

비정형 XML 문서에서의 의미정보 검색을 위한 이동에이전트 시스템

공용해*, 최인석**, 이경수*
*순천향대학교 정보기술공학부,
**홍성기능대학 전자계산학과
e-mail:cis@kopo.or.kr

A Mobile Agent System for Meaningful Information Filtering for XML Documents

Yong Hae Kong* and In Seok Choi** Kyeung Soo LEE*

*Division of Information Technology Engineering,
Soonchunhyang University

**Department of Computer, HongSung Polytechnic College

요약

본 연구에서는 인터넷에 존재하는 다양한 XML 문서에서 의미정보를 검색하기 위한 의미정보수집 이동에이전트 시스템을 개발하였다. 이동에이전트는 실행 프로그램이 이기종 분산 환경의 네트워크에서 자율적으로 이동 및 반응하며 실제로 데이터가 존재하고 있는 장소로 이동하여 목적을 수행한다. 의미정보수집 이동에이전트 시스템의 연구를 위하여, 정보를 개념화하고 포괄적 DTD를 자동으로 생성할 수 있는 DTD생성기를 개발하였으며 의미정보를 추론할 수 있는 추론알고리즘을 연구하였다. 개발된 의미정보수집 이동에이전트 시스템은 정보가 존재하는 원격지 사이트에 파견되어 비정형 XML 문서를 대상으로 포괄적 DTD와 추론엔진을 이용하여, 의미정보를 추출하고 전송하는 임무를 수행한다. 따라서, 의미정보수집 이동에이전트 시스템을 이용한 정보수집은 정보의 질을 향상시키고 네트워크의 부하를 감소시킨다.

1. 서론

사용자에게 다양한 양질의 정보를 제공하기 위하여, 많은 인터넷 사이트에 존재하는 XML 문서들을 검색할 필요가 있다[1]. 그러나, 인터넷상에 존재하는 XML 문서들은 동일한 정보에 대하여 서로 다른 구조와 속성을 지니고 있다. 만일, 사이트가 제공하는 XML 문서들이 동일한 구조를 따를 경우 정보 검색은 매우 쉽고 정확해질 수 있지만, 현실적으로는 그렇지 못하다. 현재 W3C 등에서 XML 문서의 구조를 정의하는 DTD의 표준화 작업을 진행하고 있으나 표준화된 DTD가 보편화되기까지는 많은 시간과 노력이 필요하며, 수많은 정보에 대해 모든 DTD를 표준화하기는 사실상 어려

운 일이다[2].

만약 특정 정보가 적절히 개념화되어 표현된다면, 이러한 개념 수준의 정보를 이용하여 비정형 XML 문서에 포괄적으로 적용할 수 있는 DTD를 생성함으로써 구조적 제약을 극복할 수 있고, 이러한 포괄적 DTD를 이용하여 비정형 XML 문서들에 대한 접근이 가능하게 된다. 또한 개념 수준의 정보는 사용자의 요구사항을 개념적 질의로 변환함으로써, 문서내의 포함된 의미정보를 추출해 낼 수 있게 한다.

원격지에 존재하는 모든 XML 문서들을 호스트로 전송 받아 정보 추출작업을 실행한다면 많은 시간 지연과 네트워크의 부하를 유발하므로 본 논문에서는 사용자를 대신하여 이기종 분산 환경의 네트워크에서 자율적으로 이동 및 반응하는 소프트웨어인 이동에이전트를 활용하였다[3]. 이러한 이동 에이전트는 실행 프로그램이 실제로 데이터가 존재하고 있는 장소로 이동하여 목적을 수행한다[4].

본 연구는 한국소프트웨어진흥원의 IRTC 사업에 의해 수행된 것임.

인터넷상에 넓게 분포되어 있는 비정형 XML 문서들에 대한 의미정보수집 이동 에이전트 시스템의 개발 과정은 다음과 같다. 먼저 다양한 XML 문서에서 의미 정보를 추론하기 위하여 특정 정보에 대한 일상의 표현을 개념화한다. 이러한 개념 수준의 정보를 이용하여 포괄적 DTD를 자동으로 생성하는 DTD생성기를 개발하고 또한 사용자의 질의와 정보의 의미를 추론하여 의미정보의 검색이 가능하게 하는 추론 엔진을 개발한다. 그림 1은 의미정보 검색을 위한 이동에이전트 시스템의 구성도이다.

본 논문의 구성은 2장에서 개념 수준의 정보에 대해서 기술하며, 3장에서는 개념수준의 정보로부터 포괄적 DTD를 생성하는 방법과 의미 추론에 관하여 설명한다. 4장에서는 비정형 XML 문서에서의 의미정보 검색을 위한 이동에이전트 시스템을 구현하며, 5장에서 결론을 맺는다.

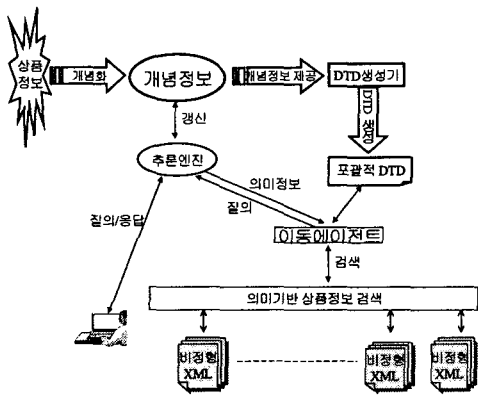


그림 1 의미정보 검색을 위한 이동에이전트 시스템의 구성도

2. 개념 수준의 정보

인터넷상의 많은 사이트에서 제공되는 XML 문서들은 서로 다른 구조와 속성을 가지고 있기 때문에 정보를 검색하는데 어려움이 있다. 그러나 개념화된 정보를 이용하면 비정형 XML 문서에 접근할 수 있으며, 정보의 의미를 추론할 수 있어 정보 검색이 수월하게 된다. 또한 개념 수준의 정보는 사용자와 응용 시스템들 사이에 전달될 수 있는 정보 영역의 공유 개념과 공통된 지식을 제공하여, 웹 응용프로그램들 사이의 지식을 공유하고 일관된 처리를 가능하게 한다. 개념 수준의 정보는 지식의 공유와 재사용을 촉진하기 위하여 인공 지능 분야에서 사용되고 있다[5]. 정보를 개념화하기 위해서는 계층적 구조화와 속성화가 필요하다.

계층적 구조화는 정보를 표현하기 위한 일상적인 개념의 포함관계를 계층적 구조로 표현한다. 그림 2는 음악 CD를 표현하는 계층적 구조의 예이다. 이 계층적 구조는 사람(가수)으로 분류된 사이트뿐만 아니라 배포 회사별로 분류된 사이트에서도 정보를 검색할 수 있도록 표현한 개념적 계층 구조이다. 그림 2에서 보여지듯이 개념 'Person'은 하위 개념으로 'Singer', 'SongWriter', 'Composer'를 가지며 하위개념은 상위개념 'Person'의 모든 속성을 상속받는다.

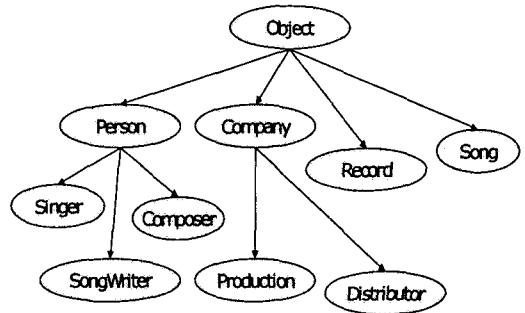


그림 2 상품정보의 개념적 계층구조의 예

속성화는 계층적 구조로 표현된 각 개념들이 가질 수 있는 속성과 상호관계를 표현한다.

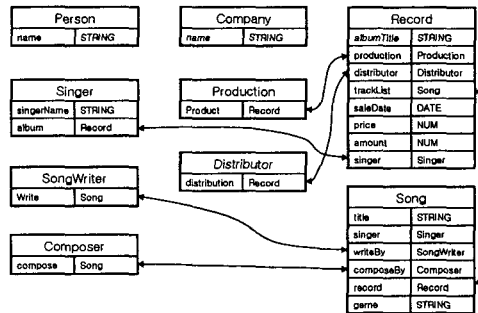


그림 3 정의된 개념의 속성과 상호관계

그림 3의 예는 정의된 개념들 사이의 속성과 상호관계를 정의하였다. 개념 'Record'는 'trackList'를 비롯한 여러 개의 에트리뷰트(attribute)와 그에 대한 값으로 정의되어 있다. 'trackList'는 'Record'의 에트리뷰트로서 개념 'Song'을 값으로 갖는다. 개념 'Record'의 에트리뷰트 'production'과 개념 'Production'의 에트리뷰트 'product'는 상호 관계를 규정하여 같은 정보로 해석된다.

3. 개념 수준 정보의 활용

특정 정보를 개념화함으로써 비정형 XML 문서의 구조에 독립적으로 접근할수 있는 포괄적 DTD를 생성할 수 있으며, 개념을 추론하여 의미정보를 추출할 수 있으므로 정보의 검색이 수월해진다[6].

3.1 개념 수준 정보로부터 포괄적 DTD 생성

DTD생성기는 개념 수준의 정보에서 제공되는 각 개념들의 상호 관련성과 정의된 속성을 토대로 포괄적 DTD를 생성한다. 우선 XML 문서의 계층적 구조를 정의하고 엔터티(ENTITY)에 할당할 수 있는

```

<!-- 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" >
<ENTITY % Singer "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;)* >
<ELEMENT compose (#PCDATA | %Song;)* >
<ELEMENT companyName (#PCDATA) >
<ELEMENT product (#PCDATA | %Record;)* >
:
:
    
```

그림 4. 개념 수준의 포괄적 DTD

실제 값의 속성을 정의한다. 그 다음, 엘리먼트가 가져야할 관련성, 항목, 속성들을 정의한다. 개념정보를 기반으로 생성된 포괄적 DTD는 유효한 정보를 가진 다양한 형태의 XML 문서에 구조적인 제약을 받지 않고 적용할 수 있으며, 접근을 용이하게 해준다. 그림 4는 개념 수준의 정보를 이용하여 생성한 포괄적 DTD이다.

3.2 개념 수준 정보로부터 의미 추론

같은 종류의 정보를 갖는 사이트의 비정형 XML 문서에서 동일한 정보를 검색하기란 어려운 일이다. 그러나 개념 수준의 정보에서 정의된 에트리뷰트의 관계를 이용하여 새로운 정보를 추론할 수 있다[7]. 예를 들면 “ ‘김갑동’은 ‘밀레니엄’이라는 타이틀로 음악 CD를 발표했다.” 라고 정의했을 때, 이 정의에서 어떤 사람이 음악 CD를 발표한 사람을 찾는 것은 어려운 일이 아니다. 그러나, 위의 정의에서 ‘가수’라는 직접적인 정보가 없기 때문에, ‘가수’를 검색하고자 했을 때에는 ‘김갑동’을 찾지 못한다. 하지만, 음악 CD타이틀을 발표했다는 사실에서, ‘김갑동’이라는 사람이 ‘가수’라고 추측할 수 있다. 즉, “음악 CD타이틀을 발표한사람은 당연히 가수다” 라는 일상적인 개념이 정의가 있다면 “ ‘김갑동’은 가수다” 라는 새로운 개념을 추론하여 정의할 수 있다. 이처럼 새로운 개념을 정의하고 추론정보를 만들어내는 것을 추론 엔진이라 한다. 추론 엔진은 클라이언트에서 요구하는 질의를 적당히 추론하여 인터넷상에 존재하는 비정형 XML 문서에서 표현되지 못한 의미정보를 추출할 수 있게 한다.

4. 의미 정보 수집 이동에이전트 시스템 구현

이동 에이전트란 사용자를 대신하여 이기종 분산 환경의 네트워크에서 자율적으로 이동 및 반응하는 소프트웨어를 의미하며 자율성, 지능성, 이동성과 대행성, 비동기성 등의 특징을 가지고 있다. 원격지의 큰 데이터를 호스트에 저장할 때 문서의 전체를 전송하는 것보다 데이터들을 지역적으로 처리한 뒤 필요한 정보만을 보내는 것이 더욱 효율적이다. 이동 에이전트 환경에서는 실행 프로그램이 실제로 데이터가 존재하고 있는 장소로 이동하여 목적을 수행하게 되며 일반적으로 수행 완료 후에는 소멸하거나 홈 사이트로 돌아오게 되므로 효율적인 네트워크 활용이 가능하다. 또한 파견된 이동에이전트는 원격지에서 수행되므로 호스트에 대해서 독립적

으로 실행된다.

본 연구에서 개발한 DTD생성기는 특정 개념에 대한 모든 XML 문서에 적용할 수 있는 포괄적 DTD를 생성하여 구조적인 제약을 극복하였다. 개념 수준의 정보로부터 생성된 추론엔진은 문서에 내포된 의미를 추론하여 보다 다양하고 정확한 정보 검색을 지원한다. 이동에이전트는 포괄적 DTD와 추론엔진을 원격지 사이트에 이동시켜서 임무를 수행하게 하며, 검색된 의미정보를 원래의 호스트로 전송한다. 모든 정보 수집 활동이 완료되면 호스트로 귀환 또는 소멸한다. 그림 5는 의미정보수집 이동에이전트를 도식화하였다.

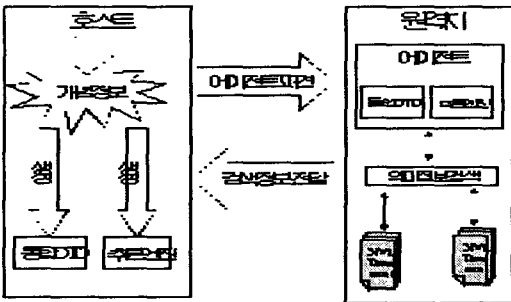


그림 5 의미정보수집 이동 에이전트 시스템

5. 결론

최근 인터넷에는 XML 문서를 기반으로 하는 많은 사이트가 있으며, 정보를 표현하는 XML 문서의 구조도 매우 다양하다. 따라서, 정보 검색자가 원격지의 사이트에서 제공되는 비정형 XML 문서들로부터 동일한 의미의 정보를 검색하기 위해서는 정확한 분석과 해석이 필요하다. 본 논문에서는 특정 내용에 대한 비정형 XML 문서에서 정보를 검색하기 위하여 정보를 개념화하였으며, 포괄적 DTD를 생성해주는 DTD생성기와 의미를 추론할 수 있는 추론엔진을 개발하였다. 생성되는 포괄적 DTD는 XML문서에 접근할 때 발생하는 구조적인 제약을 극복할 수 있었으며, 또한 추론 엔진을 이용한 의미 정보 검색은 보다 다양하고 신뢰성 있는 정보를 검색할 수 있게 했다. 또한 이동에이전트를 이용하여 이기종 분산 환경의 네트워크에서 자율적으로 이동 및 반응하면서 표현되지 않은 정보를 유추하여 검색할 수 있으므로 키워드 검색 방법보다 효율적이다.

의미 정보수집 이동 에이전트 시스템은 사용자의 요구에 의하여 원격 사이트에서 필요한 의미 정보만

추출하여 전송하므로 최소한의 호스트의 실행시간만 필요하며, 네트워크의 부하를 줄일 수 있었다.

[참고문헌]

- [1] Richard Anderson 외, "Professional XML", Wrox Press, 2000.
- [2] W3C, Extensible Markup Language(XML) 1.0 Specification, <http://www.w3.org/TR/REC-xml>, 1998.
- [3] 석황희, 김인철 "이동 에이전트의 개념과 응용", 한국멀티미디어학회, 제3권, 제2호, pp29-39, 1999
- [4]<http://aicore.chungbuk.ac.kr/~dobest/document/Agents/AmEC/amec.html>
- [5] A. Gomez-Perez and R. Benjamins "Applications of Ontologies and Problem Solving Methods," AI Magazine, Vol. 20, No. 1, pp. 199-222, 1999.
- [6] M. Erdmann and R. Studer, "How to Structure and Access XML Document with Ontologies," Data & Knowledge Engineering, Vol. 36, No.3, pp. 317-335, 2001.
- [7] <http://www.aifb.uni-karlsruhe.de/WBS/broker>