

Literature Review of XBRL Semantic Research

Hongqin Li, Jun Zhai

Transportation Management College, Dalian Maritime University

Dalian, China

weiqin0725@126.com

Abstract—Development of business intelligence cannot leave XBRL. In order to exert the potential of XBRL, semantics must be added to realize the combination of XBRL and semantic web. It's important to understand the research status of domestic XBRL semantic research. So the article takes XBRL as inscription and contains ontology for retrieving conditions, searches China journal full-text database (CNKI). At last the article selects the representative journals, conferences and research papers to make the literature review of XBRL semantic research in China. Then the article introduces the abroad newest progress of XBRL semantic research, causing the thinking of XBRL semantic research trend. Finally the article puts forward the measures for promoting the development of the XBRL semantics research. This paper has certain significance for XBRL semantic research.

Keywords—XBRL; Semantic Web; Ontology; Linked Data

I. INTRODUCTION

A. XBRL Basic Conception

Since 1998, the American CPA, Charles Hoffman proposed XBRL (eXtensible Business Reporting Language), XBRL has achieved rapid development. So far XBRL has been used in more than 570 projects in 27 countries around the world. In China the study of XBRL standard began in 2002. In 2004 Shanghai Stock Exchange selected 50 Shanghai local listed companies as XBRL report pilot. Since 2008, Shanghai Stock Exchange required all listed companies to submit XBRL reports and disclose on its website for users. Shenzhen Stock Exchange started the XBRL pilot in February 2005 and in February 2009 required all listed companies to submit annual report using XBRL instance document, and provide the download function of instance documents on its website. But the taxonomy of the two Stock Exchange is inconsistent, and thus causes a hotspot in the follow-up research.

XBRL is a language based on XML, aimed at promoting the release, exchange, and analysis of the information the financial statements contain. However XBRL and XML has some limitations—lack of formal semantics and reasoning mechanism. Semantics and reasoning is very useful to solve business problems in the use of XBRL financial statements. So if you want to play the potential of XBRL, XBRL must be combined with semantics. In fact, the proponents of XBRL, Charles Hoffman also has realized the importance of XBRL semantic research, and conducted the research of FRO (Financial Report Ontology) development in the field of accounting. In its prototype websites, the reasons of

developing FRO is pointed out, that is, the higher the semantic level is, the computer can do more for the user. So what is the semantic level of structure? Fig.1 answers this question.

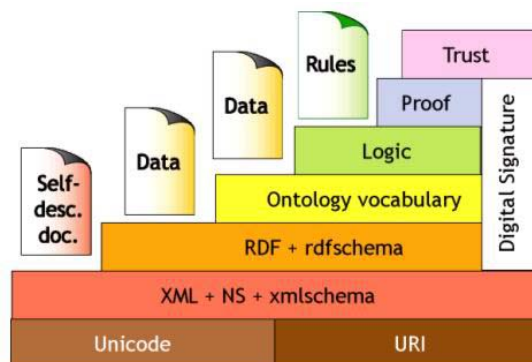


Fig. 1. semantic web architecture

B. Semantic Web

In view of the difficulty of current web data mining and limited to display information for specific domain, the semantics of the display information will not be able to understand. In 1999 the founder of the World Wide Web, Tim Berners-lee formally put forward the concept of semantic web. Semantic web makes computer understand information through the reconstructed network by mixing its meaning in the information.

Fig.1 contains seven layers, the semantic web architecture provides the technical support for realizing the semantic web vision through a bottom-up hierarchical functional extension. The first layer is the foundation of semantic web. Semantic web realizes the identification of network resources via URI (uniform resource locator). The Second layer realizes the separation of structure and contents of network information through XML, and XBRL is implemented based on XML, so financial statements using XBRL can meet the requirements of this layer. In order to implement XBRL semantics, one should be on the basis of this layer to explore. The core technology of the third layer is RDF (Resource Description Framework). RDF is a data model about the resource and the relationship between resources. RDF provides simple semantics for the data model. The fourth floor Ontology reveals resource and the complex and rich semantic information between resources. In order to realize semantics we should use ontology development language—OWL(Web Ontology Language). Ontology plays a core role in the semantic web. So in the literature retrieval, Ontology is as a search term. The fifth to

seventh layer is mainly responsible for the realization of the rule and reasoning mechanism.

II. LITERATURE REVIEW OF XBRL SEMANTIC RESEARCH IN CHINA

Before this, there are a number of scholars who have summarized XBRL research status. Minglai Zhuang et al[1] reviewed the XBRL research during 2001-2010 in China. This paper divided XBRL research into four aspects: the XBRL taxonomy and accounting application research, XBRL application research in corporate governance and audit, XBRL application research and promotion strategy in financial analysis and process reengineering, and other aspects. Yi He[2] got the XBRL research about education, auditing, internal control, security and the direction of empirical analysis. Above two articles both reviews from the angle of the overall development of XBRL. With the in-depth development of XBRL research, related research literatures emerge in endlessly. All aspects of XBRL research are also more and more active, the XBRL progress in one area of research are reviewed. Kun Gao[3] reviewed the domestic research on XBRL taxonomy from the taxonomy design, the participation main body, evaluation system in recent years. Gu Qing et al[4] generalized the audit issues research under XBRL environment from 2004 to November 2013. There is no corresponding literature review in XBRL semantic research.

The author has searched China journal full-text database based on the theme of "XBRL" and got 2205 results. On this basis, "ontology" is as keyword to search the results, after further reads the results, and finally selects 35 representative journals, conference, and academic theses. The classification of the literature is showed in Fig.2 and Fig.3.

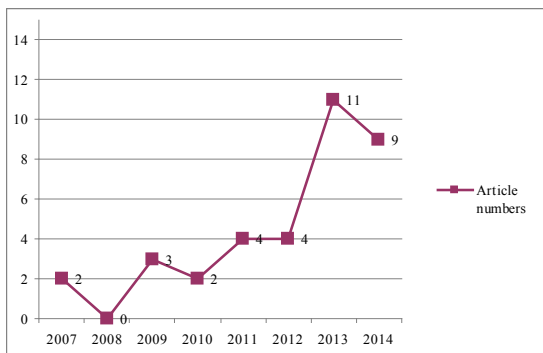


Fig.2. according to the year summary of search result of XBRL as theme and ontology as keyword

From Fig.2, we can see that in China XBRL semantic research starts late and has a big progress in 2013. In 2007 Jinping Gao first mentioned ontology in doctoral dissertations, and based on the ontology to construct the theory of XBRL financial reporting taxonomy. Another paper is Changchuan Liu's master degree thesis, this paper draw ontology basis that the computer could understand accounting semantics, and put forward an innovative view: how to make computer understand and learn the integrated use of accounting, finance, securities investment, cost management and auditing. This view

coincides with the current development trend of foreign XBRL semantics.

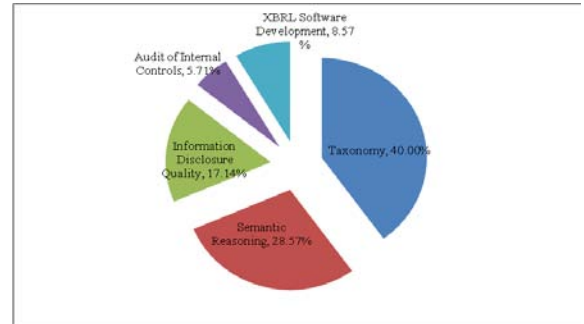


Fig.3. according to the research content summary of search result of XBRL as theme and ontology as keyword

As shown in Fig.3, in China the research contents of the taxonomy, semantic reasoning and information disclosure quality research are more, in respect of XBRL software development and the internal control auditing study are less. This illustrates that XBRL application has not been popular, potential users such as regulators, enterprise management and the investing public are not familiar with XBRL. The following generalizes from five aspects: "taxonomy", "semantic reasoning", "information disclosure quality", "internal control auditing" and "XBRL software development".

A. Taxonomy

The taxonomy research can be divided into three categories: taxonomy design (8 articles), taxonomy mapping (3 articles), and XBRL GL taxonomy design (3 articles). Taxonomy design refers to the evaluation and quality improvement of taxonomy, taxonomy extension of the industry and enterprise, taxonomy authentication. Taxonomy mapping is in order to solve the problem of interoperability between taxonomy. As mentioned earlier, Shanghai Stock Exchange and Shenzhen Stock Exchange uses different taxonomy, in China, at present there have three taxonomies: CAS (general taxonomy of China), the SSE (Shanghai Stock Exchange taxonomy) and SZSE (Shenzhen Stock Exchange taxonomy). An enterprise may need to submit XBRL instance document in accordance with different taxonomy at the same time. This leads the study of taxonomy mapping. Research on XBRL GL, first of all to be clear, according to the regulation of XBRL international organization, XBRL is divided into XBRL FR (XBRL Financial Reporting) and XBRL GL (XBRL Global Ledger), XBRL GL corresponding to XBRL FR, is mainly used to record companies daily trading, as the standard of enterprise internal report. At present the XBRL taxonomy in most literature actually refers to XBRL FR taxonomy. In China XBRL GL hasn't yet to have a specific taxonomy.

- First, taxonomy design

In terms of taxonomy design, Jinping Gao(2007), Zhengzheng Li(2013) and Xiao Pan(2013) made efforts in the quality improvement of taxonomy. Jinping Gao[5] pointed out that the information elements defined in XBRL financial reporting

taxonomy had differences with the rules of accounting, information disclosure and firm's preferred reporting practices. These differences would have economic consequences. So this paper used the idea of ontology to build the theory frame of XBRL taxonomy.

Changyin Huang(2012), Shuo Wang(2013) made efforts in taxonomy extension of the industry and enterprise. Changyin Huang[6] pointed out that the structure of taxonomy was the application of hierarchical structure of ontology in the field of XBRL financial reporting. The hierarchical structure of ontology from high to low was: meta ontology, top ontology and domain ontology. So the corresponding structure of taxonomy included general taxonomy and extended taxonomy. According to the characteristics of hierarchical extension, extended taxonomy was divided into general level, industry level and enterprise level.

Feng Liu(2012), Fan Sun(2012) explored the XBRL model including XBRL taxonomy. Feng Liu[7] built the XBRL technical model on semantic web basis and made use of the idea, system architecture and implementation techniques of semantic web to reconstruct specification, taxonomy and instance documents. Fan Sun[8] analyzed the lack of the XBRL technical architecture in semantic formalization and pointed out that the preparing of the instance documents should adopt the method of ontology engineering in order to implement the ontology of the instance documents and make the instance documents for computer automated processing.

Zhounan Yang et al[9] tried to establish the methodology system of XBRL taxonomy recognition based on the theory of the ontology.

- Second, taxonomy mapping

In terms of taxonomy mapping, Hui Fu[10] compared the architecture of CAS(2010) taxonomy and IFRS(2010) taxonomy, put forward the differences between the two standards, and then regarded taxonomy as a specific domain ontology to construct the ontology solving these differences, which provided a feasible scheme for automated mapping. Jimei Li et al[11] designed the algorithms of determining and resolving concept heterogeneity between XBRL taxonomies on the analysis of taxonomy heterogeneous formation and expression of ontology. Zhongsheng Wu et al[12] proposed unified financial report domain ontology and constructed XBRL financial report transformation model to solve taxonomy mapping.

- Third, XBRL GL taxonomy design

In terms of XBRL GL taxonomy design, Xinzi Hao[13] built the XBRL GL data model of three layers structure with the aid of ontology: the underlying ontology recorded the actual native information of firm's trading business, middle for domain-specific

ontology such as accounting ontology, the upper for a particular purpose such as XBRL FR. Zhongsheng Wu et al[14] established a unified XBRL ontology model on the supply chain. The process of construction mainly included building the global ontology, local ontology (XBRL FR and XBRL GL), defining the mapping (global ontology concepts and their relations and local ontology concepts and their relations).

B. Semantic Reasoning

From the architecture of the semantic web, we can see, in order to do reasoning, we must first to construct the ontology. Yongjian Li[15] built the XBRL meta model, and transformed the meta model into ontology, then added rules so as to realize the semantic reasoning. Changchuan Liu[16] put forward how to make computer understand and learn the integrated use of accounting, finance, securities investment, cost management and auditing. Ding Pan et al[17] pointed out that the lack of semantics restricted the thorough analysis of XBRL financial information. This paper discussed the semantic integration method of XBRL and BI based on ontology. Zhanying Huang[18] pointed out the importance of BI based on semantic web in financial risk management, internationalization and management by the European commission comprehensive project MUSING. Meijie Du et al[19] pointed out that on the basis of semantic web, computer could understand the meaning of XBRL documents, query, analyze, mining, automatically process the preparation and transformation of business report. Yong Wang[20] used XSD2OWL Translator to transform XBRL schema file into ontology, XML2RDF Translator to transform XBRL instance documents into RDF file, and then used RDF query technology to realize the integration of XBRL reports, providing the basis for financial information exchange all over the world.

C. Information Disclosure Quality

In terms of information disclosure quality, Xiaolan Zhong [21] extracted the undefined elements in taxonomy based on the matching result of Shanghai Stock Exchange XBRL taxonomy and 2010 annual report of listed companies on the basis of ontology, so as to reduce the differences between XBRL taxonomy and information disclosure requirements of financial report. Wen Liu[22] evaluated the quality of accounting information from the perspective of taxonomy extension, and used the seven steps of ontology theory to analyze the company's PDF annual report, compared with general taxonomy, then proposed taxonomy extension methods according to the seven steps.

D. XBRL Software Development

In XBRL software development, ZhenQing Yang[23] developed intelligent financial analysis system through the establishment of concept ontology, analysis model ontology, analysis indicators ontology and resource ontology to realize the classification of financial statement analysis, specific financial indicators analysis and comprehensive financial capacity analysis.

E. Internal Control Auditing

There are only two articles in the internal control auditing. Li Dai[24], discussed the method of the internal control auditing under XBRL environment, used the seven steps of ontology theory to design the internal control auditing taxonomy under XBRL environment.

III. XBRL SEMANTIC RESEARCH PROGRESS IN OTHER COUNTRIES

About XBRL semantics research in other countries, the author mainly spreads from taxonomy mapping and semantic reasoning, emphasizes on the latest research results.

A. Taxonomy Mapping

Susan Marie Thomas et al[25] realized different multilingual jurisdictional taxonomies were mapped by finding correspondences between concepts based on the diversity of languages and XBRL taxonomy. The taxonomy mapping process includes four stages: Phase 1 automatically converts each XBRL taxonomy into RDF by using MONNET xbl2rdf converter, then converts each XBRL concept into ontology by using SPARQL query; In Phase 2, the ontology is manually enriched and clarified. The enriched ontology is input to Phase 3, which automatically computes cross-taxonomy equality and subsumption relationships by means of logical reasoning services. Phase 4 is mapping verify, it uses recall and precision as evaluation index. The process is tested on pairs of taxonomies but could also be used for more than two.

Relative to the domestic research in China, the study abroad focuses on the borders and area problems of financial data integration, and the research in China is limited to realize the three (CAS, SSE and SZSE) taxonomies mapping. In addition, on the evaluation index, the current domestic index is too simple compared with the precision and recall. With the continuous development of XBRL, more and more organizations around the world look to XBRL data, hope to collect and share business information with the help of XBRL, and implement data integration in a multinational company, industry or a higher level of business decisions. But it is not easy to achieve the goal, the lack of semantics reduces the comparability of XBRL. XBRL organization also admits that, in response, the organization established the comparability group. The task of the group is responsible for the collection of demand of regulators, listed companies and other stakeholders. The group shows the relationship between the elements set in the form of XBRL assertions by offering a taxonomy mapping. It is enough to see the importance of taxonomy mapping.

B. Semantic Reasoning

XBRL semantic reasoning progresses rapidly in foreign country and a lot of practical systems have developed:

Romilla Chowdhuri et al[26] built OFXD (Ontology-based Framework for XBRL-mapping and Decision-making). The framework automatically built XBRLOnt Ontology by using the information extracted from the XBRL financial statements, so as to realize the interoperability between XBRL files

meeting different taxonomy. On the basis of this the framework calculated the financial indicators: profitability, financial leverage, liquidity and operating ability by mapping the semantic heterogeneous XBRL element. The framework provided a useful starting point for using the XBRL potential to benefit the individual investors and institutional investors. At the same time, the article pointed out the potential to integrate XBRL and non XBRL data. Such a model would allow investors to compare companies not only based on financial metrics, but also on other characteristics such as governance scores, stakeholder assessments, market shares. In fact, now there are a lot of similar studies.

The main method of integrating XBRL and non XBRL data is linked data. In 2006, Tim Berners-lee proposed linked data is the core and key of data of web. Linked Data is the lightweight implementation of semantic web. Sean O'Riain et al[27] proposed the idea of building global financial ecosystem throughout the business activities and business report activities of the supply chain by using linked data to integrate the XBRL and non XBRL data. Here the non XBRL data is mainly refers to the open data, including industry information provided by DBpedia, company data provided by Crunchbase, market data and financial news provided by Bloomberg and the New York Times, location from GeoNames; company categories provided by Dublin core. Benedikt Kampgen et al[28] built FIOS (Financial Information Observation System) based on RDF data cube vocabulary to convert a variety of data format into RDF, then based on the linked data in the form of RDF to analyze. FIOS helped the analysts to create their own analysis similar to Excel function through the integration of the annual and quarterly balance sheet, the daily stock quotes, and the company and industry background information.

IV. REVELATION

Through literature review of XBRL semantic research, the author puts forward the following suggestions of the development of XBRL semantic research in China.

A. Increasing the open degree of XBRL Instance Document

In the literature reading, the author finds that the foreign research all has the experimental process by using the data from official sources, such as EDGAR. Data is important that research is carried out successfully and experimental results can be effectively verified. But, in China current XBRL reports is not accessible, the Shanghai Stock Exchange does not provide download of XBRL instance document. This causes great obstacles for the scholars to research on XBRL.

B. Strengthening XBRL Personnel Training

From analysis we can see that XBRL semantic research has certain characteristics of school, which is associated with the technical requirements of XBRL semantic research. In order to speed up the XBRL semantic study, we must vigorously strengthen XBRL personnel training. During the development of XBRL in China, quite a number of accounting personnel join the XBRL learning team. But in the domestic teaching of undergraduate course phase, there is no related

course of XBRL. The students can only grasp the XBRL reporting technology by practical training in the future. Therefore, we should increase the XBRL personnel training, to improve the recognition about XBRL of highly educated professional researchers, to facilitate the further development of research on XBRL semantics.

C. Improving the XBRL Semantic Research Level

At present in the aspect of semantic reasoning, the foreign countries focus on the integration of the XBRL and non XBRL data by using linked data. The scope of research is expanded to include company background information, stock quote information of listed companies as well as the financial news and so on. The goal is to help analysts more efficiently to make decision. But so far, in China research on XBRL is still limited to information on how to fully use of XBRL, and has not yet been extended to combine with non XBRL data. Of course, it is related to the development of linked data in China. If we want to keep up with the pace of study in foreign countries, we also need to use linked data to develop more data sources available.

D. Jointing Experts in Various Field to develop XBRL Semantic Research

As mentioned above, the semantic research of XBRL is difficult. In order to add semantics to XBRL, first of all, we must be familiar with semantic web and the concept of ontology. Susan Marie Thomas et al(2014) pointed out that for most accountants ontology construction tools (such as a Protege) is very strange, so in the experiments, the concept is described with the help of accountants, not directly by the accountants. There are language problems of different taxonomy mapping, such as France accountants is not easy to explain Spain's balance sheet, even if it has a label in English. So the independent accountants in Spain and in France describe their own balance sheets, this can lead to a faster progress, and have better precision. It is obvious that the progress of XBRL semantic research need the union of experts in various fields, including each country, each region, each industry (computer science, accounting, regulatory agencies) to exert the maximum potential of XBRL.

REFERENCES

- [1] Minglai Zhuang, Miaomiao Wei. The Review during Ten Years on XBRL Research in Our Country:2001-2010. Friends of the Accounting, No. 21. pp.18-22,2011.
- [2] Yi He. The Review of XBRL Domestic Research. Friends of the Accounting, No. 4. pp.48-52,2012.
- [3] Kun Gao. The Relevant Literature Review of XBRL Taxonomy in China. Oriental Enterprise Culture. No. 2. pp.199,2013.
- [4] Gu Qing, Chaoqun Xin. The Literature Review of the audit issues research under XBRL environment. Friends of the Accounting, No.1. pp.115-118,2015.
- [5] Jinping Gao. Research on XBRL Financial Reporting Taxonomy: Quality, Economic Consequence, and Improvement. Shanghai: Shanghai Jiao Tong University, 2007.
- [6] Changyin Huang. Research on Hierarchical Extension of XBRL Financial Reporting Taxonomy. Shanghai: Shanghai Jiao Tong University, 2012.
- [7] Feng Liu. The XBRL Technical Model on Semantic Web Basis and The Application Study. Beijing: Research Institute for Fiscal Science, Ministry of Finance, P.R.China, 2012.
- [8] Fan Sun, Zhounan Yang. Analysis and Improvement of XBRL Technical Architecture Based on Linguistics. Accounting Research, No. 7. pp.13-19,2013.
- [9] Zhounan Yang, Jianguo Zhu, Feng Liu, Ying Sun, Yingji Zhao. Research on the Theory Basis and Methodology System of XBRL Taxonomy Recognition. Accounting Research, No. 11. pp.10-15,2010.
- [10] Hui Fu. The CAS-Taxonomy and IFRS-Taxonomy Mapping Study. Xiangtan: Xiangtan University,2012.
- [11] Jimei Li, Meijie Du. Information Integration algorithm of Heterogeneous XBRL Financial Reporting. Journal of Jilin University(Engineering and Technology Edition),Vol.42,pp.266-270,2012.
- [12] Zhongsheng Wu, Tianxi Zhang, Zhide Chen, Wei Du. Research on Conversion between XBRL Financial Reports based on Domain Ontology. Application Research of Computers Vol.30,pp.3643-3651,2013.
- [13] Xinzi Hao. The Study on Constructing a XBRL GL Application Framework in China. Beijing: Capital University of Economics and Business,2014.
- [14] Zhongsheng Wu, Tianxi Zhang, Zhengzheng Li. The XBRL Technology Research of Business Report under the Perspective of Supply Chain. Modern Management Science, No. 7. pp. 91-93,2013.
- [15] Yongjian Li. Ontology and Reasoning of the XBRL-in Case of Financial Information. Guangzhou: Jinan University,2014.
- [16] Changchuan Liu. The Logic System of Accounting based on Computationalism. Changsha: Hunan University,2007.
- [17] Ding Pan, Yunshan Pan. Incorporating XBRL into Business Intelligence Applications based on Formal Semantics. China Academic Accounting Association Annual Meeting, 2011.
- [18] Zhanying Huang, Ding Pan, Dong Wang. The Application of XBRL and Business Intelligence. China Academic Accounting Association Annual Meeting, 2011.
- [19] Meijie Du, Jimei Li, Houzhou Zhao. The International General Commercial Language on Semantic Web Era-XBRL. Finance & Accounting, No.6. pp. 49-51,2013.
- [20] Yong Wang. The Data Integration Method of Semantic Heterogeneous XBRL based on Ontology. Software Guide, Vol.13,pp.127-129,2014.
- [21] Xiaolan Zhong. The Research on Differences between XBRL Taxonomy and Requirements of Enterprise Financial Reports Information Disclosure. Changsha: Hunan University,2013.
- [22] Wen Liu. Research on Hierarchical Extension of Company Information Disclosure based on XBRL Universal Taxonomies.Haerbin: Harbin Institute of Technology,2014.
- [23] Zhenqing Yang.The Research of XBRL Network Financial Statement Analysis System Modelling. Communication of Finance and Accounting, No.12. pp. 113-114,2009.
- [24] Li Dai. The Study on the Implementation Method of Internal Control Auditing under the XBRL Environment. Changsha: Hunan University,2014.
- [25] Susan Marie Thomas, Xichuan Wu, Yue Ma and Sean O'Riain. Semantically Assisted XBRL-Taxonomy Alignment Across Languages. Towards the Multilingual Semantic Web,2014, pp. 277-293.
- [26] Romilla Chowdhuri, Victoria Y. Yoon, Richard T. Redmond, Ugochukwu O. Etudo. Ontology based Integration of XBRL Filings for Financial Decision Making. Decision Support Systems, No.68. pp. 64-76,2014.
- [27] Seán O'Riain, Edward Curry, Andreas Harth. XBRL and Open Data for Global Financial Ecosystems:A Linked Data Approach. International Journal of Accounting Information Systems, No.13. pp. 141-162 ,2012.
- [28] Benedikt Kampgen, Tobias Weller, Sean O'Riain,Craig Weber, and Andreas Harth. Accepting the XBRL Challenge with Linked Data for Financial Data Integration.The Semantic Web: Trends and Challenges, vol. 8465, pp. 595-610,2014.