A Mobile Agent System for Meaningful Information Filtering of XML Documents

Yong Hae Kong [1] and In Seok Choi [2]

[1] Division of Information Technology Engineering, Soonchunhyang University
646 Eupnae-ri, Shinchang-myun, Asan-si, Choongnam, Republic of Korea
Tel. +82-41-530-1320, Fax. +82-41-530-1548
e-mail : yhkong@sch.ac.kr

[2] Department of Computer, HongSung Polytechnic College
344 Namjang-ri, hongsung-up, hongsung-gun, Choongnam, Republic of Korea
Tel. +82-41-630-3581, Fax. +82-41-632-5144
e-mail : cis@kopo.or.kr,

Abstract: A mobile agent system that searches for meaningful information from web XML documents is developed. For this, a domain information was conceptualized and a universal DTD generator is developed. A mobile agent with the universal DTD is dispatched to remote sites, extracts meaning information, and transmits the information back to a host. This way of information filtering is able to enhance the quality of information and reduce the network transmission overhead.

1. Introduction

Mobile information filtering can minimize a host processing time and the transmission overhead of network channel. The latter advantage is mainly because documents are processed remotely and only a very small amount of the filtered information is transmitted. Therefore, this mobile information extracting way is more efficient than conventional host document processing methods such as web-spiders. We utilizes mobile agents that can autonomously travel and respond in heterogeneous distributed network on behalf of the user [1]. Such a mobile agent moves to the actual place where the data exists and processes its objective [2].

Searching meaningful information from web XML documents becomes very necessary[3]. But, XML documents may have different structures and properties even for the same information. If all the XML documents conform to the same structure, information search would become significantly easier and more accurate. However, this is not the case at the present time. Although there is an effort to set standards for DTD at W3C and other organizations, it will take much time and effort until the standards are widespread[4]. Besides, it is practically impossible to make standards for all the application areas.

If a particular information area is to be properly conceptualized, we can use this conceptual information to generate a DTD that can be comprehensively applied to various XML documents overcoming their structural limitations.

This work was supported by the supporting program for Information Technology Research Center, Korea Institute of Multimedia Contents Contents & Software in Korea.

Following is the development process of the meaningful information filtering mobile agent system: First, we conceptualized the general representation for application domain information in order to deduce the meaningful information in various XML documents. Then, we constructed a DTD generator that automatically generates a universal DTD based on this concept level information, and an inference engine that deduces user's query makes information search possible. Figure 1. is the configurattion of the semantic information search mobile agent system.

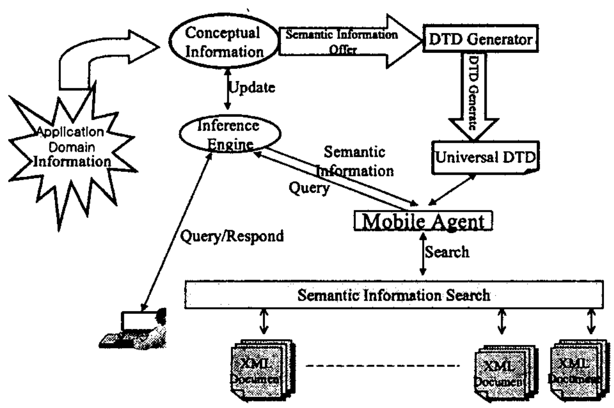

Figure 1. Mobile Agent System for Meaningful Information Filtering

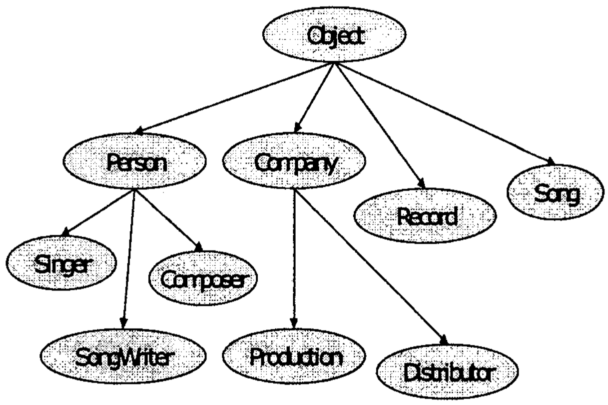

Figure 2. Concept Hierarchy of 'Music CD'

2. Forming Conceptual Information

Information of conceptual level enables sharing common information for many application programs. Concepts can be represented by a hierarchical structure and their attribute relations[5]. The example of Figure 2 shows a hierarchical structure of 'Music CD' concepts. The concept 'Person' has sub-concepts 'Singer', 'SongWriter', and 'Composer'. Sub-concepts inherit all the attribute of a super-concept 'Person'.

Attributes are defined by atomic values STRING and NUM, and/or relations to other concepts. The example of figure 3 shows that attributes represents a relation to other concepts. The value of attribute 'trackList' is defined to concept 'Song'. With these attribute relations, concepts are semantically inter-related.

Person	
name	STRING
gender	STRING
skill	STRING

Singer	
singerName	STRING
song	Song
recordsAlbum	Album

Solo

Group	
member	Singer
position	STRING

Composer	
compose	Song

SongWriter	
write	Song

Song	
songTitle	STRING
sungBy	Singer
recoreded Album	Album
composeBy	Composer
writeBy	SongWriter

Album	
albumTitle	STRING
singer	Singer
trackList	Song
distribution	STRING

Figure 3. Attributes of 'Music CD' Concepts

3. Generating Universal DTD from Conceptual Information

We implemented a universal DTD generator that derives conceptual relations from the given concept hierarchy and their relations. The DTD generator consists of two processors as in figure 4. The first processor is Ontology processor. Ontology processor parses concept information as input. It creates concept and attribute classes. The second processor is Generating DTD processor. The processor generates a universal DTD from concepts and attribute classes.

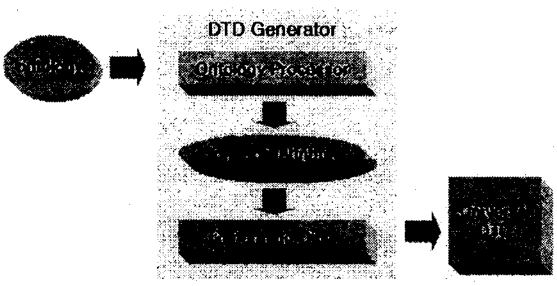

Figure 4. DTD Generator

The general idea of DTD generator is as follows: Each concept is mapped to an element type in the DTD. For each attribute of these concepts, DTD generator defines a sub-element and an XML attribute for the concepts element. If the attribute represents a relation to another concept, the attribute element has content as the respective concept element, otherwise its content model is simply PCDATA. The DTD generator generates a universal DTD by the following four operations.

1. ENTITY generation from the hierarchical structure.
2. ELEMENT generation from the concepts and the attributes.
3. ATTLIST generation from the attributes of concept.
4. ELEMENT generaton from attribute and relation of the other concepts

By applying the above operations, complex attribute relations are derived and finally a universal DTD is generated. Generated universal DTD can be applicable to several documents that have different structures. Figure 5 is a partial list of the generated universal DTD from the above conceptual information.

```
<!—Generated DTD base on the ontology -->

<!-- entities for realizing the is-a hierarchy -->

<!ENTITY % Object "Object | Song | Record | Company | Distributor |
Production |
   Person | Composer |SongWriter | Singer" >
<!ENTITY % Person "Person | Composer | SongWriter |Singer" >
                    :
                    :
<!-- element declaratioins for ontology concepts -->
<!ELEMENT Object (#PCDATA)* >
<!ELEMENT Person (#PCDATA | name)* >
<!ELEMENT Singer (#PCDATA | name | singerName | album)* >
                    :
                    :
<!-- ATTLIST declaratioins for ontology attributes -->
<!ATTLIST Object >
<!ATTLIST Person
        name      CDATA    #IMPLIED >
<!ATTLIST Singer
        name      CDATA    #IMPLIED
        singerName         CDATA    #IMPLIED
        album     CDATA    #IMPLIED >
                    :
                    :
<!-- element declaratioins for ontology attributes -->
<!ELEMENT name (#PCDATA) >
<!ELEMENT singerName (#PCDATA) >
<!ELEMENT album (#PCDATA | %Record;)* >
<!ELEMENT write (#PCDATA | %Song;)* >
                    :
                    :
```

Figure 5. Generated Universal DTD (Partial)

4. Inference Relation of Attribute and Concept

Eventhough XML documents have dissimilar structures, meaning search is possible if they are from the same concepts. We can infer their relation from conceptual level and attribute. For example, if we know the fact 'James album Millennium', we are able to search "Which person P have album about Millennium?". Even though the given fact does not have 'Singer' keyword in it, we can infer 'James' must be a 'Singer' from the fact 'James album millennium'. That is, the given fact contains information about 'singer'. We can infer a new fact 'Millemnnium singer James'. This inferred implicit relation between attributes of concept is shown in figure 6.

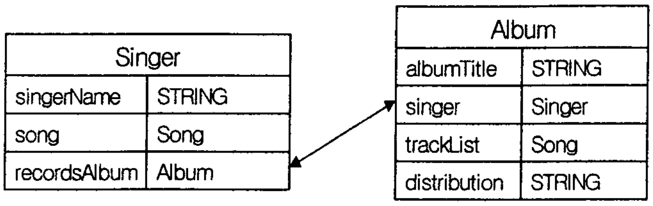

Figure 6. Inferred Relation between Attributes and Concepts

Such inferred knowledge provides a way to search meaningful information from XML documents. Concepts are attribute-related among concepts as in figure 7.

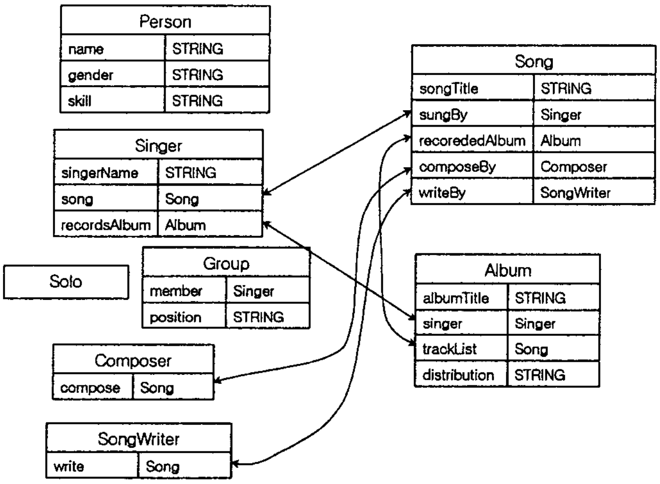

Figure 7. Relational Attributes of Concepts

5. Implemention of Meaning Filtering Mobile Agent System

Mobile agent is a program that moves and reacts autonomously on hetrogeneous distributed network. A mobile agent is created with the name and address, which is called a context. The same mobile agent from the original one can be cloned having the same context. The cloned one may be dispatched from the current context to a target context and runs remotely. An agent may be deactivated, activated, disposed or removed if necessary. Moblie agent operations are illustrated in figure 8.

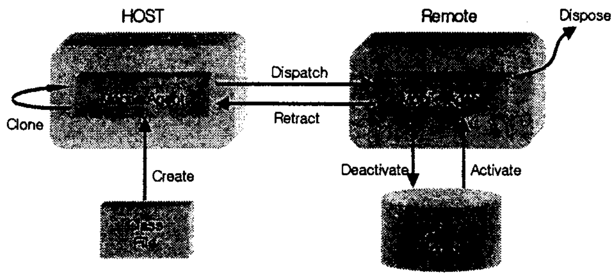

Figure 8. Mobile Agent Operations

Figure 9 is a diagram of information filtering mobile agent system. The mobile agent is dispatched to remote sites with universal DTD, filters information, and return to the original host with collected information. The generated DTD can be broadly applied in filtering various XML document types. For examples, the universal DTD covers both of the following two XML documents having different structures.

```
<Album>
  <albumTitle>Hell Freezes Over</albumTitle>
    <singer>Eagles</singer>
    <distribution>Universal Music</distribution>
    <trackList>
      <Song>
        <songTitle>Hotel California</songTitle>
        <writeBy>D.Felder, D.Henly</writeBy>
        <composeBy>G.Frey</composeBy>
      </Song>
    </trackList>
</Album>

<Singer>
  <singerName>Eagles</singerName>
  <recordsAlbum>
    <Album albumTitle="Hell Freezes Over">
      <distribution>Universal Music</distribution>
      <trackList>
        <Song songTitle="Hotel California">
          <writeBy>D.Felder, D.Henly</writeBy>
          <composeBy>G.Frey</composeBy>
        </Song>
      </trackList>
    </Album>
  </recordsAlbum>
</Singer>
```

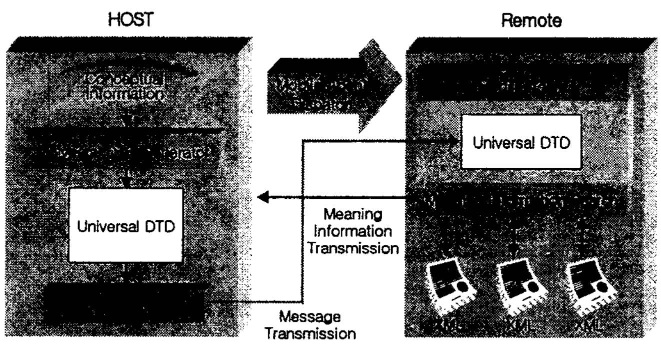

Figure 9. Meaning Filtering Mobile Agent System

Moblie information filtering minimizes the host document processing time and the network transmission overhead through distributed information processing and minimum data transmission

6. Conclusions

Nowadays, the need for web application programs to access XML documents is ever inceasing. But, XML documents of a domain may contain a lot of varieties. Conceptual approach can overcome the structural dissimilarity in a certain degree. Therefore, we conceptualized an application domain and generated a univcrsal DTD that can be applicable to broad range of the domain documents. The generated universal DTD is able to overcome the structural limitation in filtering meaningful information of XML documents. We also used a mobile agent system in filtering remote XML documents to minimize the transmission overhead on networks as well as the processing overhead of the host.

We are now pursuing a further study on evaluating the efficieny of using mobile agents in filtering remote data and on developing DTD-to-DTD matching algorithm.

Reference

[1] H.suk and C.Kim, Agent Concept and Application, Korea Multimedia Society Vol.3, no.2, pp29-39, 1999.

[2]http://aicore.chungbuk.ac.kr/~dobest/document/Agents/AmEC/amec.html

[3] Richard Anderson, Professional XML, Wrox Press, 2000.

[4] W3C, Extensible Markup Language(XML) 1.0 Specification, http://www.w3.org/TR/REC-xml, 1998.

[5] A. Gomez-Perez and R. Benjamins "Applications of Ontologies and Problem Solving Methods," AI Magazine, Vol. 20, No. 1, pp. 199-222, 1999.