

Method and practice of comparing indicators from standards

Kenan Gong^{1,2}, Keqin Gan¹ and Aixian Li^{1,a}

¹China National Institute of Standardization, Beijing, 100191, China

²Research Institute of Information Technology, Tsinghua University, Beijing, 100084, China

Abstract. Chinese standards and foreign standards differ in technical indicators and requirements, which result in the poor recognition of products that comply with Chinese standards. It is crucial to set up a universal method to compare standard indicators, so that the inconsistency between standards could be eliminated, and ultimately improve Chinese products competitiveness. This paper proposed a method for indicator comparison from standards. Based on Semantic Web technology, the method introduced fragmentation of standard documents in a manner of 'product - stylistic rule - indicator' triplet. The processed fragments were annotated by the knowledge-based organization and used for the following indicator comparison. The method and procedure were generalized to develop standard indicator comparison system. The system provided a great experience in representing technical indicators as well as the comparison of indicators from standards. Further practice of the system suggested that repeated standard development and conflicted indicators in Chinese standards are still common.

1 Introduction

Standard is a vital technical component in international cooperation, playing an irreplaceable role as the pass and bridge connecting technological achievements and industry practice. With the intensification of global competition, the standard has become a strategic approach for enterprises to improve their international competitiveness. In response to the changes in the international market, and enhance enterprise strength in the meantime, Europe, United States, Japan, South Korea and other well-developed countries have adopted a variety of measures to promote adoption of their national or regional standards.

Comparison of standard indicators is to compare the technical indicators of the same or similar objects in different standards so that the technical differences between standards could be determined. The performance of indicators ensures products that comply with standards are safe, reliable and of good quality. It is noteworthy that the level of consistency between Chinese standards and international or foreign standards is still not seen as enough. By comparing the differences of indicators from standards, we can further improve the depth and competitiveness of Chinese standards. Therefore, research on the comparison of standard indicators has become a hot topic in Chinese industries.

Standards are essential in promoting public goods, enhancing the competitiveness of industries, and contributing to a liberalized global trading system. Therefore, various regions and countries actively develop standards strategy so that they could be used by diverse

interests to meet their objectives at the national and international level. By signing and implementing the Vienna agreement with the International Organization for Standardization (ISO), European Committee for Standardization (CEN) makes European standards recognized at international level by improving the exchange of information and cooperation with ISO, thus promoting the benefits of European standards to international trade and markets harmonization [1]. On the other hand, by encouraging enterprises to participate in the development of international standards, EU maintains the competitive edge in the field of international standardization activities [2, 3]. To promote and facilitate enterprises competitiveness through the development and use of standards, the American National Standards Institute (ANSI) works cooperatively with standards developing organizations (SDOs) to enhance the global competitiveness of U.S. business [4]. For example, standards developed by the American Society for Testing and Materials (ASTM) and the Institute of Electrical and Electronic Engineers (IEEE) are globally recognized leaders in developing standards, part of which have become de facto standards and are adopted worldwide [5]. The government of Japan has established a high-level coordination mechanism for standardization through the Intellectual Property Strategy Headquarters [6]. Headed by Prime Minister, the Intellectual Property Strategy Headquarters developed International Standard Strategy, which included establishing Prime Minister Award and other initiatives to encourage enterprises to take part in international standardization activities. South Korea has established a National Standards Council, chaired by the Minister of Ministry of Knowledge Economy (MKE). The

^a Corresponding author: liax@cnis.gov.cn

National Standards Council is composed of a few private sector experts and Deputy Ministers of different ministries, responsible for reviewing and coordinating major standards-related issues [7]. In the latest National Standards Plan approved by the National Standards Council, the South Korean government emphasized the importance of national standards development to support future prioritized industry and relevant international standards activities, including the strategy that proactively participates in international standards for infra-technologies [8].

China is not only the world's largest manufacturer and leading producer but also has become an export juggernaut in recent years, which makes products "Made in China" available everywhere. However, due to several factors, including the indicator difference between Chinese standards and the standards implemented in developed countries, Chinese products recalled from foreign markets have been rapidly increasing in the last few years. For example, the U.S. Consumer Product Safety Commission (CPSC) ordered 279 recalls in 2017, 134 of which was produced in China, accounting for 48% of the total number of recalls in the year [9]. China has taken several measures to the growing problem in standardization. In 2015, the State Council of China released the "Reform Plan for Deepening Standardization" to help standardization play the fundamental and strategic role in the national governance system and governance capability, to boost the sound and sustainable economic and social development [10]. In the same year, the Office of the Leading Group for the Belt and Road Initiative issued the Action Plan on Belt and Road Standard Connectivity, with the intent to speed up the development and implementation of Chinese standard "going global" work plan [11]. Under this circumstance, standardization work has drawn unprecedented attention.

Due to different technological levels, industrial processes and production methods, Chinese standards and foreign standards differ regarding technical indicators, requirements and standard systems, which is also one of the major limiting factors of competitiveness of Chinese products and technologies, resulting in the poor recognition of products complied with Chinese standards. By performing the comparison of standard indicators, the technical consistency of the standard can be confirmed. It not only gives the conclusion of comparisons from the professional point of view, pointing out the inadequacies in Chinese standards but also clears up misunderstandings of Chinese standards due to the differences in standard systems and production methods. To date, there are no specific methods and techniques for comparing technical indicator from standards. The increased demand for standard comparison amongst Chinese enterprises has enabled some relative research to be conducted, but no common method has been demonstrated yet. Therefore, it is imperative to study universal methods to compare indicators between Chinese and foreign standards in various industries.

2 Theoretical basis for comparison of standard indicators

The theoretical basis for the comparison of standard indicators is the Semantic Web technology. Semantic Web, often known as the web of data, enables data to be linked with each other so that the meaning is machine readable, which helps increase the efficiency in many sophisticated tasks [12]. Only by understanding the meaning of data can we make more efficient use of the underlying data. In most cases, there is a lack of meaning in the information itself, which needs the utilization of user commands or complex programming codes to reveal the specific meaning. For example, HTML tags are used to specify the formatting of the text on the webpage. The header tag <h1> tag defines the main heading, which means the text surrounded by <h1> tag is more important than others in semantics. However, these tags are merely isolated keywords which lead to a lack of connection, thus failing to provide a more defined context. It can only be used for exact matching due to the weakness of semantic information. Similarly, in terms of databases, the semantic cues that assist in comprehending remain limited, provided that tables and columns from databases are well specified.

Semantic Web provides a common framework that allows information to be represented as a set of statements. It consists of three parts: subject, predicate, and object, known as a triple. The subject denotes the object that is to be described, and the predicate expresses the relationship between the subject and the object. These data interchange can be achieved through the Resource Description Framework (RDF) [13]. To further extend RDF vocabulary to better specify the data structure, RDF Schema (RDFS) was released to provide mechanisms for describing groups of related resources and relationships [14]. The comparison of standard indicators adopts the semantic data descriptions. By establishing the RDF applicable to the standard indicators, this paper developed a "standard indicator comparison method" including fragmentation of standard documents, knowledge reorganization, and standard indicator comparison.

3 The method used for comparison of standard indicators

3.1 Indicator extraction and annotation from standard documents

It is necessary to perform text recognition, content structuring, content classification, and extract indicators from standard documents for subsequent indicator comparison. First, text recognition of standard documents in PDF format was applied as the precondition for the follow-up work. We used commercial Optical Character Recognition (OCR) software for the text recognition. Second, section structure and content from standard documents were extracted automatically by content structuring analysis. Third, the extracted content was identified and classified by indexing and classification. Eventually, the indicator extraction was carried out. With

the use of computer-aided manual review, the indicators were organized according to the indicator classification system and annotated to the corresponding content (Figure 1).

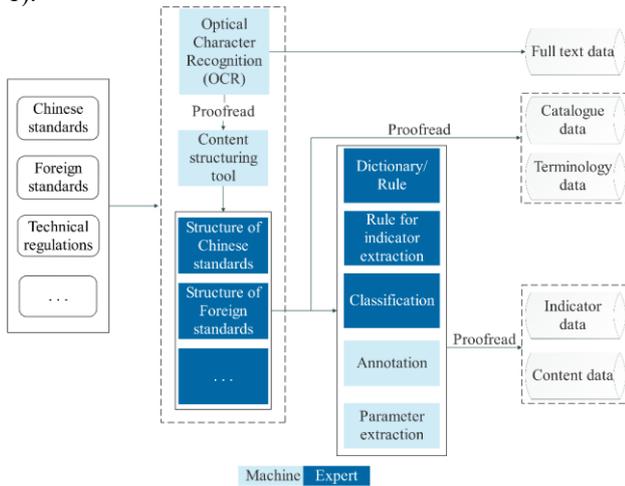


Figure 1. The workflow to extract indicator data and associate it with corresponding content.

3.2 Fragmentation of standard documents

The fragmentation of standard documents established a description framework in the form of “product - stylistic rule - indicator,” learned from RDF triples. RDF uses syntax to describe the metadata as a data model. An RDF file may include multiple resource descriptions, which are made from statements. The statement is written in the form of resource - property - value, corresponding to an RDF triple which is a sequence of subject, predicate, and object. In “product - stylistic rule - indicator” framework, the product mainly summarizes objects mentioned in the standard; the stylistic rule is used to generalize structure of standards that fall in the same sector or industry; The indicator represents the content of indicators describing the corresponding stylistic rule of product, which includes indicator name, value, unit, scope and so on. The product, stylistic rule, and indicator can be regarded as the subject, predicate, and object in a standard document, respectively.

3.3 Knowledge reorganization based on fragmentation of standard documents

In linguistics, hyponymy is used to describe the relationship between words or phrases, that is, hyponyms are specific instances included in a generic term, known as their hypernym [15]. When we compare standard indicators, the product is broadly seen as a hypernym that has a class of hyponyms. For example, as a generic term, transformer includes autotransformer, constant voltage transformer (CVT), booster transformers and other specific transformers. Thus, the comparison of transformer standards should include standards from all specific transformers. To meet this demand, it is necessary to set up knowledge reorganization based on standard fragments, which mainly focuses on the reorganization of the terms and definitions from products. By incorporating knowledge reorganization, the relationship between

hyponyms and hypernyms could be summarized for the following comparison.

The stylistic rule is another dimension that needs knowledge reorganization. As a type of normative document, the development of Chinese standards not only follows GB/T 1.1-2009: Directives for standardization Part 1: Structure and drafting of standards to comply with certain stylistic rule and structure, but also in line with GB/T 13016-2009: Principles and requirements for preparing diagrams of standard system, to establish standard system based on certain principles and requirements. Stylistic rules also need to be generalized and reorganized based on the synonyms, as it might vary in different standards. For instance, “range” and “scope” should be considered the same in most situations.

The indicator is the minimum granularity of standard document fragments. To make indicators comparable, the names and units of the indicators need to be generalized and reorganized as well, such as the normalization of m^3 and the cubic meter.

3.4 Comparison of standard indicators

Based on requirement analysis and previous research results, there are two methods in comparing standard indicators. Namely standard-based comparisons (two or two sets of standards) and comparisons based on products (cross-standard comparisons) respectively. The standard-based comparison mainly relies on the layout and stylistic rule that standards used and are compared within the same sections. The comparison based on products adopts the “product - stylistic rule - indicator” triple method, which allows thorough comparisons of the same product in different standards. This method is applicable to compare national standards and industrial standards, or between Chinese standards and foreign standards.

4 Standard indicator comparison system

After the indicator data and annotation were extracted and stored, this paper designed and implemented a system to compare indicators from standards. The hardware topology, data model and key technology adopted by the system will be illustrated.

4.1 Hardware topology

The hardware topology for the system contains three servers. The database server hosts the Microsoft SQL Server database, the processing system server deploys the processing system for standard indicators, and the application system server deploys the system to compare the indicators. The structure allows users and experts to query comparison of standard indicators service and standard indicators processing service using computer terminals through Browser/Server architecture (Figure 2). The system is built by using .Net Framework 4.0, with DotNetNuke platform as the content management system and SQL server 2012 for data storage and retrieval.

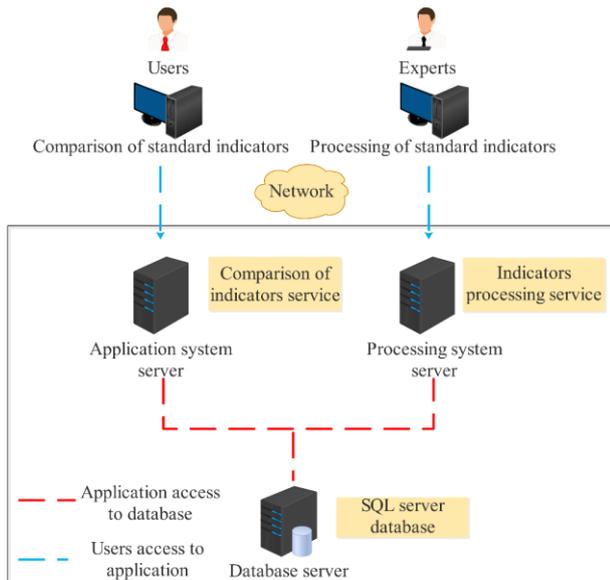


Figure 2. The illustration of hardware topology deployed for comparison of standard indicators system.

4.2 Data model

Based on the traditional relational data model, the design focused on the implementation of fragmentation and knowledge reorganization. That is, based on product - stylistic rule - indicator triple structure, the knowledge reorganization to explain semantic relations for product and stylistic rule should be applied. The semantic relations include hyponyms, hypernyms, and synonyms, while the indicators include names, values, units, and qualified classes, as shown in Figure 3.

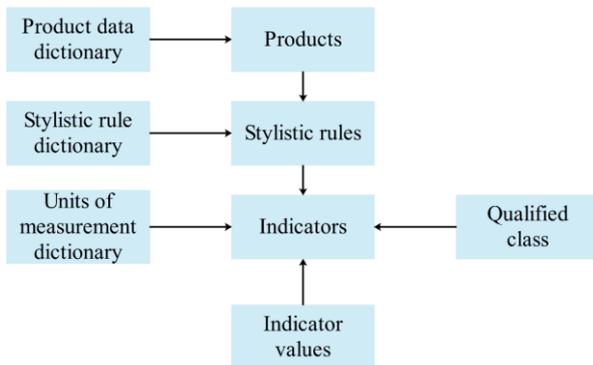


Figure 3. The data model of comparison of standard indicators system.

4.3 Key technologies

The development of the system adopted three key technologies, including content structuring, content classification, and space-time tradeoff method.

To improve the extraction efficiency of standard documents, a process was elaborated for content structuring. 1) Find nodes of all the sections in a standard document tree. 2) Mark the section nodes in a fixed format with XML tags, including section numbers, titles, text contents, figures, tables and reference links. 3) Extract marked texts and store in the database with a fixed

database format.

Because the body and structure vary from standard to standard, the way of expression as well as the section that technical content lies in can be inconsistent between standards. Therefore, annotated text classification was carried out to enable a more accurate standard indicator comparison. The classification method was established based on the characteristics of the body structure of standard documents, which included the representation of standard document text, standard classification, and the mapping of text and classification. Based on the clusters predefined by experts and classified standard documents as the training set, the classification model was derived from training [16]. The model was continuously updated and optimized for the standard documents to be classified (Figure 4).

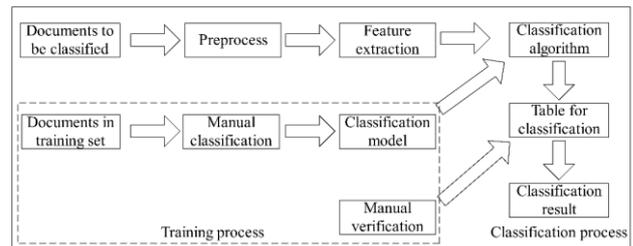


Figure 4. The illustration of content classification.

Due to the complexity of the table structure and the large number of standard indicators data, the comparison of standard indicators would be difficult to respond in time when a multiple table query is requested. In order to solve this problem, by making use of space-time tradeoff, the system added a scheduled task to merge multiple tables into one lookup table to support query requests. The lookup table contains ten tables, including standard information, ontology information, stylistic rule information, primary indicator, indicator information, qualified class information, content revelation information and other tables that allow for easy data querying. By optimizing and merging the table layout, this system not only simplified programming but also solved the problem of inadequate response of the query page. The response time of a simple query with high computational complexity is reduced from 60 seconds to within 1 second (the response time may differ given different search conditions).

5 Application examples

The indicator value of a product from the different level of standards can be found to be different or conflicted due to the complexity of Chinese standards. It is vital when it comes to hygiene and sanitation since hygienic, sanitary and safety issues are connected with the welfare of people's lives. Therefore, this paper analyzed Chinese standards from the food industry by collecting ontology concepts and stylistic rules in related fields, and three examples were given to illustrate the approach to compare standard indicators by using the system.

We first applied the comparison between the different level of Chinese standards for the same product. By comparing technical indicators from the different level of

standards for the same ontology, the similarities and differences between standards could be identified. In the first example, we chose wine as the product, and the related standards are comprised of three levels of stylistic rules. The first level determines the type of standards, which includes product standards, and standards for entry-exit inspection and quarantine. For product standards, the second level covers ten stylistic and structural rules, ranging from scope to appendix, and the third level further extended the provision set by level two. For example, the technical requirements in the third level are stipulated in sense, physical and chemical, and hygiene indicators. On the aspect of indicators, it contains the final factors based on the related term in level 3. In this example, there are, respectively 3 and 12 indicators in sensory requirements and physical and chemical indicators (Figure 5).

We chose benzoic acid and sodium benzoate as the indicator for comparison, which was essential to the quality of the wine. Two related standards were found, which were GB 15037-2006: Wines, and GB 2760-2014: National Food Safety Standard, Standard for Uses of Food Additives, respectively. The result indicated that specified indicator values in GB 2760-2014 are more stringent compares with the GB15037-2006. GB 15037-2006 stipulated that benzoic acid and sodium benzoate in wines must be lower than 50mg/L, whereas no benzoic acid nor sodium benzoate were allowed as an additive in wines represented in GB 2760-2014.

The confliction between indicator values set by different standards for the same product is not uncommon in Chinese standards. In the second example, we compared the Dichlorodiphenyltrichloroethane (DDT) level from guava, two current industrial standards that stipulated DDT limit in guava were reviewed. In the first standard, NY/T 518-2002: Guavas, CODEX STAN 215-1999, Codex standard for guavas, MOD, the standard dictate that the product shall comply with the maximum residue limits for pesticides established by the Codex Alimentarius Commission, which demands that DDT should not be found in guava samples [17, 18]. On the other hand, by

specifying the scope of tropical and subtropical fruits, NY/T 750-2011: Green food – Tropical and subtropical fruits, adopts a relatively loose condition, which requires the DDT level to be less than 0.05 mg/kg. DDT was one of the first chemicals developed as insecticide, and studies have shown that a range of human health effects is linked to DDT and its breakdown product DDE [19, 20]. Even though the two standards were both published by the Ministry of Agriculture of the People's Republic of China, the different technical committees and scopes led to the conflicted perspective on standard indicators. Therefore, the standard indicators should be unified in case of conflict between different standards for the same product.

Since the hyponymy pattern has been adopted in the system, the relationship between hypernym and hyponym could be concluded. This function realized the comparison between indicators from different hyponyms included in a generic term. In the third example, we chose three hyponyms of milk as research objects, namely evaporated milk, sweetened condensed milk and liquid milk. Milkfat and protein were compared among them. According to the definition by CODEX, evaporated milks are milk products which can be obtained by the partial removal of water from milk by heat, or by any other process which leads to a product of the same composition and characteristics. Sweetened condensed milks are milk products which can be obtained by the partial removal of water from milk with the addition of sugar, or by any other process which leads to a product of the same composition and characteristics [21, 22].

Evaporated milk, sweetened condensed milk, and liquid milk have lots of varieties. By including the hyponyms of each object, the indicators of milkfat and protein content can be compared (Table 1). The industrial standard NY/T 657-2012: Green food – Dairy product, has specified physical and chemical indexes for dairy products separately. With the advance of the system, we can infer from the result that both evaporated milk and sweetened condensed milk are required to have similar milkfat composition, which is in line with the fact that the

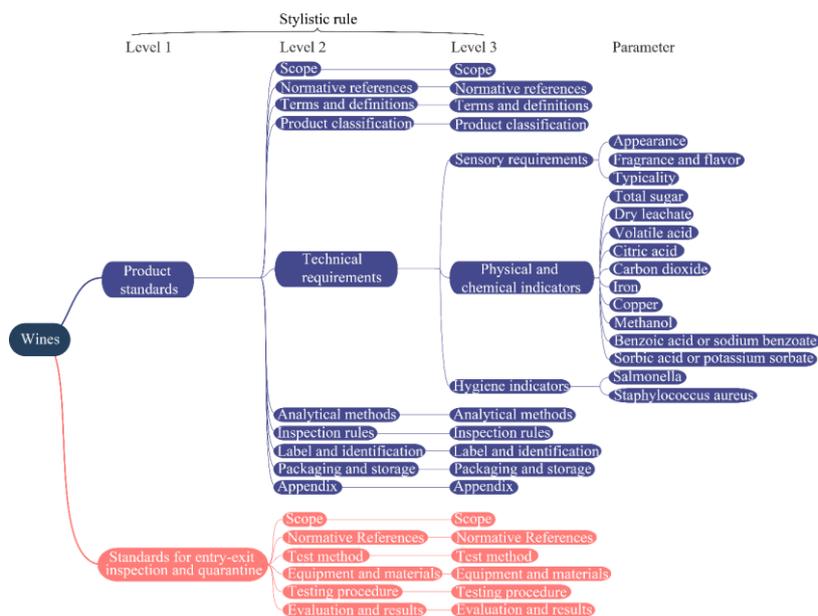


Figure 5. The layout of the product-stylistic rule-indicator triplet in wine standards.

Table 1. Physical and chemical indexes illustrated by industrial standard NY/T 657-2012.

Products	Items	Milkfat (x)/(g/100g)	Protein (y)/(g/100g)
Evaporated milk	Evaporated high-fat milk	$X \geq 15.0$	
	Evaporated milk	$7.5 \leq X < 15.0$	
	Evaporated partly skimmed milk	$1.0 < X < 7.5$	
	Evaporated skimmed milk	$X \leq 1.0$	
Sweetened condensed milk	Sweetened condensed high-fat milk	$X \geq 15.0$	$\geq 34\%$ of the milk solids-not-fat
	Sweetened condensed milk	$7.5 \leq X < 15.0$	
	Sweetened condensed partly skimmed milk	$1.0 < X < 7.5$	
	Sweetened condensed skimmed milk	$X \leq 1.0$	
Liquid milk	Pasteurized milk	≥ 3.1	≥ 2.9
	Sterilized milk		

difference between these two products is whether or not sugar is a permitted ingredient. On the other hand, the milkfat must not be lower than 3.1g in 100g liquid milk. For the aspect of protein content, the requirement for protein varies between condensed milk and liquid milk, where the content of protein in both evaporated milk and sweetened condensed milk must not be lower than 34% of the milk solids-not-fat, while protein content must not be lower than 2.9g per 100g liquid milk.

6 Conclusion

Based on Semantic Web technologies, this paper develops a standard indicator comparison system using “product-stylistic rule-indicator” triplet. By introducing fragmentation of standard documents, knowledge reorganization, and standard indicator comparison, this system is suitable to compare indicators from a class of products specified by different standards.

In the practice of standard indicator comparison, the indicator comparison system provides a great experience in data retrieval and comparison. This system has the following advantages: First, it demonstrates all the available indicators in a specific product, which are reflected by various features and properties. Secondly, the overview supports the inclusion of all the qualified indicators from a certain class of products. Thirdly, it allows for the comparison of indicators from different standards to determine the differences from a product or within a class of products. Thus, by classifying the products and corresponding indicators, the batch query of technical indicators from one class of products could be realized.

By applying system practices to compare Chinese standard indicators, we find that there are still problems remaining in standard development and standardization, such as repeated standard development or conflicted indicators in standards. This suggests that the future standard-setting process should be backed by better resource allocation and information exchange to avoid

standard duplications or conflicted indicators from happening. Furthermore, to make Chinese standards at the level of world-class specifications for products, services, and systems to ensure quality, safety and efficiency, the translation of foreign standards is required to support the indicator comparisons between Chinese standards and foreign standards.

References

1. International Organization for Standardization, Agreement on Technical Cooperation between ISO and CEN (Vienna Agreement). Retrieved from <http://isotc.iso.org/livelink/livelink?func=ll&objId=4230458&objAction=browse&sort=subtype> (2001)
2. Directorate-General for Internal Market, Industry, Entrepreneurship and SMEs, EY, *Independent Review of the European Standardisation System*. EU Publications. Belgium (2015)
3. European Commission, *Standardization for a competitive and innovative Europe: a vision for 2020*. Retrieved from <http://www.anec.eu/attachments/Definitive%20EXPRESS%20report.pdf> (2010)
4. United States Standards Strategy Committee, *United States Standards Strategy - A Revision of the National Standards Strategy for the United States*. Retrieved from http://www.ansi.org/standards_activities/nss/uss (2015)
5. S. Hans, A. Nayyar, *2014 Fourth International Conference on Advanced Computing & Communication Technologies*. 372-376 (2014)
6. Intellectual Property Strategy Promotion Bureau, *Overview of the "Intellectual Property Strategic Program 2017"*. Retrieved from http://www.kantei.go.jp/jp/singi/titeki2/kettei/chizaik_eikaku2017_gaiyou_e.pdf (2017)
7. D.G. Choi, *A Primer on Korea's Standards System: Standardization, Conformity Assessment, and*

- Metrology*. Retrieved from <http://www.nist.gov/publications/primer-korea146s-standards-system-standardization-conformity-assessment-and-metrology> (2013)
8. Study on the establishment of the 4th National Standards Plan by the Ministry of Trade, Industry and Energy. Retrieved from http://www.prism.go.kr/homepage/theme/retrieveTheMeDetail.do?leftMenuLevel=110&cond_brm_super_id=NB000120061130100005269&research_id=1450000-201700082 (2015)
 9. U.S. Consumer Product Safety Commission, CPSC recall list in 2017. Retrieved from <http://www.cpsc.gov/Recalls> (2017)
 10. American National Standards Institute, *China's State Council Announces Plan for Standardization Reform*. Retrieved from http://www.ansi.org/news_publications/news_story?menuid=7&articleid=61bd0ca1-6bfd-4c5c-b780-ddb45316bba0 (2015)
 11. Belt and Road Portal, *Action Plan on Belt and Road Standard Connectivity (2015-17)*. Retrieved from <http://eng.yidaiyilu.gov.cn/zchj/qwfb/35977.htm> (2017)
 12. G. Antoniou, F.V. Harmelen, *A Semantic Web Primer, 2nd Edition (Cooperative Information Systems)*, (2008)
 13. S. Powers, *Practical RDF - solving problems with the resource description framework*, (2003)
 14. D. Allemang, J. Hendler, *Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL*, (2011)
 15. M. Stede, In: *Proceedings of the first international conference on Natural language generation*, 93-99 (2000)
 16. T. Lahtinen, *Automatic indexing: an approach using an index term corpus and combining linguistic and statistical methods*. University of Helsinki. Helsinki (2000)
 17. Codex Alimentarius Commission, *CODEX STAN 215, CODEX Standard for Guavas*. Retrieved from http://www.fao.org/input/download/standards/341/CXS_215e.pdf (1999)
 18. Codex Alimentarius Commission, Maximum Residue Limits (MRLs). Retrieved from <http://www.fao.org/fao-who-codexalimentarius/codex-texts/maximum-residue-limits/en/> (2017)
 19. M.P. Longnecker, W.J. Rogan, G. Lucier, *Annu. Rev. Public Health* **18**, 211-244 (1997)
 20. B. Eskenazi, J. Chevrier, L.G. Rosas, H.A. Anderson, M.S. Bornman, H. Bouwman, A. Chen, B.A. Cohn, C.D. Jager, D.S. Henshel, *Environ Health Perspect* **117**, 1359-1367 (2009)
 21. Codex Alimentarius Commission, *CODEX STAN 281, CODEX Standard for Evaporated Milks*. Retrieved from http://www.fao.org/input/download/standards/172/CXS_281e.pdf (1971)
 22. Codex Alimentarius Commission, *CODEX STAN 282, CODEX Standard for Sweetened Condensed Milks*. Retrieved from http://www.fao.org/input/download/standards/173/CXS_282e.pdf (1971)