

시멘틱웹 환경에서의 이동 에이전트 기반 자동 정보 공유 시스템

노경신^o 유영훈 조근석

인하대학교 컴퓨터공학부

nohks^o@eslab.inha.ac.kr, yhyu@eslab.inha.ac.kr, gsjo@inha.ac.kr

Automatic Information Sharing System Based on Mobile Agent in Semantic Web

Kensin Noh^o Young-Hoon Yu Geun-Sik Jo
School of Computer Science & Engineering, Inha University

요 약

현재의 웹은 동일한 답변을 요구하는 중복된 질문들과 온톨로지의 부재로 인해 심각한 자원 낭비의 문제를 안고 있다. 이러한 문제를 해결하기 위하여 이동 에이전트를 이용한 정보 공유 시스템이 대안으로 제시 되었으나, 서버 수의 급증에 따른 자료 처리시간 증가와 사용자의 의도와는 전혀 다른 답변을 가져오는 등의 문제를 여전히 안고 있다. 따라서 본 논문에서는 이를 해결하기 위해 RDF와 RDFS 기반의 온톨로지를 활용한 시스템을 제안한다.

본 논문에서 제안하는 시스템은 자바 기반의 이동 에이전트와 RDF, RDFS 기반의 온톨로지로 구성된다. 분산 이형 환경에서 독립적으로 운영되던 질의 응답 게시판을 검색할 때, 이동 에이전트는 코드의 이동으로 네트워크의 부하를 방지하고 구조화된 온톨로지는 질문에 대한 도메인을 확인함으로써 효율적인 이동 에이전트의 라우팅을 제공한다. 본 논문에서 제안한 시스템을 전자상거래에서 빈번하게 발생하는 상품에 대한 소비자들간의 질의 응답 게시판에 적용하여 시멘틱웹 환경에서 도메인 추론을 통한 에이전트 라우팅과 가중치를 부여하지 않은 클래스의 속성 비교를 이용한 검색 방법이 시간의 단축과 네트워크 부하의 감소, 그리고 기존 시스템의 가장 큰 문제점이었던 정확도(precision)와 재현율(recall)을 개선하는데 효과적이고 타당하다는 것을 보여준다.

1. 서 론

현재 인터넷 상에는 수많은 커뮤니티 사이트가 있고, 커뮤니티 안에는 독립적으로 운영되고 있는 질의 응답 게시판을 존재한다. 이러한 게시판은 서로 다른 플랫폼을 이용하여 독립적으로 운영하고 있는 경우가 많기 때문에 수많은 같은 질문이 반복될 수 있고 답변의 효율적인 공유와 활용이 불가능하다. 이러한 문제의 해결을 위하여 분산 이형 환경에서의 이동 에이전트 기반의 정보 공유 시스템 [1]에서는 효율적인 질의/응답 게시판을 공유 문제를 해결하기 위한 하나의 방법론을 제시한다. 그러나 기존의 시스템은 질문자의 질문에 해당하는 답변들을 검색하기 위하여 키워드 사전(Keyword Dictionary)을 통해 추출된 키워드를 불리언 검색(Boolean Search)에 적용하였기 때문에 사용자의 질의와 의도가 같은 질의에 대한 정확한 응답을 찾아내는 데 한계를 가지고 있다. 또한 공유되고 있는 게시판을 많아질수록 각각의 서버를 에이전트가 돌아다녀야 하기 때문에 검색 시간이 늘어나게 된다.

이에 본 논문에서는 기존 시스템이 가지고 있는 위와 같은 단점을 개선하기 위해 최근 대두되고 있는 XML과 RDF (Resource Description Framework), RDF schema로 구성된 온톨로지(ontology)를 이용한 시멘틱웹(Semantic Web) 기술 기반으로 전자상거래에서 빈번하게 발생하는 상품에 대한 질의 응답의 효과적인 공유와 검색 방법을 제시한다.

2. 관련 연구

2.1 이동 에이전트의 특성

이동 에이전트는 특정 목적을 수행하기 위한 자율적인 프로세스의 하나이다. 비단 수동적인 작업 뿐 아니라 환경에 따라 능동적 혹은 독립적인 활동을 수행한다. 이동 에이전트의 특성은 크게 자율성(Autonomy), 반작용성 (Reactivity), 사회성(Social Ability), 이동성(mobility), 지능(intelligence) 선행성(Pro-activeness) 등으로 나눌 수 있다.

또한 이동 에이전트는 코드 이동을 통한 트래픽 감소, 네트워크 대기(network latency)극복, 에이전트 통신 프로토콜(ATP : Agent Transfer Protocol)에 의한 캡슐화, 다양한 환경에 대한 능동적인 반응과 같은 장점을 가진다 [4].

2.2 시멘틱웹(Semantic Web)

웹 자원의 지식화를 통한 효과적인 정보의 표현과 검색의 기술로 주목되고 있는 시멘틱웹은 현재 W3C의 주도하에 그 개발이 활발하게 이루어지고 있다. 시멘틱웹과 함께 사용의 중요성이 대두되고 있는 웹 온톨로지 언어인 RDF(Resource Description Frame Work)는 데이터가 가진 요소와 속성간의 관계를 기계가 이해할 수 있게 하기 위해 설계된 웹 자원 기술의 일반 언어이다. RDF는 주어(Subject), 목적어(Object), 술어 (Predicate)의 개념을 이용하여 자원을 기술한다. RDF에서 확장된 RDFS(RDF Schema)는 클래스 속성과 도메인 및

자원의 제어를 위해 설계 되었으며 현재 웹 자원의 지식화에 근간을 마련하는 데 많은 가능성을 제시한다 [5, 6].

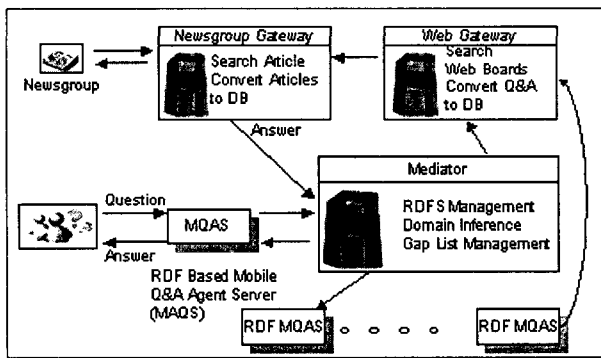
2.3 질의 응답 검색 시스템

질의 응답 검색 시스템으로는 SMART information retrieval system을 이용하여 데이터베이스에 저장된 FAQ를 자연어로 검색하는 시스템인 FAQ(Frequently Asked Question) Finder System 이 있다 [2, 3].

FAQ Finder System은 USENET뉴스 그룹의 자료들을 여과하고 요청 계층(Request Hierarchy)에 의해 질문의 종류를 분류하며 Word Net의 단어 사전을 이용하기 때문에 보다 정확한 답변을 얻을 수 있다. 하지만 사람의 의견을 중요시 하는 전자상거래의 질문과 답변에 적합하지 않고, 모든 답변을 정제해 놓아야 한다는 단점을 가지고 있다.

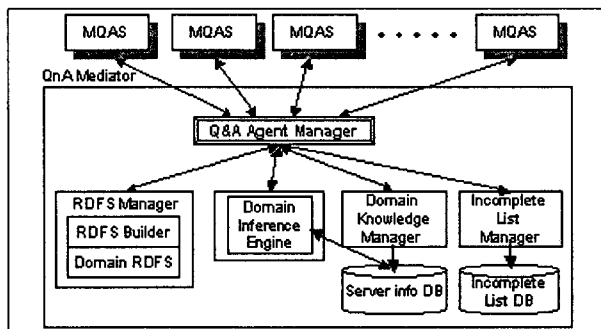
3. 시스템 구조

3.1 Question Finder System의 구조



[그림 1] Question Finder System

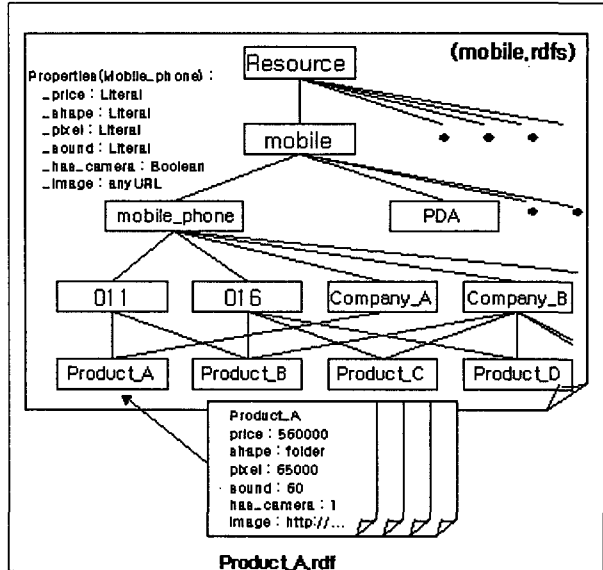
[그림 1]에서와 같이 사용자가 질문을 던지면 Agent Server는 RDF형식으로 재구성한 다음 중계기(Mediator)로 전달한다. 또한 중계기는 사용자의 질문에 대한 Domain을 추론하고 해당 MQAS(Mobile Agent Q&A Server)로 Agent를 보내 답변이 존재하는 이전 사용자의 가장 흡사한 질문을 검색한다. 중계기 구조는 [그림 2]와 같다.



[그림 2] Question Finder의 중계기(mediator) 구조

3.2 클래스의 RDFS Domain Knowledge구성

[그림 3]에서 Mediator의 Domain knowledge에 정의 되는 Mobile 클래스를 간단하게 정의하였다.



[그림 3] Mobile Class에 관한 RDFS 트리구조

mobile.rdfs는 핸드폰에 관한 클래스와 속성들을 정의 하며, Product_A.rdfs는 Product_A에 대한 인스턴스를 나타낸다. Mediator는 사용자의 질문을 이용하여 질문에 대한 도메인(Class)을 찾아내고 해당 서버로 에이전트를 보낸다. 식(1)에서는 속성을 이용한 클래스의 추론을 베이지안 접근법으로 보여준다 [7].

$$P(C_j | B_i) = \sum_{B_k \in \text{children}(B_i)} P(C_j | B_k) \cdot \frac{P(B_k)}{P(B_i)} \quad (1)$$

P = 해당 클래스가 도메인이 될 확률
 C_j = 속성을 포함하고 있는 클래스
 B_i = leaf node가 없는 RDFS의 property
 B_k = 질문에 포함된 속성

3.3 질문의 유사도 결정

RDF 형식으로 인스턴스 된 사용자 질문은 클래스와 속성(Property)으로 나뉘게 된다. 식(2)에서는 사용자의 질문에 포함된 클래스와 속성을 가지고 문서간의 유사도를 구하게 되는데 속성의 가중치와 빈도 수를 고려하지 않는다. 고려 사항은 클래스의 유무와 속성 노드의 배치이다. 속성 비교값은 총 4 가지로 구분할 수 있고 식(3)에서와 같이 속성 유사도를 계산할 수 있다.

$$Sim = \sum_{i=1}^{n_c} \frac{1}{n_c} \left(n_p - \sum_{j=1}^{n_p} D_j \right) \quad (2)$$

Sim = 질문의 유사도
 n_c = 질문에 명시된 클래스의 수
 (단, 질문에 클래스가 없으면 n_c = 1)
 n_p = RDFS에 정의된 속성의 총 개수

여기서 D값은 두 질문이 가지고 있는 속성의 유사도를 판단 하기 위한 척도이며,

$$D_j = \left(\frac{D_{qi} - D_{ai}}{n_p + 1} \right)^{|D_{qi} - D_{ai}|} \quad (3)$$

D_{qi} = 현재 질문에 속한 클래스의 i 번째 속성노드
 D_{ai} = 지난 질문에 속한 클래스의 i 번째 속성노드

와 같이 서술할 수 있다. 여기서 Sim 의 값이 최대일 때 사용자의 질문과 가장 흡사한 문서를 찾을 수 있다.

4. 실험 및 결과

4.1 실험 환경 및 데이터 집합

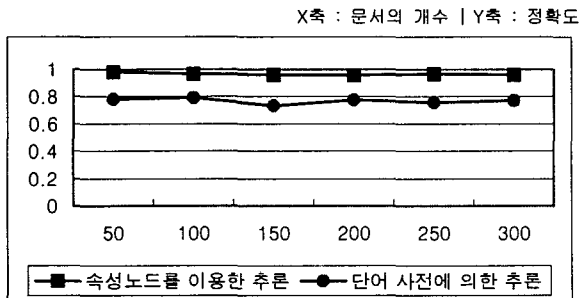
실험을 통한 성능 평가를 위해 Windows2000, JAVA2, JSP, jdk1.4, jena2.0, Tomcat4 Access2000, Aglet2.0을 사용하여 구현 하였으며, 실험 및 평가는 펜티엄4 2GHz, 256MB RAM의 시스템에서 하였다. 또한 실험 데이터는 Naver 지식인 (<http://kin.naver.com>)에서 최근 1주간의 핸드폰과 PDA 구매에 관한 질문 300개를 수집하고 각 클래스와 속성들을 추출하여 RDF문서로 구성하였다 [표 1].

[표 1] RDFS의 클래스와 속성 수

	핸드폰	PDA
Classes 1	12	11
Classes 2	22	14
Property	9	10

4.2 실험 결과

[그림 4]는 도메인을 추론하는 방법으로 단어사전을 이용한 카테고리 추론과 본 논문이 제안하는 속성을 이용한 추론의 결과를 나타낸다. 본 논문이 제안하는 방법에서 중계기의 도메인 판단에 대한 정확도는 96.5% ± 3%의 신뢰구간을 가진다. 빠르고 정확한 에이전트의 라우팅을 위한 필요조건인 높은 도메인의 추론율은 [표 2]에서와 같이 서버의 수가 증가할 경우 응답 시간을 크게 향상시킨다.



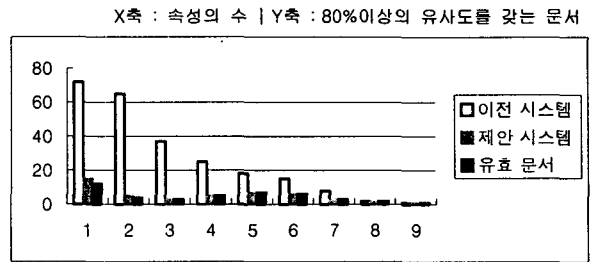
[그림 4] 사용자 질문 수에 따른 Domain 정확도

[표 2] 에이전트 서버의 수에 따른 응답시간 (단위 : ms)

	1	2	3	4
이전 시스템	75	130	176	211
제안 시스템	80	87	88	88

[그림 5]는 각각 속성의 개수를 달리하고 유사도가 80% 이상인 문서를 검색한 결과를 나타낸 것인데, 속성에 따른 재현율의 큰 차이를 보여준다. 예를 들어 "핸드폰" 클래스에 has_camera의 속성만을 비교하였을 경우 본 논문이 제안하는

속성 노드의 탐색법을 이용한 검색과 달리 불리언 검색에서는 카메라를 원하는 사용자의 모든 질문에 대한 답변을 가져오기 때문에 문서의 양은 많지만 필요 없는 경우가 대부분이었다.



[그림 5] 속성 수에 따른 추천 문서의 개수

5. 결론 및 향후 연구

본 논문에서는 Semantic Web 환경 하에서 이동 에이전트와 온톨로지를 이용한 질의 응답 게시판의 공유 및 검색시스템을 제안하였다. 이 시스템은 기존 시스템 보다 정확도(Precision)와 재현율(Recall)을 효과적으로 개선할 수 있었다. 또한 실험을 통하여 RDF와 RDFS를 이용한 온톨로지가 중계기의 에이전트 라우팅에 효율적으로 쓰일 수 있음을 보였다.

하지만 여전히 사용자의 질문에 대한 가중치와 제약 속성(Constraint Property)을 판단함에 있어 효과적인 자연어 처리를 통한 RDF문서 구성이 요구되며, 향후 연구 과제인 자연어 처리와 효과적인 온톨로지 구성을 통한 보다 정확한 검색엔진의 개발과 함께 최근 이슈화 되고 있는 Mobile Web 환경으로의 시스템 확장이 필요하다.

감사의 글

본 연구는 한국과학재단 지정 인천대학교 동북아 전자 물류 연구센터의 지원에 의한 것임.

6. 참고 문헌

- [1] 박제복, "본산 이형 환경에서의 이동에이전트 기반의 정보 공유 시스템", 정보 과학회, 제 29권 제1호, 2002.2
- [2] Buckley, C., "Implementation of the SMART Information Retrieval [sic] System", Technical Report 85-686, Cornell Univ. 1985
- [3] Robin D. Burke, Kristian J. Hammond, & Vladimir A. Kulyukin, "Question Answering from Frequently-Asked Question Files: Experiences with the FAQ Finder Systems", Technical Report TR-97-05, Chicago Univ. 1997
- [4] Michael N. Huhns, Munindar P.Singh, "Readings in AGENTS", Morgan Kaufmann Publishers, Inc., 1998
- [5] Roger L. Costello, David B. Jacobs, "XML Design (A Gentle Transition from XML to RDF)", RDF tutorial DARPA, 2002
- [6] Roger L. Costello, David B. Jacobs, "Inferring and Discovering Relationships using RDF Schemas", RDFS tutorial, DARPA, 2002
- [7] Sushama Prasad, Yun Peng, Tim Finin, Zhongli Ding, "A Tool For Mapping Two Ontologies Using Explicit Information", AAAI Workshop on Meaning Negotiation, 2002
- [8] Kit Hui, Stuart Chalmers, Peter Gray, Alun Preece, "Experience in using RDF in Agent-mediated Knowledge Architectures", AAAI, 2003