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

Yoo-Jin Moon, Kijoon Choi, Kyongho Min, Wan Pyong Kim, Youngho Hwang, Pankoo Kim and Youngse Mun

Abstract

International standards on product catalogue are converging on the UNSPSC. Adopting this standard, we designed an architecture for the product search agent employing Semantic Web and semantic networks.

The architecture is based on the central repository model of product catalog management with distributed updating processes. It also includes the perspectives of buyer's and supplier's. And the consistency and version of product information is controlled by a numeric product code system. The multi-lingual product names are resolved by semantic networks and product name ontologies, which make the design expandable to the Semantic Web applications.

Contents

Ⅰ. *Introduction*

► The technological infrastructure has several layers, which include network protocols, application architecture, and business process technologies.

► The architecture issues address the impending problems without losing flexibility and potential appropriation.

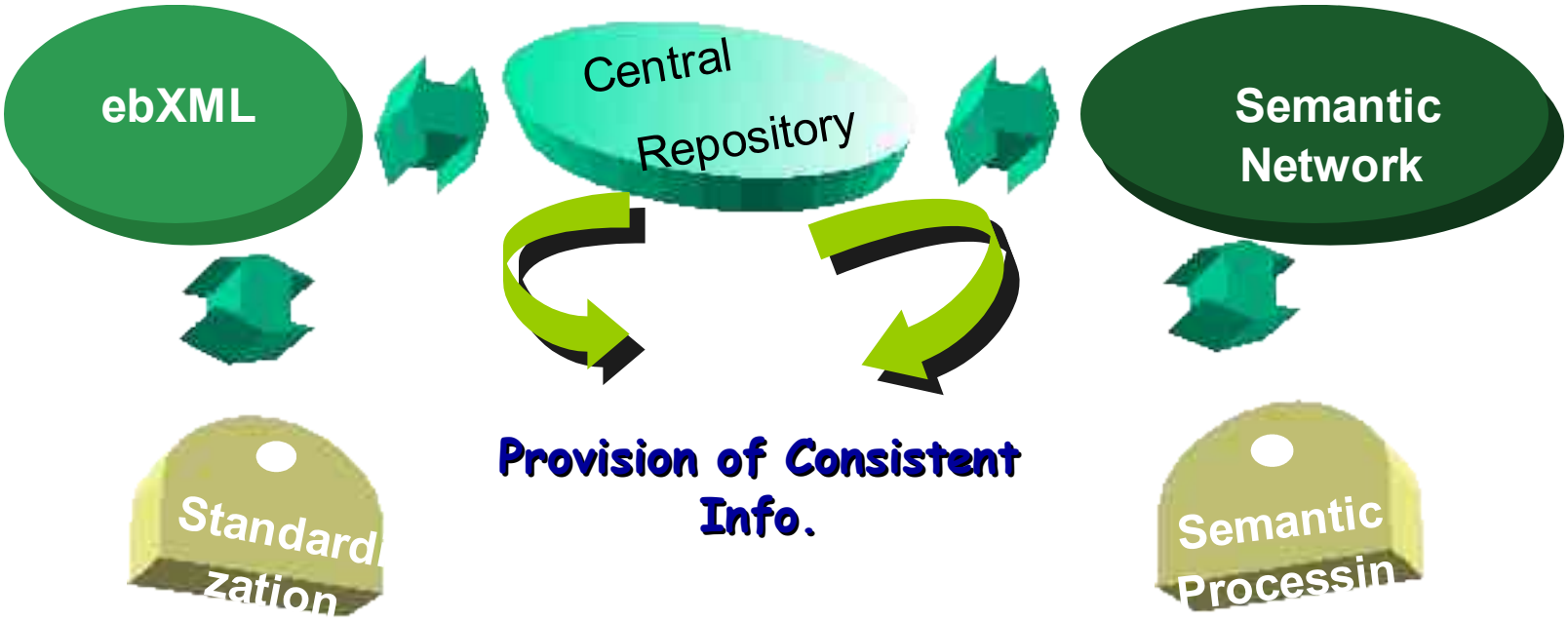
► A key process of the architecture issues is a product search or a retrieval service for potential buyers in Internet.

► It is important to standardize e-catalogs of the product classification system.

► We need to integrate product information across industries and languages for a global e-commerce environment.

Ⅱ. *Literature Review*

Semantic Web & Semantic Network



・ To process product information in Internet for e-commerce we need to develop XML-based knowledge representation.

・ XML has definite limits on semantic representation of the structure in a consistent way.

・ We need a Semantic Web technology to deal with various product information and a multilingual problem involved in global e-commerce environments, even when we use the e-catalogs.

・ RDF (Resource Description Framework) is a cornerstone of the Semantic Web, in which the RDF represents semantic relations by subjects, predicates, and objects.

・ The RDF provides minimal consistency without losing flexibility, and it delegates domain-dependent representation to ontologies.

・ Some part of the ontology is contained in an e-catalog system.

· The Semantic Web will enable the computer-operated WWW data to be automated, integrated, and reused .

· Semantic networks represent hypernyms of word senses in the form of isa-hierarchy. WordNet is an English semantic network based on a psycholinguistic theory.

· Standardization can contribute to reduce the development cost, and to facilitate interoperability between heterogeneous systems.

· Most frameworks are based on XML and Table 1 summarizes the most popular standardized frameworks such as BizTalk by Microsoft, RossettaNet by IBM and Intel, ebXML by UN/CEFACT and OASIS.

Table 1.

Framework	Structure	Applicable Industry	Major Groups
ebXML	horizontal	general	OASIS, UN/CEFACT
RossetaNet	vertical	IT/EC/SM	IBM, Intel, and so on
BizTalk	horizontal	general	Microsoft
eCo	horizontal	general	CommerceNet

Table 1. General Characteristics of E-Commerce Framework

- 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.

Ⅲ. *Design of Multilingual Product Retrieval Agent*

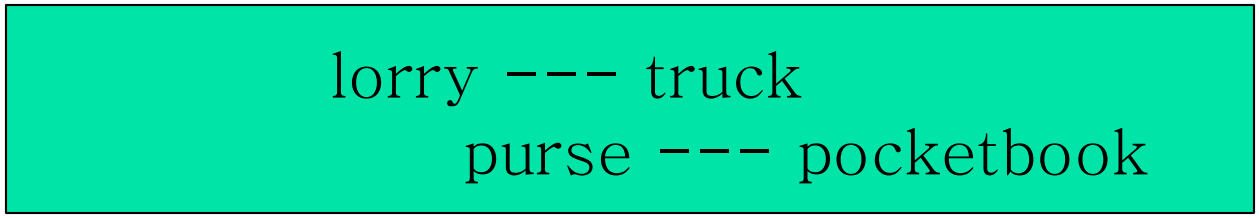


Fig. 1. Examples of Synonyms in Product Names

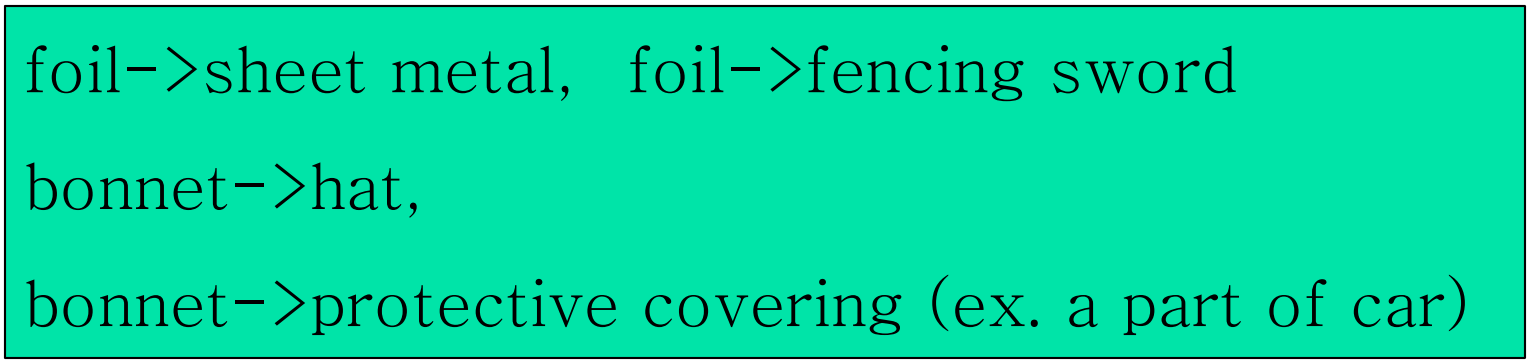


Fig. 2. Examples of Homonyms and Polysemys in Product Names

An Example of RDF

```
<rdf:RDF
Xmlns:rdf=http://mislab.hufs.ac.kr/
Xmlns:catalog=http://catalog.hufs.ac.kr/
<rdf:Description rdf:about="A_Desk">
<catalog:Color_of>Brown</catalog:color>
</rdf:Description>
</rdf:RDF>
```

Table 2.

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

Table 2. Examples of UNSPSC Codes for Products

Figure 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:Code_Of>53111601</a:Code_Of>
  <a:Is_A>Shoes</a:Is_A>
  </rdf:Description>
</rdf:RDF>
```

Fig. 3. XML Representation of Properties for Skate Shoes

- 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

- The Semantic 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

Figure 4.

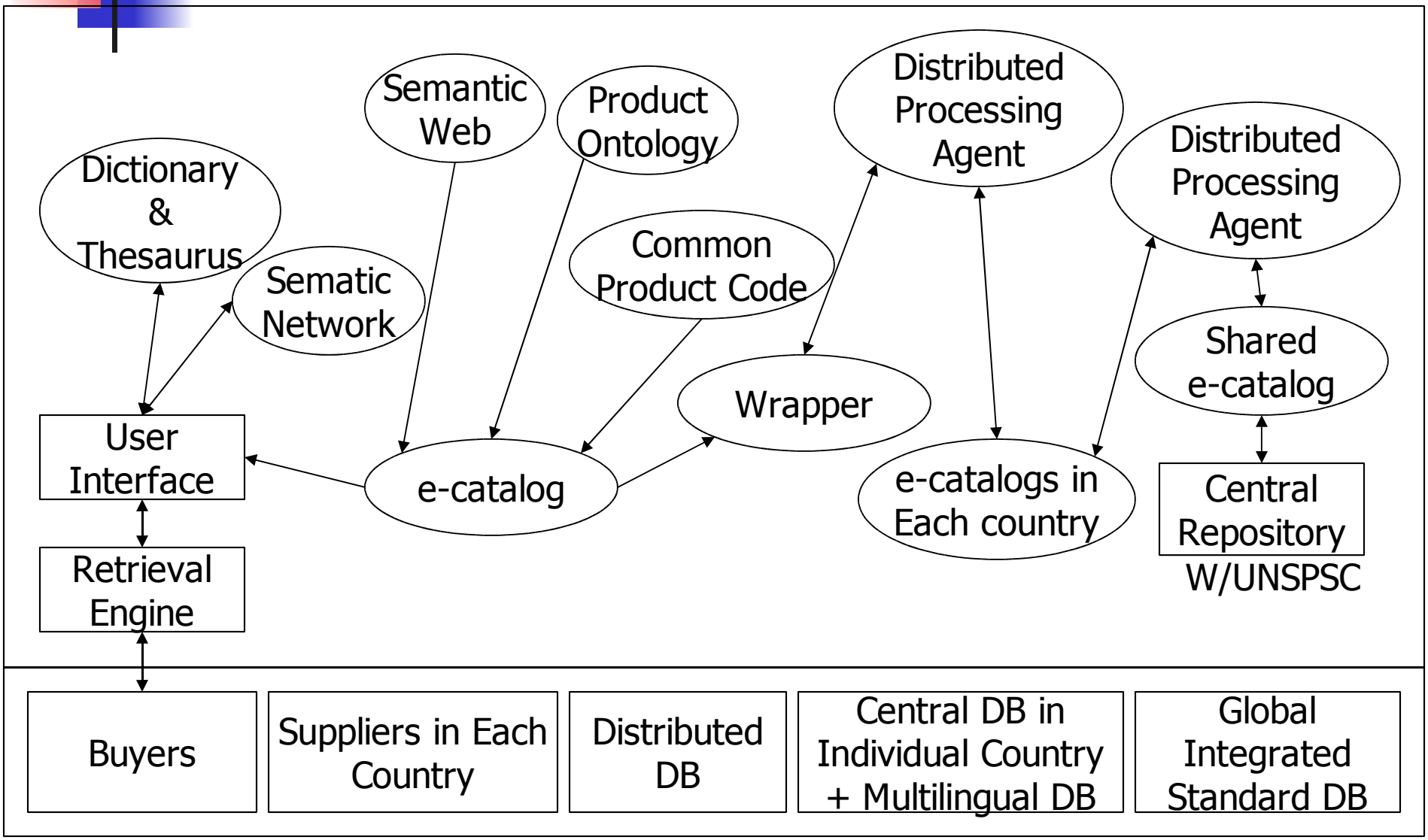


Fig. 4. Architecture of A Product Retrieval Agent

Ⅳ. *Implementation of the Agent for Product Retrieval*

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.

► We define several functions of the algorithms related to the multilingual product retrieval agent.

- **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 for 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) from SEMANTIC_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

Table 3

	# of product retrieval	# of success retrieval	Success Ratio(%)
PN in UNSPSC	83	83	100
Synonyms of PN	27	23	85.19
Homonyms of PN	24	24	100
Polysemys of PN	21	19	90.48

Table 3. Success Ratio of Product Retrieval

Ⅴ. 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.

► Semantic web mechanisms, presented in RDF graphs, provide new opportunities for web product retrieval.

► 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.

Future Works

► To update synonym sets of the semantic networks for convenient interfaces, and to update coined words in the UNSPSC system and the semantic networks. .

► The product names in the UNSPSC system should be completed with cooperation of suppliers and venders.

► Each country should keep its own UNSPSC system translated into its language.