projects and heads are lost during full text indexing and search. Yellow pages normally lack the information of the responsibilities of the listed persons. The major

goal of the corporate brains project was to reduce the time spent on finding the right contact and to enable communication between experts.

2.2. Buzzwords, Technical Terms and Acronyms

New employees are often lost in the wording microcosms of their new employer. A lot of buzzwords and technical terms come across IT professionals and people are lost in the jungle of acronyms. Therefore the application should be backed with a glossary that is semantically linked to the stored knowledge.

2.3. Technical Requirements

Orthogonal to this two business requirements stood the technical requirement to display the knowledge dynamically to the users. Users should be able to search the repository via a simple interface. And the double administration of core attributes of persons like phone numbers or room locations should be avoided. Therefore the application had to be integrated with the existing Lightweight Directory Access Protocol ([LDAP](#)).

3. Ontology

Six entities have been identified among which heads are certainly the central entity with a lot of different associations to the other entities. Common to all entities is the need for description and the ability to have several names which are handled as synonyms. Everyone of the defined entities is backed by an association to the underlying glossary to enable fast discovery of technical terms.

3.1. Persons

Persons are the cornerstones of every organization. They deal with the topics, are involved in projects and accumulate a lot of knowledge in different fields of interest. In situations where questions arise these persons just have to be found. Therefore persons are associated with topics, organizational units and projects in the role of contact person. Persons have several attributes beside a short textual description. These attributes are all derived from the existing [LDAP](#) to ensure data consistency.

3.2. Topics

Topics are one central entity in the ontology. The word topic is not used in the very general way the topic map standard defines what a topic e.g. subject is but in a much more concrete definition. A topic in the project context is any field of interest in the company's domain, a task or collection of tasks for which anyone in the company has the responsibility. There are some more associations to other entities where topics play a role. Each topic has an associated contact person and one responsible organizational unit. And a topic can be the context for certain projects and therefore projects itself are subject of being context for a topic. Examples for topics could be "Web Enablement", "IT Security", "Knowledge Management".

3.3. Projects

Projects are a temporal institution which is represented in the knowledge repository to allow the discovery of key tasks. Projects are put in context to one or more subjects as some kind of topical categorization. For example this allows the retrieval of all projects dealing with knowledge mangement.

3.4. Departments

The hierachical organizational structure of the company has been modeled with departments. Departments can have one superordinate and several subordinate departments. Departments are involved in projects and the business leader of a certain project can be identified with a distinct association.

3.5. Terms

Terms are used to define technical terms. Terms have a validity scope to allow different departments different interpretations and definitions of a certain term. This also allows new employees to make queries like "What are the words to know in my new department?".

3.6. Categories

Categories are used to categorize logical types of terms. Categories appear in a hierarchical structure e.g. supercategory "IT Terms" with the two subcategories "software" and "hardware". They are only used as part of the glossary.

4. Possible solution paths

Two different approaches have been considered. The first one is a solution with a topic map engine and direct adoption of the topic map paradigm and the second is a classical O/R based approach with topic map ideas in mind.

4.1. Topic Map approach

The topic map paradigm as standardized by ISO has some advantages over the classical approach, though it is burdened with some black holes in experience. Using a topic map engine offers the advantage of a full featured engine that encapsulates the complete RDBMS backend. The engine allows the retrieval of topics in several manners using the toolset these engines offer. New entities can easily be introduced to the ontology without changing the database. Inference rules can be applied for non trivial knowledge retrieval. And the more sophisticated features like scope allow a new user experience in application usability and constraining search criteria.

4.1.1. Application Architecture

Where a classical object relational approach would lead into a three tier architecture with client, application server and RDBMS backend the preferred topic map approach introduces a fourth, the topic map tier. In the first step the application has been developed adhering to the ISO topic map standard and using the open source topic map engine TM4J. The Apache Lucene full text search engine has been plugged into TM4J to allow full text search in basenames and inline occurrences. A few other commercially available tools have been evaluated but considered to be too expensive. The business layer itself is downsized to a thin façade to the topic map engine where all processing logic is implemented. The application tier can be understood as a constraining façade which prevents the user from using all topic map features and binds him to the defined ontology. The view is bound to the business layer interfaces and is independent from the topic map or database approach. It makes heavy use of the Jakarta Struts framework and the newly introduced tiles component. All basic components in the view are configurable via parameters and plugged together in a configuration file. A closed user group is allowed to edit the content. Anybody who has administrator rights can edit any content. This requires a trustful, cooperative work of a rather small group of administrators.

4.1.2. Topic Map Approach Problems

Already in the requirements analysis phase some disadvantages of the topic map paradigm showed up. First of all the lack of an standardized ontology language. This imposed proprietary definition of the ontology and prevented an automatic verification of the topic map at runtime. There are some schema languages proposed by several vendors and the bond university but there is no built in support for runtime verification in TM4J. Secondly there is no standardized query language which bounds the application to a certain topic map engine and prevents easy exchange. At least TM4J and Ontopia are working hand in hand and both are using tolog, a prolog like language for queries. The topic naming constraint imposes recognizable constraints to application logic which will be discussed in the next section.

A lot of the problems that arose with the TM4J topic map engine were bugs in the engine itself. TM4J provides three different backends which are an in memory implementation, an Ozone OO Database implementation and a relational database backend using the Hibernate O/R library. TM4J was not able to deal with scopes properly which because of the Topic Naming Constraint were heavily used in the application. The TNC says that every topic which has the same basename in the same scope is deemed to be automatically merged. This implies that every person

Elmar Seestädt

Senior Consultant, Solution Manager
EMPRISE Consulting Düsseldorf GmbH
Düsseldorf
Germany
elmar.seestaedt@emprise.de

Elmar Seestädt studied Applied System Science and Information Management at the universities of Osnabrück and Münster. At university he participated in several projects including a submission of an X3D proposal and prototype implementation an conforming browser. After a short episode in an international consulting company he decided to join EMPRISE Consulting Düsseldorf GmbH. Here he is involved in the fields of Web Applications and a leading consultant for Knowledge Management.