Design and Implementation of the Multilingual Product Retrieval Agent through XML and the Semantic Networks in EC [*]

Yoo-Jin Moon [1], Kijoon Choi [1], Kyongho Min [2], Wan Pyong Kim [1], Youngho Hwang [3], Pankoo Kim [4], and Youngse Mun [5]

[1] MIS Department, Hankuk University of Foreign Studies
270 Imun-dong Tongdaemun-Gu
Seoul 130-791, Korea
yjmoonmis@hanmail.net, kjnch@orgio.net, wan100@chollian.net
[2] School of Computer and Information Sciences
Auckland University of Technology
Auckland 1020, New Zealand
kyongho.min@aut.ac.nz
[3] Kunsan University, Kunsan, Cheonbuk 573-701, Korea
[4] Chosun University, Kwangju 506-741, Korea
[5] Korea National Defense University, Seoul 122-875, Korea
yhwang@kunsan.ac.kr, pkkim@chosun.ac.kr

Abstract. Retrieval for products is an important task for e-commerce, since it represents an interface of the customer contact to e-commerce. And e-commerce should provide customers with easily accessible processes in searching. Especially, the product information on the World Wide Web needs integration and standardization to keep the pace of rapid expansion with wide reachable ranges. International standards on product catalogs are converging on UNSPSC (Universal Standard Products and Services Classification). With adoption of this standard, we designed the architecture of a multilingual product retrieval agent. The architecture is based on the central repository model of product catalog management with a distributed updating process. It also includes the perspectives of buyers and suppliers. In addition, the consistency and version management of product information are controlled by UNSPSC. The multilingual product names are resolved by semantic networks, a thesaurus, and product name ontology, which enable the present architecture to be expanded to the Semantic Web applications.

1   Introduction

One prerequisite to facilitate web-based e-commerce is a technological infrastructure. The technological infrastructure has several layers, which include network protocols,

---

application architecture, and business process technologies. Recently many researchers move toward issues of the application architecture. The architecture issues address the impending problems without losing flexibility and potential appropriation. A key process of the architecture issues is a product search or a retrieval service for potential buyers in Internet. From the buyers' point of view, the searching process on the World Wide Web is tightly related with efficiency of the electronic transactions.

A classification system of the product plays a key role in developing a solution of product retrieval using a standard framework in e-commerce [1]. Thus, it is important to standardize e-catalogs of the product classification system. In reality, each industry develops its e-catalogs independently. Therefore, we need to integrate product information across industries and languages for a globalized e-commerce environment. The complexities of language translation and product information representation are mixed in web-based e-commerce sites. And endeavors to solve these complexities lead to Semantic Web implementation in e-commerce. The advance of integration technologies and mechanisms based on the world wide web paved the way to design a multilingual product retrieval agent with a semantic processing capability. We propose the architecture of the multilingual product retrieval agent, based on the Semantic Web, that tackles the complexities of the language translation and the product information representation.

2   Literature Review

To process product information in Internet for e-commerce we need to develop XML-based knowledge representation [2]. Self-describing features of XML trigger an integration problem among web-based documents [3]. Each developer may use a different set of XML tags and structures, which results in the chaos of semantic interpretations. This limitation of XML invokes Semantic Web activities [4]. The Semantic Web can give a formal structure to the web-based documents, and enables software agents to automate semantic processing on behalf of users.

XML has definite limits on semantic representation of the structure in a consistent way [5]. So it may be used for an integration mechanism at the conceptual level for the product retrieval agent. Therefore, we need a Semantic Web technology to deal with various product information and a multilingual problem involved in globalized e-commerce environments, even when we use the e-catalogs [6], [7].

RDF (Resource Description Framework) is a cornerstone of the Semantic Web, in which the RDF represents semantic relations by subjects, predicates, and objects [8], [9]. The RDF provides minimal consistency without losing flexibility, and it delegates domain-dependent representation to ontologies [10], [11]. Some part of the ontology is contained in an e-catalog system [7]. The Semantic Web will enable the computer-operated WWW data to be automated, integrated, and reused [12], [13].

Ontology is used for knowledge representation of facts or states defined by entities and their relationships. The ontology is an enabler of Semantic Web technologies [14]. The ontology provides relevant domain knowledge for RDF [15], [16], which has relational structures among resources, properties, and values. The well-defined

ontology on the relevant domain is a prerequisite for semantic applications in e-commerce [17].

Semantic networks represent hypernyms of word senses in the form of isa-hierarchy. WordNet [18] is an English semantic network based on a psycholinguistic theory. It represents lexical concept in the form of isa-hierarchy of each word sense. Synonym sets (synsets) is a set of lexical concepts of English nouns, adverbs, adjectives, and verbs. Each synset represents one underlying lexical concept. This paper refers to WordNet for nouns as a knowledge base to disambiguate word senses in English product names.

The semantic network of Korean nouns, called Korean Noun Networks (KNN) [19], has been built in the form of ISA hierarchy. The ISA hierarchy consists of a node and an edge. The node represents synonym sets composed of Korean nouns and English WordNet. The edge represent hypernymous relations between the nodes. In this paper, KNN are utilized to automatically extract the sets of hyponymous concepts. This paper refers KNN as a knowledge base to disambiguate word senses in Korean product names.

There are many kinds of agents for e-commerce, for example, agents for recommendation, agents for negotiation, agents for comparison shopping, agents for e-catalog management, and etc. The agents proposed in this paper have common characteristics with those for recommendation and comparison shopping. The agents for recommendation need to meet the conditions as follows. First, acquisition of users' information should follow the semantic relationships of web-based documents or contents. Second, the agents have to show flexible responses to changes in web site contents. Third, they need to accommodate, incrementally and persistently, the change of a user's requirement. Fourth, they should have a natural user interface. Fifthly, they need implicit methods to draw user requirements. Finally, they should allow continuous learning for performance enhancements. With these features enabled, we need a distributed agent managing product codes used in e-catalogs [6].

Standardization can contribute to reduce the development cost [1], and to facilitate interoperability between heterogeneous systems [20]. Several propositions about standardization have been published within interoperable B2B business frameworks. Most frameworks are based on XML and Table 1 summarizes the most popular standardized frameworks such as BizTalk by Micoroft, RossettaNet by IBM and Intel, ebXML by UN/CEFACT and OASIS.

Table 1. General Characteristics of E-commerce Frameworks

Framework	Structure	Applicable Industry	Major Groups
ebXML	horizontal	General	OASIS, UN/CEFACT
RossetaNet	vertical	IT/EC/SM	IBM, Intel, and other commercial corp.'s
BizTalk	horizontal	General	Microsoft
eCo	horizontal	General	CommerceNet

There are two kinds of product code for e-commerce: a product classification code and a product identification code. The product classification code enables categorical searches by grouping similar products, and the product identification code matches a specific product to a code. An UNSPSC (Universal Standard Products and Services Classification) code system has been developed especially for e-commerce. It is the one of the best code systems for alignment of diverse systems in distributed environments. The characteristics of the UNSPSC code system are as follows. First, it fits for e-commerce, ERP, spending analysis, and etc. Second, product categories are adjustable to most industries. Third, the classification is detailed and publicly accessible without restriction.

A central repository model is for the alignment of e-catalog storages by ebXML and UDDI (Universal Description, Discovery and Integration) technologies. There are other e-catalog alignment models such as the single server model, the virtual catalog model, and the mediator model.

3   Design of a Multilingual Product Retrieval Agent

The architecture utilizes the central repository model for e-catalog management. The central repository model is a pull model that the users register their product codes voluntarily, and the repository maintains the catalog information providing access to anyone. We used UNSPSC (Universal Standard Products and Services Classification) as a product code system that provides consistent digit-based universal product, and service codes across language variants.

The proposed architecture reflects both buyers' and suppliers' perspectives in product information integration. Structures built in the directories of product information may be maintained individually. The digit-based product code system gives the consistency across different mechanisms involved.

Ambiguity and uncertainty of product names, that are caused by synonyms, homonyms, and polysemys, can be processed for product names by a semantic network, a thesaurus, and an ontology dictionary of product names. Fig. 1 illustrates examples of the synonyms which will be processed by the semantic network of each language. Fig. 2 illustrates examples of a pair of homonyms and a pair of polysemys which will be processed for a convenient user interface by the ontology dictionary and the thesaurus for products, from the viewpoint of buyers and suppliers. Referencing the thesaurus may provide relevant terms and concepts, but it can not provide semantic differences in computer-operable ways. So, the ontology dictionary of product names is introduced for differentiation of polysemys illustrated as in Fig. 2.
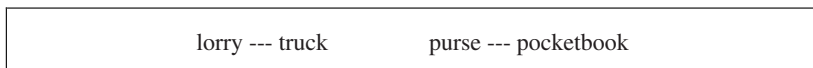
lorry --- truck   purse --- pocketbook

Fig. 1. Examples of Synonyms in Product Names

foil -> sheet metal,	foil -> fencing sword
bonnet -> hat,	bonnet -> protective covering (ex. a part of car)

Fig. 2. Examples of Homonyms and Polysemys in Product Names

RDF is a representation of metadata for resources [19], [20]. Generally, it applies to information on web sites. Ontology on product information may be integrated with RDF modelings to facilitate product retrieval on the web sites. In Fig. 3, 'Shoes' is represented by a RDF graph. The UNSPSC code of Shoes is 53111600: 53 for a segment, 11 for a family, 16 for a class, 00 for a commodity. Table 2 shows the composition of the UNSPSC codes of 'Shoes', 'Men's Shoes', 'Women's Shoes', and 'Skate Shoes'.

Table 2. Examples of UNSPSC Codes for Products

Product	Segment	Family	Class	Commodity
Shoes	53	11	16	00
Men's Shoes	53	11	16	01
Women's Shoes	53	11	16	02
Skate Shoes	49	15	16	02

Properties or predicates can be used to represent the relational property of product information as illustrated in Fig. 3. In this example productCode_Of is a predicate element. It may also be represented as an object.
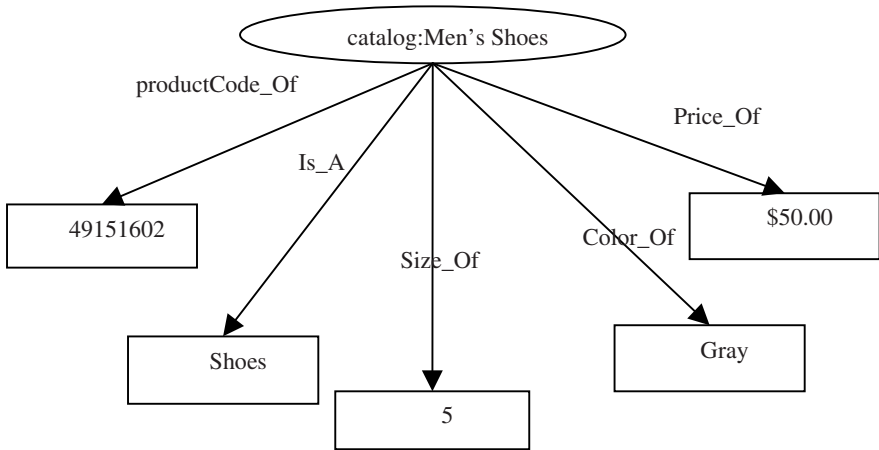

Fig. 3. An Example of Properties of Men's Shoes in a RDF Graph

Fig. 4 shows an example of RDF representation of properties of Men's Shoes in Fig. 3.

```
<RDF xmlns:a="http://mislab.hufs.ac.kr/catalog#">
<rdf:Description ID="Men's Shoes">
   <a:Price_Of>$50.00</a:Price_Of>
   <a:Color_Of>Gray</a:Color_Of>
   <a:Size_Of>5</a:Size_Of>
   <a:Is_A>Shoes</a:Is_A>
   </rdf:Description>
</rdf:RDF>
```
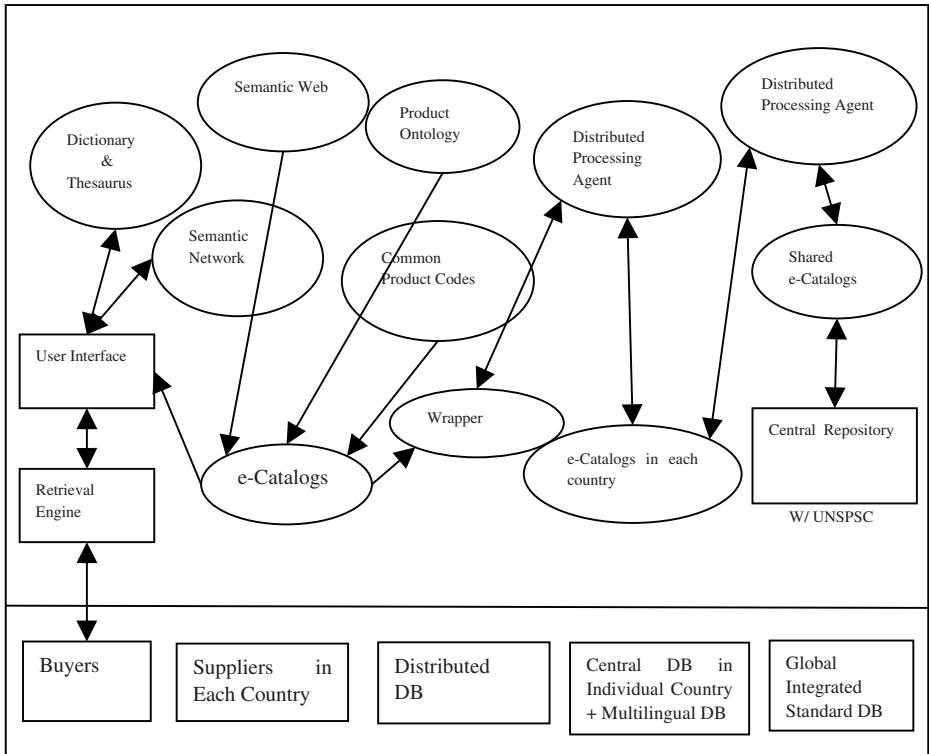
Fig. 4. RDF Representation of Properties of Men's Shoes

Fig. 5. Architecture of A Product Retrieval Agent

Fig. 5 illustrates the architecture in which our proposed retrieval agent works. The main characteristics of the proposed product retrieval agent are as follows:

- Knowledge bases: a common product code (UNSPSC), product ontology, common e-catalogs, a bilingual dictionary, multilingual semantic networks, and a thesaurus
- Perspectives: buyers' and merchants' perspectives
- Management model of e-catalogs: a central repository model with distributed processing environments
- Level or range: buyers, merchants, suppliers, product information processing procedures in each country, globally integrated standard catalogs

In this paper we used the semantic network to resolve ambiguities caused by synonyms, homonyms, and polysemys in the system implementation. The architecture covers overall interactions among components involved in product retrieval processes. The architecture is based on the central repository model of e-catalog management using globally integrated standardization of product information. The central repository model is an alignment model for industry e-catalog storages using ebXML and UDDI. And the classification structures are to be implemented in XML, specifically in Semantic Web.

Data consistency is maintained by 8-digit (optionally 10-digit) coded UNSPSC. The UNSPSC code system is well suited for external interoperability of product information, and is specifically developed for electronic commerce. The problems of a static product code system can be alleviated by the semantic networks, that support a semantic-based retrieval process from the buyers' perspective.

4   Implementation of the Agent for Product Retrieval

In this chapter, we describe the main implementation algorithms of ambiguous product names, which are not included in the UNSPSC code system for convenient multilingual product retrieval processes. And we define several functions of the algorithms related to the multilingual product retrieval agent, as follows:

- UNSPSC(user): a set of UNSPSC list translated into a language used by a user;
- UNSPSC_CODE(pn): a UNSPSC code of pn (i.e. product names) in UNSPSC list;
- SYNSET(pn): a set of synset of pn extracted from semantic networks;
- HYPERNYM(pn): an ordered list of pn's hypernyms extracted from an ontology dictionary;
- HYPONYM(pn): an ordered list of pn's hyponyms extracted from an ontology dictionary;
- SEMANTIC_NETWORKS(user): a set of semantic networks built by the user's language;
- THESAURUS(user): a set of thesaurus for product names written in language used by the user;
- RELATED(pn): an ordered list of a product name related to pn in the thesaurus;
- TRANSLATED(pn): translated pn into English using a bilingual dictionary;

An algorithm retrieving synonyms of product names selects a product name and its code from the UNSPSC code system in each language, using the semantic networks and the ontology dictionary of product names.

```
RetrievalForSynonyms(pn, user)  :
{Input of pn(a product name) from the user};
if pn is found in UNSPSC(user)
   then return the UNSPSC code
   else if SEMANTIC_NETWORKS(user) is not found
           then { pn = TRANSLATED(pn);  user=English;}
       EndIf;
EndIf;
find SYNSET(pn) from SEMANTIC_NETWORKS(user);
if  SYNSET(pn) is null   then exit();
for each i in SYNSET(pn) {
  if  i  is found in UNSPSC(user)
      then return  i &UNSPSC_CODE(i);
  for each j in HYPONYM(i) from SEMANTIC_NETWORKS(user)
    if   j is found in UNSPSC(user)
         then display i&j;
  for each k in HYPERNYM(i) fromSEMANTIC_NETWORKS(user)
    if  k is found in UNSPSC(user) then display  i & k;
}EndFor

Let the user select what he wants to retrieve from the
displayed lists;

if the user selects one item,
then return the item and UNSPSC_CODE(item);
EndRetrievalForSynonyms
```

An algorithm retrieving homonyms of product names selects the product name and its code from the UNSPSC code system using the ontology dictionary of product names.

```
RetrievalForHomoyms(pn, user)   :
{Input of pn(a product name) from the user};
If pn is found in UNSPSC(user) more than once
   then find all pairs of pn & UNSPSC_CODE(pn)
   else exit();
EndIf;
for  each i in all pairs of pn & UNSPSC_CODE(pn) {
     find  HYPERNYM(i)  from an ontology dictionary;
     display i & HYPERNYM(i);
}EndFor
```

```
Let the user select what he wants to retrieve from the
displayed lists;

if the user selects one item,
then return the item and UNSPSC_CODE(item);
EndRetrievalForHomonyms
```

An algorithm that retrieves polysemys of product names selects the product name and its code from the UNSPSC code system using the ontology dictionary and the thesaurus of product names.

```
RetrievalForPolysemys(pn, user)  :
{Input of pn(a product name) from the user};
If pn is found in the entry of  THESAURUS(user)  more
than once
  then find all pairs of pn & UNSPSC_CODE( RELATED(pn))
  else exit();
EndIf;
for each i in pairs of pn & UNSPSC_CODE(RELATED(pn))
   display i & RELATED(i);

Let the user select what he wants to retrieve from the
displayed lists;

if the user selects one item,
then return the item and UNSPSC_CODE(item);
EndRetrievalForPolysemys
```

The suggested algorithms were tested for the agents of keyword-based product retrieval. If the retrieved product name exists in the UNSPSC code system in either English or other languages, the agent returns the exact UNSPSC code number. Then the product retrieval is considered to be performed successfully.

Experiments show that 21 out of 24 synonyms of the product names have succeeded in product retrieval, and that 20 out of 20 homonyms of the product names have succeeded in product retrieval. Synonyms of the product names trigger failure in product retrieval when they do not exist in the semantic network. Polysemys of PN (product names) trigger failure in product retrieval when they are coined words.

5 Conclusion

The interoperability can be assured by developing XML-based e-catalogs on products. E-catalogs can represent structured product information, but may not be used to

integrate heterogeneous product information for efficient buyers' product searches. Recently proposed semantic web mechanisms, presented in RDF graphs, provide new opportunities for web product retrieval. The keyword-based retrieval is efficient as long as the product information is well structured and organized. But when the product information is displayed across many online shopping malls, especially when it is described in different languages with different cultural backgrounds, buyers' product retrieval needs language translation of the product information disambiguated in a specific context. We presented a RDF modeling case that resolved semantic problems in the representation of product information and across the boundaries of the language domain.

With adoption of the UNSPSC code system, we designed and implemented an architecture for the multilingual product retrieval agent. The architecture was based on the central repository model of product catalog management with distributed updating processes. It also included the perspectives of buyers and suppliers. And the consistency and version management of product information were controlled by the UNSPSC code system. The ambiguities of multilingual product names were resolved by the semantic network, the thesaurus, and the ontology dictionary of product names. The suggested algorithm would be applied to multilingual product retrieval in any language, as long as there exists its bilingual dictionary or its semantic network.

Classification standards representing products in electronic commerce require researches on design and implementation issues to validate the industrial guidelines involved in the standard-setting procedures. The UNSPSC code system should be standardized in the product code system and in the integration of product information across the industries.

Future works are, firstly, to update synonym sets of the semantic networks for convenient interfaces, and to update coined words in the UNSPSC system and the semantic networks. Secondly, the product names in the UNSPSC system should be completed with cooperation of suppliers and venders. Thirdly, each country should keep its own UNSPSC system translated into its language.

References

1. Zhao, Y.: Develop the Ontology for Internet Commerce by Reusing Existing Standards. Proceedings of International Workshop on Semantic Web Foundations and Application Technologies. (2003)
2. Anutariya, C., Wuwongse, V., and Akama, K.: XML Declarative Description with Negative Constraints. Proceedings of International Workshop on Semantic Web Foundations and Application Technologies. (2003)
3. Choi, M. Y., and Joo, K. S.: A Unified Design Methodology based on Extended Entity-Relationship Model for XML Applications. 2nd International Conference on Computer and Information Science (ICIS 2002). (2002) 136–141
4. Fu, Fengli, Song, Y., and Kim, Y.: XML and Bioinformatics. 2nd International Conference on Computer and Information Science. (2002) 147–152
5. Chao-Min, C.: Reengineering Information Systems with XML. Information Systems Management, Vol.17:4. (2000) 40–54

6. Baron, J. P., Shaw, M. J., and Bailey Jr., A. D.: Web-based E-catalog Systems in B2B Procurement. Communications of the ACM, Vol.43:5. (2000) 93–100
7. Stanoevska-Slabeva, K., and Schmid, B.: Internet Electronic Product Catalogs: An Approach Beyond Simple Keywords and Multimedia. Computer Networks, Vol.32. (2000) 701–715
8. Fensel, D., Hendler, J., Lieberman, H., and Wahlster, W.: Spinning the Semantic Web. MIT Press (2003)
9. Geroimenko, V., and Chen, C.: Visualizing the Semantic Web. Springer-Verlag (2003)
10. W3C: Resource Description Framework(RDF) Model and Syntax Specification. (1999)
11. W3C: Resource Description Framework(RDF) Schema Specification 1.0. (2000)
12. Berners-Lee, T.: The Semantic Web. Scientific American (2001)
13. Berners-Lee, T.: Weaving the Web. HarperCollins (2000)
14. Desmontils, E., and Jacquin, C.: Indexing a Web Site with a Terminology Oriented Ontology. (http://www.semanticweb.org/SWWS/program/full/paper5.pdf). (2003)
15. Broekstra, J., Klein, M., Decker, S., Fensel, D., van Harmelen, F., and Horrocks, I.: Enabling Knowledge Representation on the Web by Extending RDF Schema. Computer Networks, Vol.39. (2002) 609–634
16. Horrocks, I., and Hendler, J.: The Semantic Web - ISWC 2002. Springer-Verlag (2002)
17. Gruber, T. R.: A Translation Approach to Portable Ontology Specifications. Knowledge Systems Laboratory Technical Report KSL 92–71. Knowledge Systems Laboratory, Stanford University (1993)
18. Miller, G. A. , Beckwith, R., Fellbaum, C., Gross, D., and Miller, K.: Introduction to WordNet : An On-line Lexical Database. in Five Papers on WordNet, CSL Report. Cognitive Science Laboratory, Princeton University (1993)
19. Moon, Y.: Construction of Semantic Networks for the Language Information Processing. Proc. of International Symposium on Advanced Intelligent Systems. (2001) 42–46
20. Jhingran, A. D., and Pirahesh, M. N.: Information Integration: A Research Agenda. IBM Systems Journal, Vol.41:4. (2002) 555–562
21. Cruz, I. F.: Final Report of NSF Sponsored Workshop, International Semantic Web Working Symposium (SWWS). (2001) (http://www.semanticweb.org/SWWS/report/swws-report.pdf).