

# 지능형 OMDR 기반의 자동 문서 공유 에이전트를 이용한 지식서비스

김수경\*, 최호진\*

\*한국정보통신대학교 공학부

e-mail:ksk0314@icu.ac.kr, hjchoi@icu.ac.kr

## A Knowledge Service Using Automatic Document Sharing based on Intelligent OMDR

Su-Kyoung Kim\*, Kee-Hong Ahn\*\*

\*Dept of Engineering, Korea Information Communication University

### 요 약

본 연구는 온톨로지, 자연어 처리, 메타데이터 등의 시맨틱 웹 기반 기술들을 이용하여 시맨틱 웹 응용을 위한 전체적인 기술 적용과 그의 활용에 목적을 두고 있다. 이를 위해 OWL을 기반으로 조직이나 기관의 지식 주제별 도메인 온톨로지와, 기존 워드넷(WordNet)이나 더브린 코어 메타데이터(Dublin Core Meta Data)와 조직에 정의된 데이터베이스의 스키마를 MDR로 구축하여 상호 연결하여 온톨로지가 갖는 지능적 추론과 규칙 서비스와 표준화된 메타데이터의 결합 방법을 제공한다. 이는 기존에 온톨로지와 메타데이터의 재활용과 연결(Alignment)에 있어 연구적으로 높은 가치가 있다. 그리고 조직의 사용자가 문서를 작성할 때 문서의 내용에 대해 자연어 처리 기술과 온톨로지의 기술을 이용해 적합한 용어나 메타데이터를 자동으로 제공하여 작성된 문서의 공유와 재사용성을 높이고, 작성된 문서를 XML 형식으로 구성되는 XML 기반 지능 문서 데이터베이스(XMB Based Intelligent Document Database)에 저장하여 유사한 문서를 작성하거나 사용할 필요가 있는 사용자에게 문서 등록과 검색 에이전트(Document Registry and Retrieval Agent)를 통해 이러한 제공하여 문서 지식의 사유화를 최소화 하고, 유사 문서의 제작성과 또는 특정 문서의 작성에 필요한 시간이나 경비를 줄이게 된다. 또한 웹상이나 PDA 같은 개인 휴대장치를 통해서도 서 등록과 검색 에이전트를 통해 문서를 검색하고 사용할 수 있게 한다면 언제 어디서나 해당 서비스를 활용하는 유비쿼터스와 시맨틱 웹의 실질적 응용을 거둘 수도 있으리라 사료된다.

### 1. 서론

시맨틱 웹이 제시 된지 많은 시간과 다양한 관련 연구가 진행되었으나 최종 사용자(End-User)를 위한 시맨틱 웹 기반의 응용시스템에 대한 개발과 연구 부족으로 인해 현재까지도 시맨틱 웹의 활용과 발전은 부족한 실정이다. 따라서 본 연구는 현재 시맨틱 웹을 기반으로 하는 응용에 대한 현재 국내의 연구 상황 분석과 또한 시맨틱 웹 응용이 갖춰야할 핵심 기반 기술들에 대한 연구 분석을 통해 시맨틱 웹이 현재 웹의 문제점을 극복할 수 있는 킬러 애플리케이션의 역할을 할 수 있는 시맨틱 웹 응용시스템의 프레임워크를 제시하고자 한다.

시맨틱 웹의 공유성(Sharing), 지능성(Intelligence), 확장성(expansion)과 자동성(Automaticity)과 웹 서비스의 특징을 만족할 수 있도록 "지능형 OMDR 기반의 자동 문서 등록 및 검색 에이전트를 통한 지식서비스"를 제공하기 위해 추론과 규칙 서비스를 제공하는 OWL(Web Ontology Language) 기반의 MDR(Meta Data Repository)를 이용하여 사용자가 조직이나 기관에서 작성한 문서(document)의 내용 중 주요 메타데이터와 용어 그리고 용어 의미를 자동으로 추출·생성하여 개인 데스크

탑 또는 지식 문서 저장소에 등록된 뒤 해당 문서를 필요로 하는 조직 또는 기관의 사용자에게 재사용하거나 지식으로 제공하고 문서를 작성하는 사용자에게는 자동으로 조직이나 기관의 메타데이터나 용어를 참조할 수 있도록 반자동으로 제공하는 시맨틱 웹 기반의 지식서비스시스템을 제안하였다.

먼저, 본 연구는 웹 온톨로지가 갖는 특징을 만족할 수 있도록 서술 논리(Description Logic)와 규칙을 제공하는 OWL을 기반으로 조직이나 기관의 지식 주제별 도메인 온톨로지와, 기존 워드넷(WordNet)이나 더브린 코어 메타데이터(Dublin Core Meta Data)와 조직에 정의된 데이터베이스의 스키마를 MDR로 구축하여 상호 연결하여 온톨로지가 갖는 지능적 추론과 규칙 서비스와 표준화된 메타데이터의 결합 방법을 제공한다. 이는 기존에 온톨로지와 메타데이터의 재활용과 연결(Alignment)에 있어 연구적으로 높은 가치가 있다.

그리고 조직의 사용자가 문서를 작성할 때 문서의 내용에 대해 자연어 처리 기술과 온톨로지의 기술을 이용해 적합한 용어나 메타데이터를 자동으로 제공하여 작성된 문서의 공유와 재사용성을 높이고, 작성된 문서를 XML

형식으로 구성되는 XML 기반 지능 문서 데이터베이스(XMB Based Intelligent Document Database)에 저장하여 유사한 문서를 작성하거나 사용할 필요가 있는 사용자에게 문서 등록과 검색 에이전트(Document Registry and Retrieval Agent)를 통해 이러한 제공하여 문서 지식의 공유화를 최소화 하고, 유사 문서의 재작성과 또는 특정 문서의 작성에 필요한 시간이나 경비를 줄이게 된다. 또한 웹상이나 PDA 같은 개인 휴대장치를 통해서도 서 등록과 검색 에이전트를 통해 문서를 검색하고 사용할 수 있게 한다면 언제 어디서나 해당 서비스를 활용하는 유비쿼터스와 시맨틱 웹의 실질적 응용을 거둘 수도 있으리라 사료된다.

2. 관련연구

시맨틱 웹을 기반으로 하는 응용 시스템에 대한 연구는 1999년 팀 버니스리(Tim Buners Lee)가 시맨틱 웹의 패러다임을 제시한 후 다양한 분야에서 시도되었다. 국내의 경우, 시맨틱 웹에 대한 연구 초기에는 시맨틱 웹 기반의 검색 시스템의 구조에 대한 연구[1]와 이후 시맨틱 웹을 기반으로 하는 다양한 정보 검색 시스템[2, 3, 4, 5, 6]의 연구가 있다.

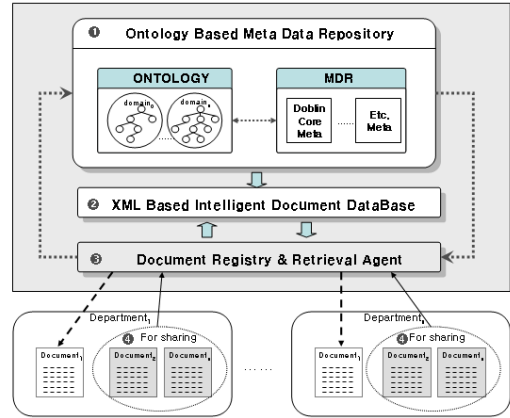
[1]의 연구는 초기 시맨틱웹을 기반으로 검색 시스템을 구현하는 데 있어 기본적인 메커니즘을 제공하였으나 실제 응용을 제공하는데 필요한 기초 기술들의 방법과 응용을 제공하지 못하였다. [2] 연구는 온톨로지와 에이전트를 이용한 시맨틱 웹 기반의 정보 검색 시스템을 제안하였으나 시맨틱 웹 기반 응용의 가장 핵심 기술인 온톨로지의 명시적인 활용과 추론에 대한 방법을 제시하지 못하고 실험에 구현된 응용 시스템의 활용 방안에 있어 최종 사용자를 위한 정보 전달의 제시 보다는 단순한 유사도 측정의 결과로서 시맨틱 웹 기반 응용 시스템을 제시하였다. [3]의 연구는 DQL(DAML Query Language)를 이용한 시맨틱 웹 검색 시스템을 제시하였다. 그러나 [3]은 온톨로지 추론과 추론 엔진을 적용하였으나 실제 사용자를 위한 검색이 아닌 온톨로지 언어 중 하나인 DALM의 표현을 통해 검색을 제시함으로써 사용성이 현저히 떨어진다.

[4],[5] 연구는 특정 분야의 검색 시스템을 시맨틱 웹을 기반으로 하는 방법을 제시하였으나 두 연구 또한 온톨로지의 지식 표현과 구축을 통한 정보 검색 방법이 아닌, 온톨로지 내 개념을 실제 온톨로지가 아닌 관계형 데이터베이스로 구축하여 이를 통해 정보 검색이 진행되도록 실제 의미론적인 검색 보다는 개념이 구축된 테이블 들간의 정합을 통해 정보검색이 진행되는 문제가 있다. 특히 국내의 경우 시맨틱 웹 기반의 응용 시스템에 대한 연구가 초기에 비해 많이 감소되고 일부에서는 시맨틱 웹의 발전 가능성에 대한 회의론을 제기하기도 하였다.

국외 연구는 국내 연구에 비해 시맨틱웹을 기반으로한 응용시스템의 메커니즘과 구축 방향 및 다양한 분야로의 접목의 시도되고 있는 실정이며, 최근에는 최종 사용자

위한 시맨틱웹 응용의 활성화가 시맨틱웹을 발전시킬 수 있는 근본적인 요인으로 발표되고 있다[6]. W3C에서는 SWAP(Semantic Web Application Platform)을 의 필요성을 구축하여 이를 통해 정보 검색이 이뤄지도록 하였고 구축과 지식 표현과 공유성의 부분시맨틱 웹의 기반 기술 연구에 대한 선도적 위치를 선점연구와 발전에 있어 의 연구와 발전에 있어 방법과 있도록 예전에 인공지능 분야에서 많은 연구가 진행되었던 온톨로지(Ontology)에 대한 다양한 연구가 활발히 진행되고 도리어 많은 연구자들이 온톨로지에 대한 연구에만 집중되어 초창기 시맨틱 웹에서 제시했던 웹의 발전 방향을 퇴색되고 있는 상황이다.

3. OMDR 기반 시스템 제안



(그림1) OMDR 기반 시스템의 전체 구조

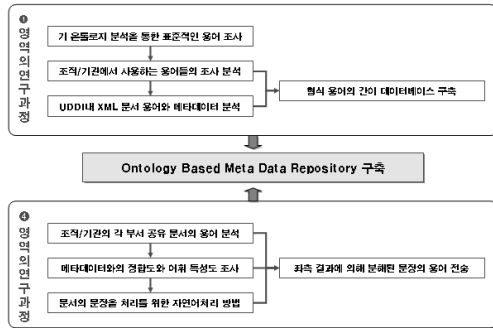
(그림 1)은 온톨로지 기반 메타 데이터 저장소에 등록된 온톨로지와 메타데이터를 기반으로 진행되는 제안 시스템의 전체 구조이다. 제안시스템은 괄호숫자 ①, ④의 부분을 중심으로 진행된다.

먼저, ①의 Ontology Based Meta Data Repository의 도메인 온톨로지와 메타데이터 저장소의 구축을 위해 워드넷이나 Cyc와 같은 온톨로지의 분석을 통해 인간의 지식 개념에 대한 표준적인 용어들을 조사하고 이를 특정 조직이나 기관에서 사용하는 용어들과의 비교를 통해 형식 용어 (Formal Term)를 온톨로지 간에 데이터베이스로 일단 구축한다. 또한 현재 UDDI(Universal Description, Discovery, and Integration)내 등록된 일반 조직이나 기관에서 웹 서비스의 표준으로 사용되는 XML 문서들에 표현된 엘리먼트(Element)나 속성(attribute)에 대한 용어를 더블린 코어 메타데이터와 같은 표준 메타데이터 들과의 정합성 분석을 통해 이를 위의 온톨로지 간에 데이터베이스에 추가한다.

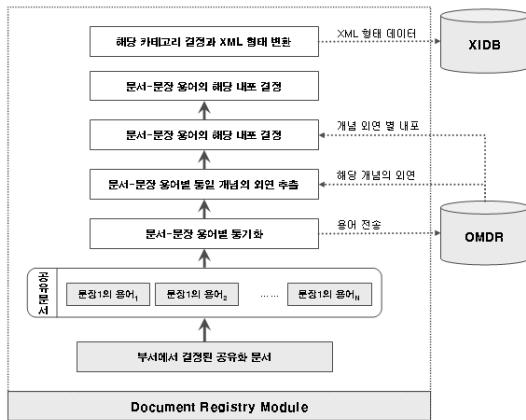
다음 ④ 영역에 나타난 조직이나 기관의 각 부서에서

사용되는 공유용 문서들에 표현되는 용어나 어휘에 대한 분석을 통해 메타데이터와의 표준화 정합도 그리고 동음이의어나 이음동어의 표현 정도를 조사하고 작성 문서에서 나타난 문장들을 처리하기 위한 자연어처리 방법을 개발한다. 이 자연어처리 방법에 의해 문서내 표현되는 용어 중 사적인 내용에 해당하는 용어들을 제거하고 분해된 문장의 용어들을 ③의 DR2A(Document Registry & Retrieval Agent)로 전송하는 방법을 연구한다.

마지막으로 ① 영역에서 구축된 온톨로지 간에 데이터 베이스와 ④ 영역의 연구 결과를 종합하여 Ontology Based Meta Data Repository를 온톨로지 서버에 구축한다. 일단 처음에 구축되는 온톨로지나 메타데이터 들은 수동으로 구축되며 추후 ② 영역과 ③ 영역의 연구에 따라 자동화 OMDR로 제정의 된다. 위에 기술된 내용들을 다시 간단하게 도식하면 (그림 2)와 같으며, (그림 2)는 ① 영역과 ④ 영역에 대한 자세한 절차를 표현하였다.



(그림 2) OMDR과 공유데이터의 접목 절차



(그림 3) OMDR과 공유데이터의 접목 절차

(그림 1)에 있는 괄호숫자 ②과 ① 부분은 다음과 같은 절차에 의해 다시 진행된다. 구축된 ①의 OMDR에 구

성된 지식들을 ③의 DR2A에서 전달되는 용어들의 발생 패턴과 확률을 의해 기계 학습하여 전달 용어들의 발생 패턴이 특정 임계치를 넘어서고 전달 용어가 ①의 지식 구성에 있어 발생 패턴의 기준에 미치지 못하거나 없는 경우, 또는 발생 패턴 기준에 비해 지식의 중요도가 높게 표현되어 있는 경우에 ①의 OMDR 내 지식들을 제정하고 정렬하여 자동으로 OMDR을 동적으로 관리하게 된다.

이를 위해서는 먼저 ③의 DR2A의 설계와 구현이 필요하다. (그림 3)은 DR2A의 상세한 구성도이다. DR2A의 구성은 크게 Registry Module과 Retrieval Module이다. Registry Module은 각 부서에서 작성된 문서들 중 공유화를 수락한 문서들에 대해 제안된 자연어 처리 방법에 따라 나타난 문장별 용어들을 OMDR로 동기적으로 전송하여 전송된 용어와 외연(Extension)이 같은 OMDR의 개념들을 추출한다. 다음 추출된 개념들의 내포들을 확인하고 확인된 결과에 의해 문서를 구성하는 용어들의 전체 분포 임계치를 계산하고 미리 기준된 임계치에 의해 해당되는 카테고리를 결정하여 XML 형태로 XIDB(XML-Based Intelligent Document DB)로 등록이 진행된다.

### 5. 결론

본 연구는 온톨로지, 자연어 처리, 메타데이터 등의 시맨틱 웹 기반 기술들을 이용하여 시맨틱 웹 응용을 위한 전체적인 기술 적용과 그의 활용에 목적을 두고 있다. 이를 위해 OWL을 기반으로 조직이나 기관의 지식 주제별 도메인 온톨로지와, 기존 워드넷(WordNet)이나 더브린 코어 메타데이터(Dublin Core Meta Data)와 조직에 정의된 데이터베이스의 스키마를 MDR로 구축하여 상호 연결하여 온톨로지가 갖는 지능적 추론과 규칙 서비스와 표준화된 메타데이터의 결합 방법을 제공한다. 이는 기존에 온톨로지와 메타데이터의 재활용과 연결(Alignment)에 있어 연구적으로 높은 가치가 있다.

그리고 조직의 사용자가 문서를 작성할 때 문서의 내용에 대해 자연어 처리 기술과 온톨로지의 기술을 이용해 적합한 용어나 메타데이터를 자동으로 제공하여 작성된 문서의 공유와 재사용성을 높이고, 작성된 문서를 XML 형식으로 구성되는 XML 기반 지능 문서 데이터베이스(XMB Based Intelligent Document Database)에 저장하여 유사한 문서를 작성하거나 사용할 필요가 있는 사용자에게 문서 등록과 검색 에이전트(Document Registry and Retrieval Agent)를 통해 이러한 제공하여 문서 지식의 공유화를 최소화 하고, 유사 문서의 제작성과 또는 특정 문서의 작성에 필요한 시간이나 경비를 줄이게 된다. 또한 웹상이나 PDA 같은 개인 휴대장치를 통해서도 서 등록과 검색 에이전트를 통해 문서를 검색하고 사용할 수 있게 한다면 언제 어디서나 해당 서비스를 활용하는 유비쿼터스와 시맨틱 웹의 실질적 응용을 거둘 수도 있으리라 사료된다.

**참고문헌**

- [1] 이강찬, 김성환, 민재홍, 박기식, 정인정, “시맨틱웹 기반의 검색 시스템 구조”, 한국전자통신연구원 주간기술동향, 2003.5
- [2] 최옥경, 한상용, “자동화된 통합 프레임워크를 위한 시맨틱 웹 기반의 정보 검색 시스템”, 2005.9, 정보처리학회논문지 제13-C권 제1호
- [3] 김제민, 박영택, “시맨틱 웹 기반 DQL 검색 시스템 설계”, 2005.2, 정보처리학회논문지 제12-b권, 제1호
- [4]이회경, 전우친, “시맨틱 웹을 활용한 초등학교 학습자료 검색 시스템”, 2006.10, 정보처리학회논문지 A 제13-A권, 제6호
- [5] 조부현, 민영근, 이복주, “시맨틱 웹을 이용한 웹 변경 탐지 시스템”, 2006.1, 정보처리학회논문지 제13-B권 제1호
- [6] 황정희, 구미숙, 이현아, 류근호, “시맨틱 웹 기반의 고객 정보 검색 시스템의 설계 및 구현”, 정보처리학회논문지 제13-D권 제4호
- [6] Karun Bakshi, David R. Karger, "End-User Application Development for the Semantic Web", 2007. MIT
- [7] MIT-LCS, 'Semantic Web Application Platform-SWAP", 2000.10, W3C
- [8] D. Fenset, C.Bussler, Y.Ding, V.Kartseva, M.Klein, M.Korotkiy, B.Omelayneko, R. Siebes, "Semantic Web Application Areas",
- [9] <http://www.twine.com/>
- [10] 김수경, 안기홍, “시맨틱웹 응용을 위한 웹온톨로지 구축기법”, 2008.2, 정보처리학회 논문지 제15-D권, 제1호