Scientific
Research
Publishing

Research on XBRL Domain Ontology Construction

Ding Pan, Di Zhang

Management College, Jinan University, Guangzhou, China
Email: tpanding@jnu.edu.cn, 1415334965@qq.com

Abstract

XBRL (eXtensible Business Reporting Language) as an application of XML (eXtensible Markup Language) technology in the field of business reporting, uses information technology to add tags for the financial reporting metadata, to achieve unstructured data processing effectively. At first, this paper analyzes the XBRL technology framework, finds out the concepts and relationships between them, and accordingly designs the XBRL domain ontology model, then based on the model we have proposed, we choose OWL DL ontology description language and Protégé3.4, take balance sheet as an example, and apply the XBRL ontology model to balance sheet to construct the taxonomy and instance, and thereby achieve the implementation of XBRL domain ontology construction. XBRL needs the integration of the accounting, engineering, computer, management and other disciplines, and the future research should be committed to solve the lack of semantic and reasoning mechanism to implement the further promotion and application of XBRL financial reporting.

1. Introduction

With the rapid development of information technology in recent years, the disclosure of financial information on the Internet gets more and more enterprises concerns, the format of internet financial reporting will become an inevitable development trend in the accounting field. Due to the fact that the format and content of traditional financial reports are different, there will be many restrictions to the acquisition, analysis and storage of financial information [1].

XBRL is an Internet-based, cross-platform computer common language which is designed especially for the

organization, disclosure and usage of financial reporting. It provides standardized definition and expression for financial data, and becomes the latest widely accepted standards and technologies [2]. Since Charles Hoffman put forward the concept of XBRL in 1998, it has gained worldwide rapid development. The US Securities and Exchange Commission (SEC) in February 2005 began to encourage listing companies to submit financial reporting on Internet voluntarily in XBRL format [3]; Canadian Securities Regulatory Commission began to implement XBRL voluntary filing program to supplement for PDF report; German bank carries out a plan to equip BARS with XBRL data input interface; New Zealand XBRL organization initiates the taxonomy making plan; Japan, Spain and other countries also launch researches on XBRL implementation one after another. The Shanghai Stock Exchange officially becomes a member of the XBRL international organization in October 2005, listing companies in Shanghai and Shenzhen are now required to use XBRL format to disclose financial reports regularly. With the application of XBRL in the accounting, auditing, securities and banking, its advantages of information processing are gradually recognized by the world, and bring great convenience to the organization, analysis and communication of financial information [4].

2. XBRL Technology Framework

XBRL technology framework mainly consists of three parts: the XBRL specification, XBRL taxonomy, XBRL instance, as shown in **Figure 1**. XBRL specification is XBRL technology master, it defines XBRL working mechanism and grammar rules, illustrates how to make taxonomy and instance according to the relevant business norms. XBRL taxonomy is a collection of accounting concepts and relationships equivalently to a terminological dictionary which is the core part of the XBRL technology. XBRL taxonomy acts as the basis for generation and interpretation of instance documents, and is established according to the accounting standards, laws and supervise provision in different countries, industries and enterprises. XBRL instance is a set of truth values in financial reporting, the fact data in instance and defined concepts in taxonomy are corresponding to each other, the application will automatically extract data from the accounting business database and generate the instance documents [5]. XBRL specification still has two optional modules: dimensions and formulas, which are used to describe the multidimensional information and cross-background calculation rules of financial data. In addition, various industries or enterprises can make an extend to the reference taxonomy to adapt to their own information disclosure preference.

The XBRL taxonomy consists of a taxonomy schema (.xsd) file and five linking library files. The scheme file defines structure and content of taxonomy, and includes the attributes of element name, data type, loan balances and event type. The elements are divided into two kinds: data items and tuples. Linking library files including label linkbase, presentation linkbase, calculation linkbase，definition linkbase, reference linkbase, describe the relationships between domain concepts using XLink and XPointer technologies, and link to the human readable documents. Label linkbase using arc characters combines human readable labels with concepts; presentation linkbase describes concept hierarchy using arc characters; calculation linkbase defines concept relationships from data calculation perspective, for example, "after-tax profits" can be defined as "pre-tax profits" minusing "income tax"; definition linkbase illustrates other relationships, such as generic and instantiation; reference linkbase defines the reference information, and points to source files by recognizing name and relevant paragraphs.

3. Ontology Concepts

British physicist Berners-Lee proposed the concept of semantic web in 1998, and pointed out that the semantic web mainly adds metadata to the data item on the Internet to make the Internet become a common information exchange medium. The architecture of the semantic web is divided into seven layers: code, grammar, resource
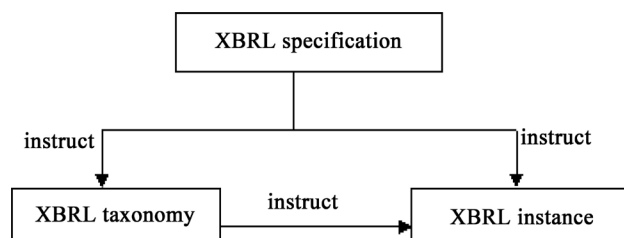

Figure 1. XBRL technology framework.

description, ontology, logic, checking and trust, the function increases gradually from bottom to top. The semantic web uses ontology as the knowledge representation model on the fourth floor of the architecture, because its powerful semantic reveal ability can provide more accurate semantic standard for accounting, auditing, and other business fields [6].

Ontology is a concept in philosophy originally, it explores the world of the "primitive" or "matrix". Since the 1990s, ontology has been applied to the computer field, in order to express and reuse the standardized knowledge, then it has obtained the in-depth research in artificial intelligence and database technology as well. Ontology still hasn't a clear definition in the academic circles, the more authoritative concept is proposed by Gruber in Stanford University in 1993: Ontology is a displaying and modeling demonstration of sharing concepts [7]. This definition mainly includes four levels: conceptual model, explicit, formalization and sharing. "Conceptual model" refers to the abstract model of the objective world phenomenon, "explicit" refers to the precise definition of concepts and relationships between them, "formalization" means that ontology is computer readable, "sharing" means that ontology expresses the widely accepted knowledge and reflects a set of sharing concepts in the related field.

Ontology actually is an entity, it captures knowledge of related fields and translates the knowledge into a set of concepts and relationships, this process is just the application of the ontology modeling method. Generally speaking, an ontology contains five basic modeling primitives: classes, relations, functions, axioms and instances, classes are also written in concepts.

Classes: collections, concepts, or object types.

Relations: relations represent the interaction between the domain classes. They are defined as a subset of n-dimensional Cartesian product: R: $C1 \times C2 \times \ldots \times Cn$.

Functions: functions are kinds of special relations. The nth element is determined by the frontal $n - 1$ elements. The formal definition is: F: $C1 \times C2 \times \ldots \times Cn-1 \rightarrow Cn$. Such as mother-of, mother-of (x, y) implies that y is the mother of x, and x can only determine its mother y.

Axioms: axioms represent true assertions, such as concept A belonging to concept B.

Instances: instances represent the elements.

According to the research level, it can be divided into four types: top-level ontology, shows the universal senses and does not depend on the particular problems or fields, such as space, time, object, behavior; domain ontology, provides jargons or vocabularies of specific area, such as aircraft manufacturing, agriculture and medicine; task ontology, defines the generic tasks and reasoning activities, such as problem judgment; application ontology, relies on both particular field and theme.

There are four kinds of ontology basic relationships, which include part-of, kind-of, attribute-of, instance-of. In the specific construction process, we can also define other relationships according to the actual situation, These relationships are usually divided into two categories, one kind is the relationship between the different logic concept levels, such as "Is-a relation", "instance-of relation", the other one is to reflect the relationships between part and whole.

Ontology languages provide users with the clear and formal description of domain concepts, and it is necessary to meet the requirements of well-defined syntax, well-defined semantics, efficient reasoning support, sufficient expressive power, convenience of expression. The existing ontology languages are divided into two categories, one is related to Web: RDF, RDF-S, OIL, DAML, OWL; the other one is related to the specific system: CycL, Ontolingua, Loom. Web Ontology Language (OWL) is an ontology description language standard in semantic web recommended by W3C, it is on the top of the W3C ontology language stack, derives from DAML + OML, and has higher semantic expression ability than Resource Description Frame (RDF). OWL contains three sub languages: OWL Lite, OWL DL, OWL FULL, users can select appropriate language according to practical situation. OWL Lite is used to provide only one classification hierarchy and simple attribute constraint for users, and supports cardinality, which only allows 0 or 1. OWL DL provides maximum expression on reasoning system for users. The reasoning system here can guarantee the computational completeness and decidability. It includes all the constraints of the OWL Lite, but can be put under specific constraints. OWL Full supports the need to maximize the expression of RDF without calculation guaranteed. It allows to increase the vocabulary to a predefined ontology, so that any reasoning software does not support all the OWL Full features.

4. Construction of XBRL Domain Ontology

At present there are lots of methods of ontology construction, such as the skeleton measure by Mike Uschold &

King, evaluation method by Gruninger & Fox, engineering method by KACTUS, the methodology method, sensus method etc. This paper draws lessons from the above-mentioned construction methods and proposes the construction process of XBRL domain ontology: 1) to determine the domain and the range of the body; 2) to consider the reuse of existing ontology; 3) to list important terms in ontology; 4) to define class and hierarchy; 5) to define the class attributes; 6) to create entities; 7) to conduct the ontology evaluation.

There are a lot of ontology development tools, this paper chooses Protégé3.4. Protégé is an ontology editing and knowledge acquisition software based on Java language and developed by Stanford University, it is independent of the particular language and can import and export different language format, such as OWL, XML, RDF (s). In the practical application, Protégé has four important tabs: class, data properties, object properties, individuals, to realize the construction of ontology.

This paper analyses the XBRL technology framework, especially the conceptual structure of taxonomy and instance, abstracts the domain concepts and their relationships, then accordingly designs the domain ontology model of XBRL [8], as shown in **Figure 2**.
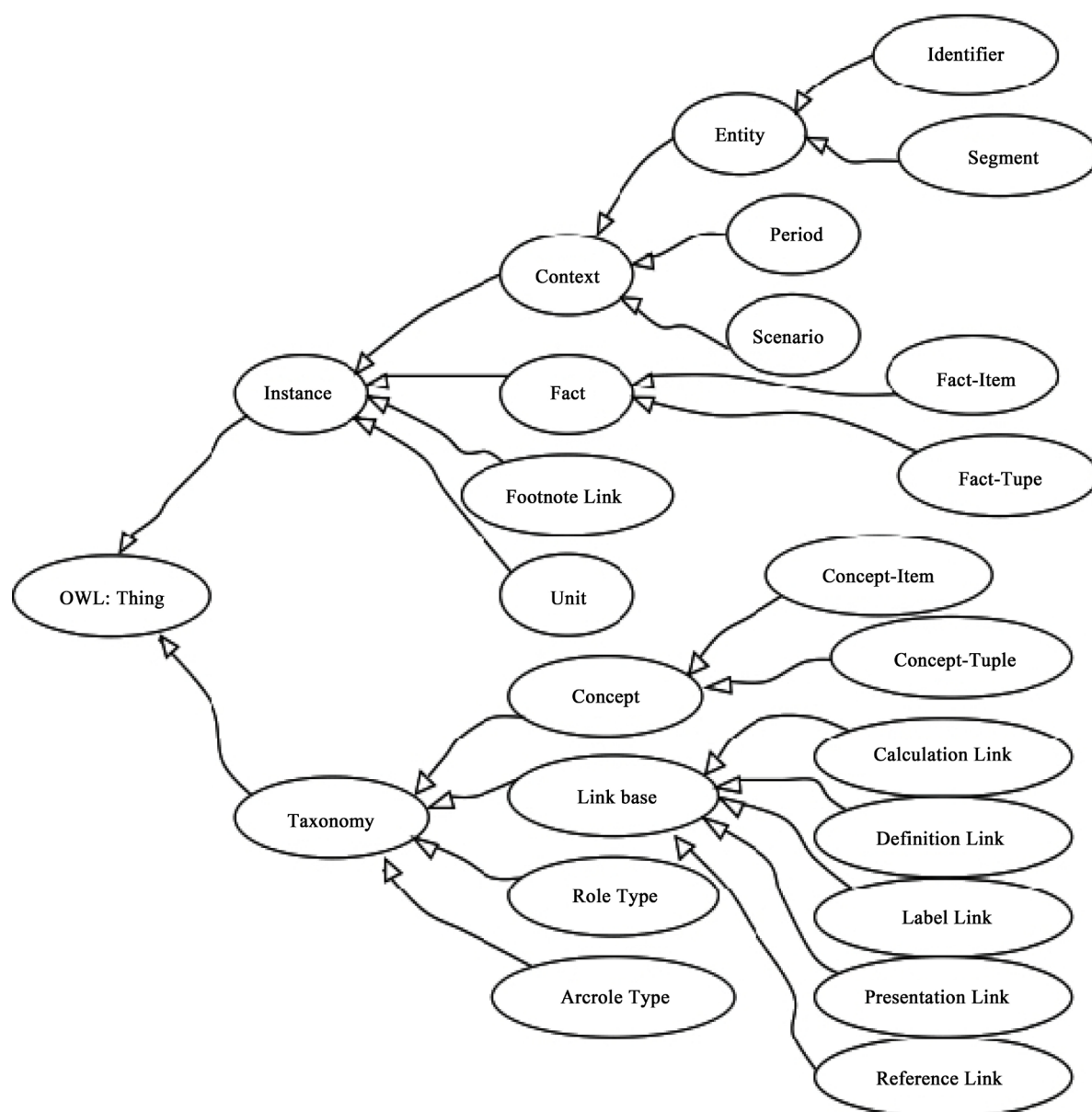

Figure 2. XBRL domain ontology model.

In the construction process, we choose OWL DL, and according to XBRL ontology model, we use class tab in Protégé to implement concept class construction, add data attributes for class in data properties tab, add object attributes in object properties tab, and then select the balance sheet as an example, add individuals for Concept-Item in individuals tag, generate the instance document, as shown in **Figure 3** and **Figure 4**.
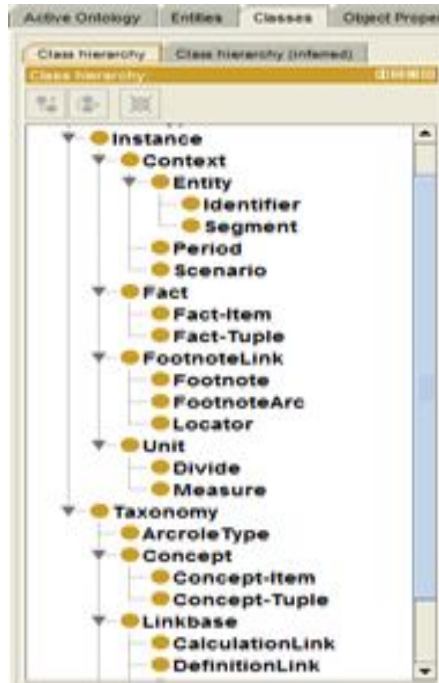

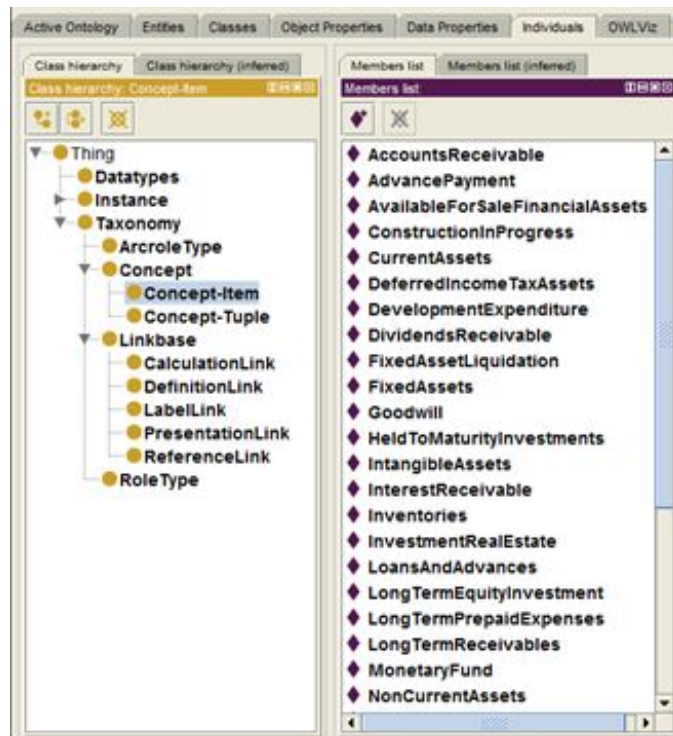**Figure 3.** Conceptual class interface.

Figure 4. Individuals interface.

5. Conclusions

XBRL domain ontology is a formal description of the entities, attributes, processes and relationships in financial reports, providing the sharing and reusing components for each system. Due to this structured expression, it also makes the reasoning mechanism possible through adding rules, and acts as the base of knowledge index and retrieval in financial reports [9]. At present, most databases use keyword and catalog retrieval technology, ontology can greatly improve the recall and precision by implementing semantic annotation for business reports as a conceptual model of knowledge sharing in the semantic level. In addition, XBRL ontology schema corresponds to a vocabulary table for information exchange in different systems, and will realize semantic integration and inter operation between heterogeneous systems.

Making good use of XBRL technology to disclose the financial reports, product and transmit business information, is an inevitable choice to participate in international competition, it shows the information direction of financial reports development under the semantic web environment. We must build a solid foundation of ontology to make XBRL become real formalization business language. XBRL ontology construction is a system engineering which calls for integration of accounting, engineering, computer, management and other disciplines. The future research should be dedicated to solve the lack of semantic and reasoning mechanism, and make further promotion and application of XBRL financial reports.

References

[1] Debreceny, R. and Gray, G.L. (2001) The Production and Use of Semantically Rich Accounting Reports on the Internet XML and XBRL. *International Journal of Accounting Information Systems*, **1**, 47-74.
http://dx.doi.org/10.1016/S1467-0895(00)00012-9

[2] Liu, B.Q. (2003) XBRL: A Convenient Internet Financial Report Language. *China Accounting Computerization*, **6**, 18-20.

[3] Pan, D. and Pan, Y.S. (2011) A Study on the Semantic Formalization of XBRL Financial Report. *Accounting Information Professional Committee of China Accounting Society*, **6**.

[4] Qu, J.L. (2004) XBRL and Its Impact on Financial Reporting. *Monthly Finance and Accounting*, **12**, 73.

[5] Zhang, T.X. (2006) Network Financial Report: The Theoretical Research on XBRL Standard. *Accounting Research*, **9**, 56-63.

[6] Lara, R., Cantador, I. and Castells, P. (2007) Semantic Web Technologies for the Financial Domain. Vol. 6, Springer, US, 41-74.

[7] Spies, M. (2010) An Ontology Modelling Perspective on Business Reporting. *Information Systems*, **35**, 404-416.
http://dx.doi.org/10.1016/j.is.2008.12.003

[8] Ying, W., Wang, D. and Huang, M. (2013) Research on the Architecture Model of XBRL Financial Report Taxonomy. *Accounting Research*, **8**, 3-9.

[9] Wang, Y.J. and Huang, R. (2011) Analysis on the Application of XBRL Technology in China. *Business Accounting*, **3**, 73-74.