

무선 인터넷 환경에서의 온톨로지 기반 정보 공유 시스템

노경신⁰ 유영훈 조근식

인하대학교 컴퓨터공학부

nohks⁰@eslab.inha.ac.kr, yhyu@eslab.inha.ac.kr, gsjo@inha.ac.kr

Information Sharing System Based on Ontology in Wireless Internet

Kensin Noh⁰ Young-hoon Yu Geun-sik Jo

School of Computer Science & Engineering, Inha University

Abstract

Due to recent explosion of information available online, question- answering (Q&A) systems are becoming a compelling framework for finding relevant information in a variety of domains. Question-answering system is one of the best ways to introduce a novice customer to a new domain without making him/her to obtain prior knowledge of its overall structure improving search request with specific answer. However, the current web poses serious problem for finding specific answer for many overlapped meanings for the same questions or duplicate questions also retrieved answer for many overlapped meanings for the same questions or duplicate questions also retrieved answer is slow due to enhanced network traffic, which leads to wastage of resource.

In order to avoid wrong answer which occur due to above-mentioned problem we propose the system using ontology by RDF, RDFS and mobile agent based on JAVA. We also choose wireless internet based embedded device as our test bed for the system and apply the system in E-commerce information domain. The mobile agent provides agent routing with reduced network traffic, consequently helps us to minimize the elapsed time for answers and structured ontology based on our proposed algorithms sorts out the similarity between current and past question by comparing properties of classes.

1. 서론

최근 휴대용 단말기를 이용한 무선 인터넷이 급속도로 발전하고 있으며, 그 응용 범위를 넓히기 위한 연구가 지속되고 있다. 본 논문은 각각의 커뮤니티 사이트 게시판과 기업의 전자상거래에서 빈번히 일어나는 상품에 대한 질의/응답의 공유에 주안점을 둔 무선 인터넷 기반의 시스템을 제안한다. 이러한 전자상거래나 커뮤니티의 게시판은 서로 다른 플랫폼을 이용하여 독립적으로 운영하고 있는 경우가 많기 때문에 수 많은 같은

질문이 반복될 수 있고 답변의 효율적인 공유와 활용이 불가능하므로 많은 검색시간이 소요 된다. 따라서 시간과 문서량의 제약을 받는 무선 인터넷 환경에서 웹에 있는 게시판을 이용하기에는 많은 불편함이 따른다. 본 논문은 웹에서의 효율적인 질의/응답 게시판 공유 문제를 해결하기 위한 분산 이형 환경에서의 이동 에이전트 기반의 정보 공유 시스템[1]의 아이디어를 이용하여 시맨틱웹과 무선 인터넷으로 그 범위를 확장하였다. 기존 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 있고, 공유되고 있는 게시판이 많아질수록 각각의 서버를 에이전트가 돌아다녀야 하기 때문에 검색시간이 늘어나는 단점을 가지고 있다.

이에 본 논문에서는 위와 같은 단점을 극복하기 위해 XML 과 RDF(Resource Description Framework), RDF schema 로 구성된 온톨로지(Ontology)를 이용한 시맨틱 웹(Semantic Web)기술과 무선 인터넷 환경을 기반으로 전자상거래에서 빈번하게 발생하는 상품에 대한 정보의 효과적인 공유와 검색방법을 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 프로토

콜 (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 W3C 의 주도하에 그 개발이 활발하게 이루어지고 있다. 시멘틱 웹과 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. RDF 는 주어(Subject), 목적어(Object), 술어 (Predicate)의 개념을 이용하여 자원을 기술한다. RDF 에서 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 무선 인터넷

무선인터넷이란 휴대폰과 같이 휴대성이 높은 통신기기 또는 PDA 와 같은 정보기기를 이용해 무선으로 인터넷에 접속해 해당 인터넷 서비스를 이용하거나 데이터 통신을 하는 것을 말한다. 본 논문에서는 WAP Forum 에서 제정한 무선망과 인터넷 연동을 위한 프로토콜인 WAP(Wireless Application Protocol)과 XML 기반인 WML(Wireless Markup Language)를 사용하였다 [8, 9].

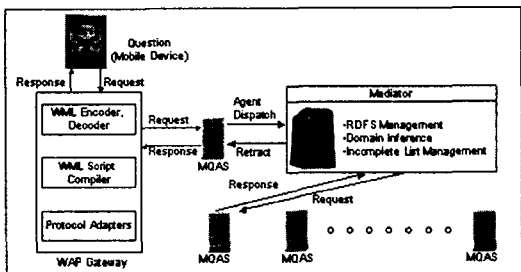
2.4 질의 응답 검색 시스템

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

FAQ Finder System 은 USENET 뉴스 그룹의 자료들을 여과하고 요청 계층(Request Hierarchy)에 의해 질문의 종류를 분류하며 Word Net 의 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.

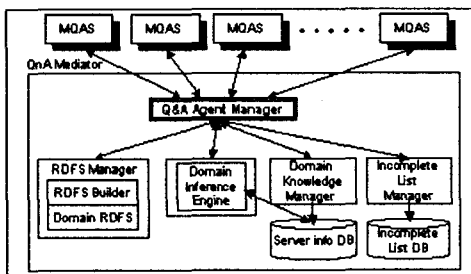
3. 시스템 구조

3.1 Wireless Question Finder System 의 구조



[그림 1] Wireless Question Finder System

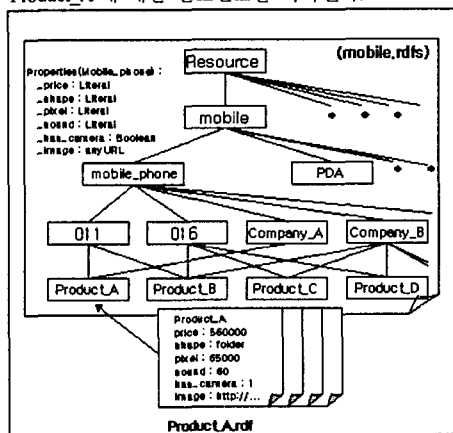
[그림 1]에서와 같이 모바일 디바이스를 사용하여 사용자가 질문을 하게 되면 WAP Gateway 로 전달된다. WAP Gateway 에서 디코딩 된 사용자의 질문은 Agent Server 를 통해 중계기(Mediator)로 보내진 후 RDF 빌더를 거쳐 RDF 형식으로 파싱된다. [그림 2]는 중계기의 구조를 나타내고 있다. 파싱된 RDF 문서는 속성을 이용한 클래스 추론으로 사용자 질문에 대한 의도를 파악하기 위해 도메인을 결정하게 된다. 이어 해당 MQAS(Mobile Q&A Agent Server)로 Agent 를 보내 답변이 존재하는 이전 사용자의 가장 흡사한 질문을 검색한다. 중계기 구조는 [그림 2]와 같다.



[그림 2] Question Finder 의 중계기(mediator) 구조

3.2 클래스의 RDFS Domain Knowledge 구성

[그림 3]에서 Mediator 의 Domain knowledge 에 정의 되는 Mobile 클래스를 간단하게 정의하였다. mobile.rdfs 는 핸드폰에 관한 클래스와 속성들을 정의 하며, Product_A.rdf 는 Product_A 에 대한 인스턴스를 나타낸다.



[그림 3] Mobile Class 에 관한 RDFS 트리구조

Mediator 는 사용자의 질문을 이용하여 질문에 대한 도메인 (Class)을 찾아내고 해당 서버로 에이전트를 보낸다. 식(1)에서는 속성을 이용한 클래스의 추론을 베이즈 접근법으로 보여준다 [10].

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

- P = 해당 클래스가 도메인이 될 확률
- C_j = 속성을 포함하고 있는 클래스
- B_i = leaf node 가 없는 RDFS 의 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 = RDFS 에 정의된 속성의 총 개수

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

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

- D_{q_i} = 현재 질문에 속한 클래스의 i 번째 속성노드
- D_{a_i} = 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 하였고 무선 인터넷 톨킷과의 연동을 위해 J2ME Wireless Toolkit 1.0.2, J2ME MIDP(Mobile Information Device Profile)를 사용하였다. 실험 및 평가는 펜티엄4 2GHz, 256MB RAM 의 시스템에서 하였으며 실험 데이터는 Naver 지식인 (<http://kin.naver.com>)에서 최근 1주간의 핸드폰과 PDA 구매에 관한 질문 300개를 수집하고 각 클래스와 속성들을 추출하여 RDF 문서로 구성하였다 [표 1].

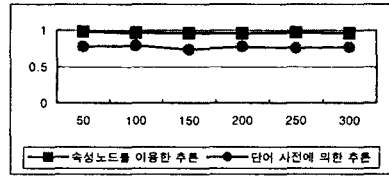
[표 1] RDFS 의 클래스와 속성 수

	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 필요 조건인 높은 도메인의 추론율은 [표 2]에서와 같이 서버의 수가 증가할 경우 응답 시간을 크게 향상시킨다.

X축 : 문서의 개수 | Y축 : 정확도

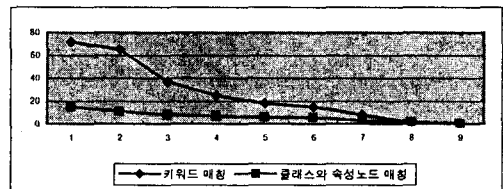


[그림 4] 사용자 질문 수에 따른 Domain 정확도

[표 2] 에이전트 서버의 수에 따른 응답시간 (단위 : ms)

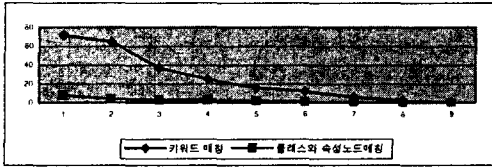
	1	2	3	4
단어사전을 이용한 추론	75	78	176	168
클래스의 속성 비교를 이용한 추론	80	87	88	88

[그림 5]와 [그림 6]은 각각 속성의 개수를 달리하고 유사도가 80%와 90% 이상인 문서를 검색한 결과를 나타낸 것인데, 속성에 따른 재현율의 큰 차이를 보여준다. 예를 들어 “핸드폰” 클래스에 has_camera 의 속성만을 비교하였을 경우 본 논문이 제안하는 속성 노드의 탐색을 이용한 검색과 달리 불리언 검색에서는 카메라를 원하는 사용자의 모든 질문에 대한 답변을 가져오기 때문에 문서의 양은 많지만 필요 없는 경우가 대부분이었다. 하지만 속성노드 탐색의 경우 속성수가 많고 유사도의 임계치(threshold)가 높을 때 불리언 검색에 비해 적절한 답변을 찾지 못하는 문제점을 가지고 있다. 따라서 질문에 따른 적절한 유사도의 임계치값을 설정하는 것이 필요하다.



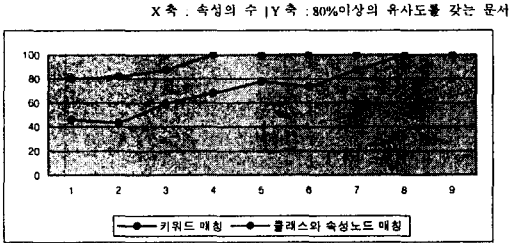
X축 : 속성의 수 | Y축 : 80%이상의 유사도를 갖는 문서

[그림 5] 검색방법에 따른 결과(1)



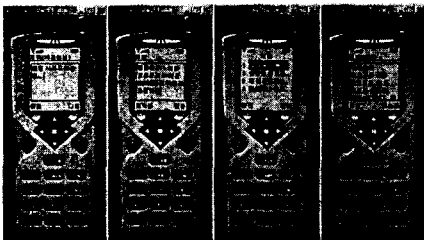
X축 : 속성의 수 | Y축 : 90%이상의 유사도를 갖는 문서
[그림 6] 검색방법에 따른 결과(2)

[그림 7]은 사용자의 질문에 따라 단어 사전과 RDFS로부터 만들어진 문서들의 검색 방법에 따른 정확도(Precision)를 보여준다. 유효문서는 3명의 사용자에게 받은 feedback으로 결정하였으며 총 300개의 문서로부터 얻어진 결과이다. 이 도표에서는 불리언 검색이 속성노드 검색과 비교할 때 더 많은 유사도를 갖는 문서를 가져왔으나, 속성 노드 검색의 경우 상대적으로 높은 정확도를 보인다.



X축 : 속성의 수 | Y축 : 80%이상의 유사도를 갖는 문서
[그림 7] 검색방법에 따른 정확도

4.3 사용자 인터페이스



[그림 8] MIDP를 사용한 시스템의 사용자 인터페이스
[그림 8]은 J2ME에서 동작하는 MIDP를 사용하여 만든 사용자 인터페이스 환경으로 여러 기종의 단말기에서 동작하는 것처럼 시뮬레이션을 할 수 있는 에뮬레이터(Emulator)이다.

5. 결론 및 향후 연구

본 논문에서는 무선 인터넷과 Semantic Web 환경 하에서 이동 에이전트 및 온톨로지를 이용한 질의 응답 게시판의 공유 및 검색시스템을 제안하였고, 실험을 통하여 무선 인터넷 환경에서도 분산되어 있는 여러 서버의

게시판을 효과적으로 쓸 수 있음을 보였다. 무선 인터넷은 일반 인터넷과 달리 시간과 문서량에 큰 제약받기 때문에 사용자 질문의 정확도와 검색 시간을 중점으로 실험을 진행하였다. 실험결과에서처럼 사용자의 질문 파악에 클래스의 속성을 이용한 추론과 비교하여 높은 정확도를 보인다. 또한 본 논문에서 제안한 유사도 알고리즘을 클래스간의 속성 노드비교에 이용하였을 때 기존의 게시판 검색을 위한 키워드 매칭(Keyword Matching)과 비교하여 높은 정확도를 가지게 된다. 하지만 여전히 사용자의 질문에 대한 가중치와 제약 속성(Constraint Property)을 판단함에 있어 효과적인 자연어 처리를 통한 RDF 문서 구성이 요구되며, OWL과 같이 다양한 문법이 제공되는 언어로의 온톨로지 확장도 필요하다. 따라서 자연어 처리와 효과적인 온톨로지 구성을 통한 보다 정확한 검색엔진의 개발을 향후 연구과제로 삼고 있다.

감사의 글

본 연구는 한국과학재단 지정 인천대학교 동북아 전자물류 연구센터의 지원에 의한 것임.

6. 참고 문헌

- [1] 박제복, “분산 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] Sumi Helal, “WAP : Present and Future”, IEEE, 2002
- [9] Chi-Wei Lan, Chun-Chou Chien, Meng-Yen Hsieh, “A Mobile e-Commerce Solution”, IEEE, 2002
- [10] Kit Hui, Stuart Chalmers, Peter Gray, Alun Preece, “Experience in using RDF in Agent-mediated Knowledge Architectures”, AAAI, 2003