

Ontology Integration Research on XML Query and Normalization

Luo Gang^{1,a}, Xu Hong-feng^{2,b}

¹Department of Computer Science, Qiannan Normal College for Nationalities. Duyun 55800 China

²College of Economics and Management Guizhou Normal University, Guizhou, Guiyang, 550001

^a3138918343@qq.com; ^b3479064@qq.com;

Keywords: ontology; integration; XML

Abstract. Nowadays the need for sharing and exchanging information is becoming more and more important. Based on ontology, this paper proposes an integration framework for manufacturing systems. With the XML has been more widely used in many fields, the development of the sophisticated XML-based data management System, is becoming the goals of many researchers. The establishments of the query methods and the normalization mechanisms which are adapted for the characteristics of XML data are the foundations making the XM-based data management system practical.

Introduction

In today's complex manufacturing environment companies are trying to respond to market demands and changes in the supply chain by providing demand visibility directly to the manufacturing floor through plant-wide systems integration. Flexible integration of resources and engineering collaboration of self-governed SMEs is expected to facilitate the strengthening of the competitive position of SMEs in their struggle for the future through acquisition of a critical mass in terms of necessary skills, capabilities and capacities[1].

There's no denying that data is one of your company's most important assets, and data integration constitutes the backbone of your enterprise's IT systems. Choosing the wrong technology for data integration can cause harmful, long-lasting effects that impact not only the IT budget, but also the productivity and responsiveness of critical business divisions within your enterprise.

With so many years invested in the older data integration technology base, it is easy to overlook some of the nagging and persistent problems that it created. These problematic issues include the following:

(1)The "hub-and-spoke" architecture, still used by many data integration environments, is costly to implement and often becomes a bottleneck that slows down the overall IT systems.

(2) Many data integration projects are still treated as independent projects, with no overall consistency, resulting in silos of integration. This is exacerbated by most integration platforms' lack of support for multiple typologies of data latency and volume; for example, few systems support high-volume data batches and trickle-feed real-time data flows within the same platform.

(3)With top-down procedural design being the norm for many integration platforms, development and maintenance becomes very costly; this lack of flexibility can considerably hinder the time to market for a complete solution.

(4)Tight coupling of data integration processes and operational systems results in a lack of flexibility; updates are much more costly and cannot be performed with the required agility of today's fast-paced business environment.

ONTOLOGY-BASED SERVICE RESOURCE BASE

The aforementioned service resource model and RDF-based resource metadata expression provide the description method of complex service resource. But we need construct various resource bases rapidly in service alliance, so that we can search and utilize the resource. Now people use relation database to store the metadata and content, but this method needs huge database structure, and needs to maintain these data. The whole work is very great. So according to the correlative theory and

method of knowledge engineering and ontology, the following method of knowledge-based service resource ontology is referred.

(1) Ontology of service resource

OIL is chosen to be the standard of constructing the ontology model of the service resource after studying AL-Log, CARIN, CLASSIC, DLR, F-Logic, Loom, OIL and OKBC. OIL is referred as a kind of ontology model language by the researchers of Amsterdam University of Holland and Manchester University of Britain. It is based on XOL namely OKBC-lite knowledge model based on XML[2]. OIL includes ontology container, class, slot, restriction of slot, value of slot and so on.

The facet method of application service taxonomy has the following class hierarchy in software component.

Ontology structure has six domains. They are component domain, implementation description domain, information domain, environment domain, style domain and edition domain. Every domain has many slots. The former four domains describe the content of a component, and the later two domains describe the relationship among components.

(2) Ontology construction of component resource

According to the aforementioned ontology construction principle, in Protégé environment, the ontology-base oriented component resource is developed. At first, the class and slot structure of resource base are designed as Figure 1.

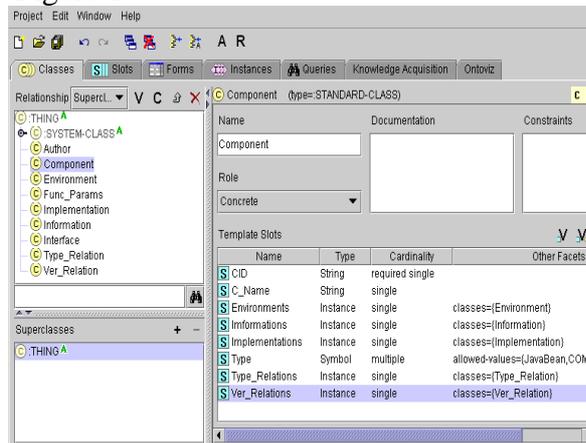


Fig.1: Class and slot structure based-ontology of Component resource in protégé environment

When constructing the class and slot structure of component resource, commonality and agility of this structure must be considered in order to describe various components easily. Implementation class is the key class, because it includes some important information such as function of component, interface and so on. Interface class is designed to be a slot of Implementation, so that it can contain complex interface information. After class and slot is designed, the component of this structure can be derived as Fig.2.

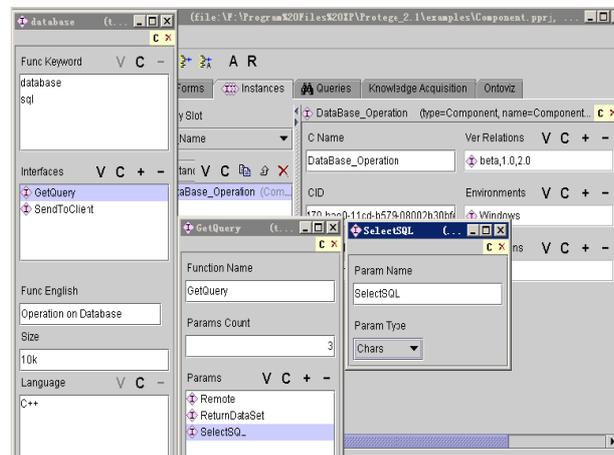


Fig.2: An example of component resource ontology in protégé environment

After constructing the resource ontology base, we can search all needed resource through resource property. And we can develop Java API and realize the integration of service resource system in protégé environment.

Requirements of the PLIBEI Platform

Figure 3 sketches the proposed framework and working mechanism of the PLIBEI platform. From Figure 3, we can find that the core of PLIBIE platform is mainly composed of two parts: web-based parts library and the part search engine.

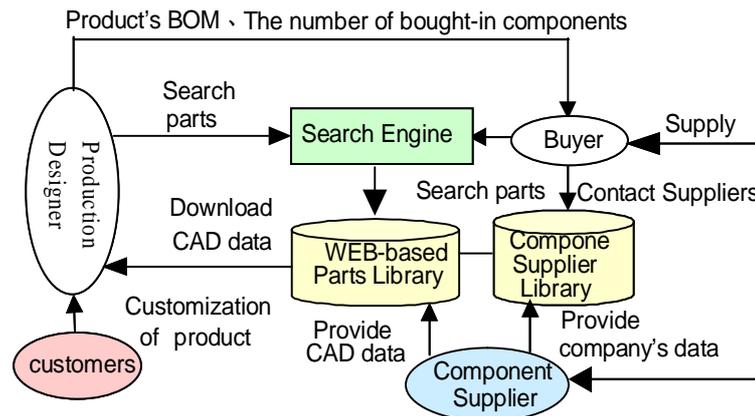


Fig.3. The framework and working and working mechanism of the PLIBEI platform.

Part information search engine: Part search engine is the bridge of information exchange and integration. It is differ from the common engine. A part search engine should provide semantic feature search ability besides the keywords search and industry classification search. For example, when a part buyer query a hexagonal machine screw, he should specify what kind of product (hexagonal machine screw), and the characteristic properties (total length, threaded length, diameter, coating etc.) The buyer and the supplier should agree on term of this product.

From previous analysis, we believe with ontologies there's now a model at hand to fit for these requirements. Ontologies serve as a means for establishing a conceptually concise basis for communicating knowledge for much purpose. In our scenario, we restrict our attention to domain ontologies-PLIB ontology. In next sections, we'll introduce what is PLIB ontology and how the PLIB ontology can meet these requirements[3].

XML Query and Normalization

Currently, the main research of XML is divided into two aspects. One is for XML, this kind of semi-structured data storage, query and management of the primary database, the data and metadata completely using XML structure, said its underlying data storage formats, such as object model and relational model, etc.). The other is a mutual conversion between it and the relational database, using the mature relational database technology to deal with XML data. Because after a direction is has practical significance, thus become a focus in the study of XML[4].

And in addition to storage solution, indexing technology is decided to a database system is one of the most important factor. If no building index structure of XML documents, then any query for XML data is likely to lead to the entire document tree traversal, with the increase of XML data sets, this overhead is unbearable. Therefore, the research of XML indexing technology has high theoretical and practical value.

Although traditional indexing technology after a long period of accumulation has been relatively mature, but, this kind of indexing technology is mainly aimed at according to the value (rather than have a certain relationship mode) positioning data record function, not too concerned about the logical relationship between data record; And the basic characteristics of XML data query is

according to the characters of model (describe the relationship between the structure of regular path expression form) extraction fits a pattern of the input data, therefore, the main content of the XML index is designed for pattern matching technology.

XML very carefully bit by bit to files or other data source and the abstract model of the XML document. If you want to determine whether the XML document is modified, so as to compare two XML documents are equal, then compare either directly (for example, a test package), or for security purpose on the comparison to the digital signature, will bring some inconvenient to do so. Through the W3C XML Canonicalization specification (c14n) to solve this problem, the specification defines the standard form of XML documents, can guarantee to provide the correct comparison, so as to ensure the consistency of the digital signature.

Tag standardization for which white space should be in the label application specific rules, a specific order of namespace declarations and neat properties. Below is my summary of establishing standardized start tag informal steps:

1. The left Angle bracket (<), followed by the element QName (prefix plus colon and local name).
2. If you have a default namespace declaration behind, and then the other namespace declarations, according to the defined prefix in alphabetical order. Omit all redundant namespace declarations (already in the ancestor element declaration and hasn't been rewritten namespace). There is a space between each namespace declaration, the equality and contain no Spaces on both sides of namespace URI.
3. All of the attributes in alphabetical order, there is a space in front of the equal sign on both sides of the double quotes, contains the value of an attribute with no Spaces.
4. The last is the right Angle brackets (>).

Acknowledgement

This research was financially supported by the Guizhou province science and technology fund projects, Guizhou LH[2014]7438.and The cooperation unit “Digital engineering center of local ethnic minorities culture, Qiannan Normal College for Nationalities. Duyun 558000 China“

This research was financially supported by the Guizhou province department of project: heterogeneous data sharing for small and medium-sized businesses in guizhou;NO[2011]2215

References

- [1] Johan List,Vojkan Mihajlović,Georgina RamÍrez,Arjen P. Vries,Djoerd Hiemstra,Henk Ernst Blok. TIJAH: Embracing IR Methods in XML Databases[J]. Information Retrieval . 2005 (4)
- [2] Jovan Pehcevski,James A. Thom,Anne-Marie Vercoustre. Hybrid XML Retrieval: Combining Information Retrieval and a Native XML Database[J]. Information Retrieval . 2005 (4)
- [3] Jaap Kamps,Maarten de Rijke,Börkur Sigurbjörnsson. The Importance of Length Normalization for XML Retrieval[J]. Information Retrieval . 2005 (4)
- [4] Benjamin Piwowarski,Patrick Gallinari. A Bayesian Framework for XML Information Retrieval: Searching and Learning with the INEX Collection[J]. Information Retrieval . 2005 (4)