

Case Study: Semantic-based Search and Query System for the Traditional Chinese Medicine Community

Zhaohui Wu and Huajun Chen, Zhejiang University, Meng Cui and Ainin Yin, China Academy of Chinese Medicine Sciences (CACMS), China

May 2007



General Description

Introduction

The long standing curation effort in the Chinese medicine community has been accumulating huge amounts of data, which are typically stored in relational data management systems such as Oracle, and published as HTML pages for public presentation. The China Academy of Chinese Medicine Sciences (CACMS) hosts much of the data. However, it has become increasingly difficult and time-consuming to manage the data, and the links to data sources from other institutions. Although they could be physically put together, but the logical links among the data are usually implicit or even lost at all. Moreover, the randomness of choosing names for relational tables, table columns, and record values make the data only understandable to the original database designer and data curator and exclusively controlled by ad hoc applications. This has caused a huge hindrance in sharing, and reusing data across databases, and organizational boundaries.

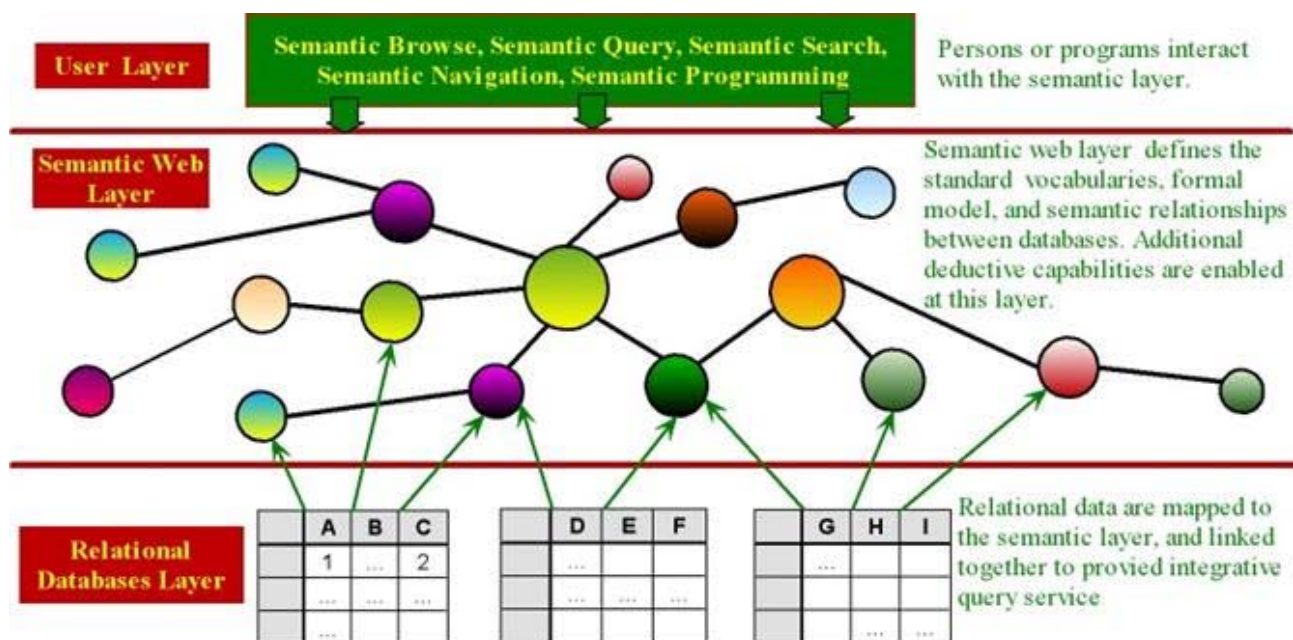


Figure 1: This figure shows the architecture of the Semantic Web layer and its role in unifying and linking heterogeneous relational data.

We have applied Semantic Web technologies to relational data to make it more sharable and machine-processable. We have also developed a semantic-based search and query system for the traditional Chinese medicine community in China (TCM Search), which has been deployed for real life usage since fall 2005. For the TCM system, a TCM ontology and the semantic layer has been constructed to unify and link the legacy databases, which typically have heterogeneous logic structures and physical properties. Users and applications now only need to interact with the semantic layer, and the semantic interconnections allow for searching, querying, navigating around an extensible set of databases without the awareness of the boundaries (Figure 1). Additional deductive capabilities can then be implemented at the semantic layer to increase the usability and re-usability of data. Besides, a visualized mapping tool has been developed to facilitate the mapping from relational data to the TCM ontology, and an ontology-based query and search portal has also been implemented to assist the semantic interaction with the system.

Mapping from relational data to semantic web ontologies

The informal approach taken for the selection of names and values within relational databases makes the data only understandable by specific applications. The mapping from relational data to the Semantic Web layer makes the semantics of the data more formal, explicit, and prepared for sharing and reusing by other applications. However, because of the inherent model difference between relational data model and the Semantic Web languages, mapping is always a complicated task and can be time-consuming and error-prone. We have therefore developed a **visualize mapping tool** to simplify the work as much as possible, as Figure 2 displays. The tool generates mapping rules that are used when a SPARQL query is rewritten into a set of SQL queries.

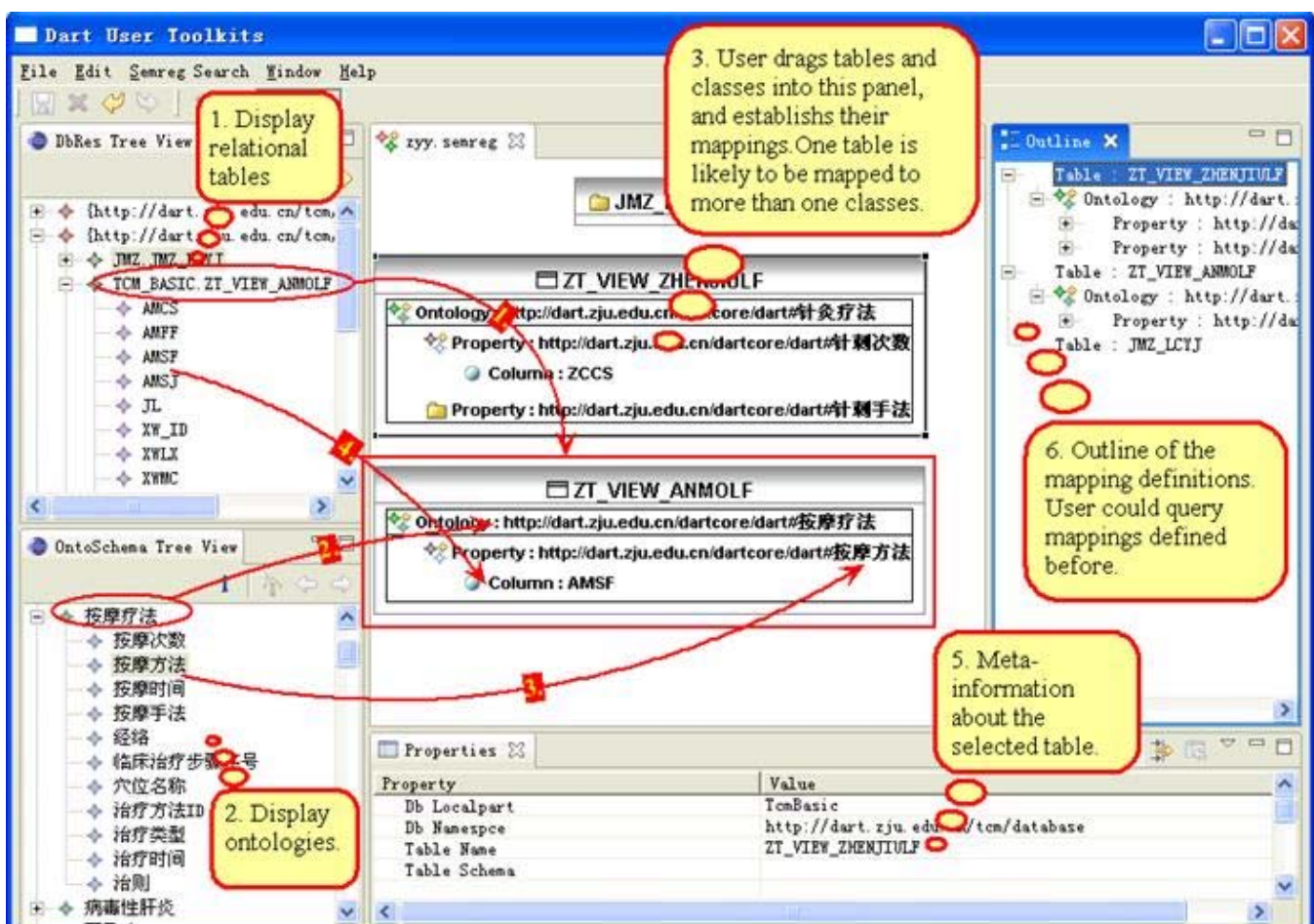


Figure 2: This figure shows a visualized mapping from a TCM relational database to the TCM ontology.

Ontology-based query and search across database boundaries

As Figure 3 displays, we have developed a semantic-based query and search portal to assist in user interaction with the system. Basically, this system consists of two components. The search component enables users to perform full-text searching through all of the integrated data sources using keywords that is similar to common Internet search engines, while the query component supplies with a means for handling more complex semantic queries posed against the semantic web ontology.

The ontology plays an important role in the mediation of the query, search and navigation. At first, it serves as a logic layer for users in constructing semantic queries. The form-based query interface is automatically generated based on the ontological structure. The constructed semantic query will then be translated into SQL queries based on the mapping rules generated by the mapping tool. At second, it enables semantic navigation across database boundaries during query and search. At third, it also serves as a control vocabulary to facilitate search by making semantic suggestions such as synonyms, and related concepts.

The image shows a screenshot of a web browser displaying a search portal for Traditional Chinese Medicine (TCM). The interface includes a search bar, a list of search results, and a dynamic form-based query interface. Annotations highlight key features:

- Ontological classes:** A callout box points to a list of search results, indicating that clicking on these classes leads to a dynamic form-based query interface.
- Based on the semantic relations defined at the ontological level, user can keep searching and navigating over the integrated databases without the awareness of the database boundaries.** A callout box points to the search results, indicating that the system allows for seamless navigation across database boundaries.
- Synonyms and Paronyms:** A callout box points to a list of suggestions, indicating that the system provides semantic suggestions such as synonyms and related concepts.
- Semantic association:** A callout box points to a list of suggestions, indicating that the system provides semantic suggestions such as synonyms and related concepts.

When full text search returns too much results, clicking the classes leads to a dynamic form-based query interface by which user could specify semantic query, thereby getting more accurate and appropriate results.

Figure 3: This figure shows the ontology-based query and search portal of the TCM search system.

Key Benefits of Using Semantic Web Technology

- Exposing of legacy data through a semantic layer so that it can be more easily reused and recombined.
- Linking data across database boundaries so as to enabling more intuitive query, search, and navigation without the awareness of the boundaries.
- The ontology serves as the control vocabulary to make semantic suggestions such as synonyms, related concepts to facilitate query and search.
- Reasoning capability such as sub-classing, transitive property can then be implemented at the semantic layer to increase the query expressiveness so as to retrieve more complete answers.
- Allows for more advanced data analysis and integrative knowledge discovery based on the huge web of data.