

집단지성을 이용한 온톨로지 생성 시스템의 설계

손종수, 류재환, 정인정

고려대학교 컴퓨터정보학과

충청남도 연기군 조치원읍, 339-700

Tel: +82-41-860-1342, E-mail: {mis026, berserkjan, chung}@korea.ac.kr

Abstract

온톨로지는 시맨틱 웹의 핵심 기술로서 2001년 시맨틱 웹이 제창된 이후 다양한 온톨로지 생성 방법의 연구가 이루어져 왔다. 그러나 온톨로지는 작성이 어렵고 난해한 면이 있어 소수의 전문가 집단에서만 사용되고 있는 것이 현실이다.

한편 웹2.0의 등장을 통해 콘텐츠의 생산 주체가 서비스 제공자에서 수요자로 바뀌게 되었으며 이로 인해 집단지성의 중요성이 부각되고 있다.

본 논문에서는 웹2.0 기반 환경에서 사용자가 생성한 온톨로지를 수집하여 대중 온톨로지를 완성하는 시스템 구조를 제시한다. 사용자 온톨로지 생성 도구는 오픈API를 이용한 웹2.0 기반 어플리케이션으로써 매시업(Mash-up) 형태로 제공된다. 그리고 온톨로지 생성 시스템은 사용자가 생성한 온톨로지를 수집하여 보편적인 형태의 온톨로지로 가공하는데 이 때 집단지성의 원리를 이용한다. 그리고 각 사용자들에게 RSS, ATOM 과 같은 출판 매커니즘을 이용하여 대중 온톨로지를 제공한다.

본 논문에서 제시하는 방법은 복잡하고 난해한 기존 온톨로지의 제작 방법의 단점을 효과적으로 해결하며 대중적으로 시맨틱 웹이 활용될 수 있는 환경을 구축할 수 있다.

Keywords:

집단지성; 웹2.0; 시맨틱 웹; 온톨로지; Web2.0, Ontology; Semantic web; Collective intelligence

1. 서론

1990년대 말 이후 인터넷의 발전과 2000년대 중반 이후 웹2.0의 대두를 통해 인터넷에는 다양한 종류의 미디어를 통해 셀 수 없이 많은 데이터가 존재하게 되었다. 그리고 웹을 통해 사용자가 활용할 수 있는 서비스도 기하급수적으로 늘어났다. 그러나 상대적으로 웹 데이터와 서비스의 분류 및

검색기법은 전통적인 키워드 매칭 방법을 벗어나지 못하고 있어 사용자가 원하는 자료를 찾기는 점점 더 어려워지고 있고 자료를 찾는 시간의 투자 또한 점점 더 많이 요구되고 있다. 팀 버너스 리에 의해 제시된 시맨틱 웹은 상기와 같은 전통적인 웹의 문제들을 극복하기 위하여 제시되었다. 시맨틱 웹은 정보자원의 의미를 컴퓨터가 스스로 인지하여 자동적으로 데이터를 처리하는 것에 그 목적이 있는데 이를 가능하게 하는 핵심기술은 온톨로지이다. 온톨로지는 객체와 객체 사이의 관계와 속성을 명시적으로 기술함으로써 시맨틱 웹의 구축이 가능하도록 하는 지식표현 언어를 지칭한다. 현재까지 시맨틱 웹을 구축하기 위하여 국내외에서 다양한 방법론이 제시되었으며 심도 있는 연구가 이루어졌지만 온톨로지의 제작 방법이 복잡한 점과 다양한 도메인의 구축된 온톨로지가 부족한 점으로 인하여 시맨틱 웹의 구축은 큰 진전을 보이지 못하고 있는 것이 현실이다. 또한 일반사용자들에게 시맨틱 웹의 활용성을 호소할 수 있을 만큼 매력적인 킬러 어플리케이션(Killer Application)을 만들어내지 못하고 있다.

한편, 웹2.0은 오라일리 미디어(O'Reilly Media)의 데일 도허티가 2000년대 초 닷컴(Dot com) 기업들의 붕괴 이후 살아남은 기업들의 특징을 정리하여 제안한 하나의 대명사를 의미한다. 웹2.0에서 제시된 특징들은 학술적으로 새로 개발된 기술들은 아니었으나 웹 기술이 발전해 나가야 할 방향을 제시했다는 것에 의미를 그 의미를 가진다.[1] 웹2.0이 제시되면서 웹 기반 서비스의 환경이 빠르게 변하고 있는데 그 중의 한가지로서 콘텐츠 생산 주체의 변화를 들 수 있다. 이전의 웹 환경에서 인터넷 콘텐츠의 생산 주체가 서비스 운영자였다면 웹 2.0에서 콘텐츠의 생산 주체는 사용자이다. 이에 따라 사용자 제작 콘텐츠를 분류하고 순위를 매기는 방법들이 다양하게 고안되었는데 이 방법들에 집단지성이 사용되어지고 있다. 한편, 웹2.0의 주요한 기술들은 XML 기반 데이터 포맷을 사용하여 시스템 및 언어의

이질성 문제를 효과적으로 해결하고 있다.[2] 따라서 사용자가 제작한 콘텐츠의 배포 및 공유가 쉬운 장점을 지니고 있으며 XML 언어를 사용함으로써 기존의 HTML 기반의 웹 데이터보다 구조적이고 가공하기 효과적이라는 장점을 가지고 있다.

본 논문에서 제시하는 시스템 아키텍처는 시맨틱 웹 구축을 위해 필수불가결한 온톨로지를 웹 2.0의 주요 기술 중 한가지인 집단지성을 이용하여 사용자가 직접 생성하고 공유하는 방법을 제시하고 있다. 집단지성을 이용하여 온톨로지를 생성하는 것은 비록 자동적인 생성방법은 아니지만 다수의 사용자가 큰 노력을 들이지 않고 온톨로지 생성이 가능하며 온톨로지를 공유하고 공동의 온톨로지를 추출함으로써 대중적으로 사용 가능한 온톨로지 집합을 구축할 수 있다. 본 논문에서 제안하는 집단지성을 이용한 온톨로지 생성 아키텍처는 웹2.0 기반 사용자 어플리케이션에 플러그인(Plug-in)과 비슷한 방법인 매쉬업(Mash-up) 방식으로 온톨로지를 생성 어플리케이션을 제작하며 RSS, ATOM과 같은 주요 웹 2.0 출판 매커니즘을 이용하여 공유함으로써 온톨로지의 생성 및 공유 문제를 해결한다.

2. 배경

2.1. 시맨틱 웹 및 온톨로지

시맨틱 웹은 기존의 월드와이드웹을 확장한 형태로서 컴퓨터가 이해할 수 있는 형태로 정보자원의 의미 및 관계를 기술하여 컴퓨터와 사람이 협동적으로 업무를 수행할 수 있는 형태의 웹을 지칭한다.[3] 시맨틱 웹을 구현하기 위하여 정보자원에 대한 의미기술이 필수적인데 이 때 사용되는 것이 온톨로지이다. 온톨로지는 정보의 의미를 기계가 이해하고 처리할 수 있는 형태로 표현하는 방법 및 마크업 언어를 통칭하며 대표적으로는 RDF, DAML+OIL, OWL 등이 있다. 무한대에 가까운 모든 사물에 대하여 온톨로지를 기술하는 것은 어렵고 복잡한 문제이므로 현재 온톨로지는 특정한 분야에 대해 한정되어 사용되고 있으며 그 구축 방법의 까다로움으로 인해 시맨틱 웹이 확산되는데 어려움으로 작용하고 있다. 이 문제를 해결하기 위한 다양한 온톨로지 생성방법이 제시되었는데 이 방법들은 본 논문의 3장에서 소개하고 있다.

2.2. 웹2.0 과 집단지성

웹 2.0은 오라일리 미디어에서 처음 사용된 용어로서 닷컴 회사들의 붕괴 이후 살아남은 기업들의 특징을 분류하는데 사용된 비즈니스적 관점에서 출발한 개념이다. 웹 2.0 기술은 참여, 공유 그리고 개방이라는 세 가지의 키워드로

표현되는데 이 세 가지의 특징을 이용하여 다수의 사용자는 다양하고 방대한 지식을 축적 할 수 있게 되었다. 이는 웹 2.0의 가장 큰 특징으로 분류되는데 이처럼 다수의 사용자가 신뢰성 있는 지식을 축적하는 것을 ‘집단 지성’ [8]이라고 표현한다. 집단지성은 'Wisdom of crowds' 혹은 'Swarm intelligence'라고도 불리며 한 개인의 지식보다는 집단 속에서 논의된 지식이 더 우수하다는 개념적인 용어이다.[9]

집단지성은 개인 사용자의 독창성과 다양성 그리고 지식의 집중으로 표현된다.[15] 개인 사용자의 독창성은 각 개인 사용자가 지식을 생산함에 있어 외부의 간섭을 최대한 배제하여야 한다는 것이고 다양성은 다양한 계층에서 지식 생산이 이뤄져야 한다는 것이다. 마지막으로 지식의 집중은 독창성과 다양성이 보장된 네트워크에서 생산된 지식을 모음으로서 전문가 지식에 못지않은 지식을 생산할 수 있다는 것이다.

그리고 또 한가지 웹2.0의 중요한 기술로서 매시업(Mash-up)이 있는데 이는 일반 어플리케이션에서 플러그인(Plug-in)과 비슷한 개념으로서 웹 어플리케이션에서 서비스 제공자가 제공하는 오픈 API를 이용하여 사용자가 직접 새로운 기능을 추가할 수 있도록 지원하는 개발 방법이다. 매시업은 많은 사용자를 보유하고 있는 블로그나 위키 등 웹 2.0의 대표적인 어플리케이션에서 사용되고 있으며 데이터의 입,출력 포맷이 XML 기반이므로 기존 웹에 비해 상대적으로 가공하고 처리하기 쉬운 형태를 지니고 있다.

3. 기존 온톨로지 생성방법 및 도구

3.1. OTKM

OTKM[4]은 OnToKnowledge Methodology for development and evolution of Ontology-based Tools for Knowledge Management의 약자로서 명명된 온톨로지 개발 방법론으로서 어플리케이션 지향 온톨로지 개발에 중점을 두고 있다. OTKM은 온톨로지 제작을 소프트웨어 제작 주기에 맞춰 제작하는 방법론으로서 폭발적으로 증가하는 웹 자원에 신속하게 대응하기 어려우며 온톨로지 생성에 방대한 비용 및 시간이 소요되는 단점을 가지고 있다.

3.2. ONIONS

ONIONS(Ontologic Integration On Naive Sources)[5]는 도메인 온톨로지를 분석 및 통합하는 목적으로 개발된 방법이다. ONIONS의 온톨로지 생성 방법은 과다한 온톨로지 정제 방지 및 지식의 상관성과 같은 온톨로지의 입수문제에 더 큰 초점을 둔다. ONIONS는 도메인 온톨로지 제작에는 유효한

방법이나 이 역시 방대하게 늘어나는 웹 자원에 대응하기에는 미흡하다.

3.3. Protege

Protege[6][7]는 스탠포드대학교 의료정보학과에서 지식기반 구조를 작성하기 위한 시스템으로서 지난 15년간의 연구기간을 거친 온톨로지 제작 도구이다. Protege는 다양한 분야의 온톨로지를 생성하고 생성한 온톨로지를 시각화하며 관리할 수 있는 환경을 제공한다.[13] 그러나 Protege는 윈도우 응용프로그램으로 제작되어 온톨로지를 하나하나 직접 제작해야한다는 단점이 있으며 대량의 온톨로지를 만들기가 어렵다.

3.4. 그 외의 온톨로지 제작 방법 및 도구

그 외에 GUI기반의 온톨로지 제작도구인 OntoEdit [10]과 KATUS[11], Ontology Development 101 [12] 등의 제작 방법론이 존재하나 이들 역시 효과적으로 대량의 온톨로지를 제작하고 공유하는 방법을 제공하고 있지는 못하다.

4. 집단지성을 이용한 온톨로지 생성 방법

본 논문에서 제시하는 집단지성을 이용한 온톨로지의 생성은 웹 2.0에서 사용되고 있는 집단지성의 개념을 온톨로지 생성분야에 접목시켜 다수의 사용자가 사용자의 지식을 추상화시켜 원시 온톨로지를 제작하고 제작된 원시 온톨로지를 모아 대중 온톨로지로 완성하는 과정을 거치게 된다. 본 장에서는 집단지성 원리에 입각하여 사용자가 제작한 원시 온톨로지를 모아 대중 온톨로지를 추출하는 모형 및 이를 구현하기 위한 아키텍처를 제시한다. 그리고 원시 온톨로지의 생성부터 대중 온톨로지의 생성까지 필요한 출판 및 공유 과정을 제시한다.

4.1. 집단지성의 원리

집단지성은 다수의 사용자가 개개인의 작업 및 지식을 공유하고 취합하여 일반적 사실을 도출해 낸다는 원리를 가지고 있다. 이 원리는 다수 사용자의 참여에 의해 어떤 사실에 대한 신뢰성을 높여나간다는 것이 그 핵심이다. 이 때 잘못된 사실이 전체 집합에 포함되기도 하는데 잘못된 사실은 전체집합에서 아주 작은 부분이므로 참과 거짓을 쉽게 구분이 가능하다. 집단지성의 원리를 수학적으로 해석하면 다음과 같다.[16]

사용자들마다 알고 있는 지식을 $X_k = a + n_k$ 로 표현할 때 a 는 옳은 지식이고 n_k 는 잘못된 지식을 의미한다. 여기서 k 는 사용자의 번호를 의미한다. 지식 X 에 대해 1부터 k 까지의 사용자가 알고 있는 지식을 모두 합하면 식 (1)과 같이 표현된다.

$$x_{CI} = \sum_{k=1}^K x_k = Ka + \sum_{k=1}^K n_k \quad (1)$$

식 (1)에서 옳은 지식은 그 양이 증가할수록 선형적으로 증가하지만 잘못된 지식은 모두 같은 지식이 아니고 상호 간섭이 적으므로 서로 독립적이라고 볼 수 있다. 따라서 잘못된 정보는 \sqrt{K} 으로 표현이 되며 이 경우 k 가 증가할수록 \sqrt{K} 배 감소하게 된다. 단 위의 표현 방법은 지식을 하나로 상정한 것이므로 지식의 수가 무한대에 가깝다고 할 때 어느 사용자 K 가 알고 있는 i 번째 지식은 $X_k[i_k] = a[i_k] + n_k$ 로 표현한다. 여기서 i_k 는 K 사용자가 i 번째 편집한 지식을 의미한다. 따라서 집단 지성에 의해 1부터 k 까지의 사용자가 편집한 i 번째 지식은 다음 식 (2)와 같이 표현된다.

$$x_{CI}[i] = \sum_{k=1}^K \begin{cases} x_k[i_k] & i_k = i \\ 0 & i_k \neq i \end{cases} = K_i a[i] + \sqrt{K} n \quad (2)$$

위 식에서 i 번째 지식에 대해 편집한 사용자가 많아지면 많아질수록 $K_i a[i]$ 가 선형적으로 증가하게 되므로 옳은 지식의 양은 커지게 되며 반면 잘못된 지식은 그 비율이 갈수록 적어지게 된다.

4.2. 시스템 설계의 요점

집단 지성을 이용하여 온톨로지를 생성하는 것은 소수의 전문가가 관련 분야의 상세한 지식을 온톨로지로 기술하는 전통적인 방법과는 달리 일반 사용자들이 어렵지 않게 개념을 도식화하고 활용할 수 있어야 한다. 따라서 본 논문에서 제시하는 온톨로지 생성 방법을 구현하기 위하여 두 가지 중요한 시스템 설계의 요점을 제시하도록 한다.

접근의 용이성 : 웹 2.0이 대두되기 시작하면서 사용자들은 개인 홈페이지를 스스로 만드는 것에서 벗어나 서비스 제공자가 제공하는 블로그, 위키 등의 개인화 서비스를 받는 것으로 전환하였다. 서비스 제공자에게 개인화 서비스를 제공받는 경우 사용자들이 스스로 자신의 공간을 커스터마이징할 수 없다. 이 단점을 극복하기 위하여 웹 2.0 서비스 제공자들은 오픈 API를 제공하고 있는데 오픈 API를 통해 사용자들이 스스로 간단한 어플리케이션을 만들거나 혹은 다른 사람이 만든 어플리케이션을 자신의 설치하여 사용할 수 있다. 따라서 집단 지성을 이용한 온톨로지의 생성도 블로그나 위키 서비스 등에서 제공하는 오픈API를 이용하여 제작하여 사용자들 자신이 받고 있는 서비스에 어렵지 않게 온톨로지 생성 도구를 추가할 수 있는 형태가 되어야 한다.

사용의 편리성 : 온톨로지 언어는 클래스와 클래스 간의 관계를 설정하고 속성을 정의하는 등 일반

사용자들이 이해하기 어려운 개념을 많이 포함하고 있다. 온톨로지의 난해함으로 인해 시맨틱 웹이 최종 사용자층까지 퍼질 수 없었던 것을 감안할 때 사용의 편리성은 본 시스템 구성에 있어 중요한 요소로 작용된다. 사용자의 편리성을 제공하기 위하여 클래스의 관계 표현은 상·하위 관계, 동치 관계, 반의 관계만으로 한정하며 제한(Restriction)은 사용하지 않는 것을 제시한다. 그리고 클래스의 생성 및 관계 표현을 쉽게 하기 위하여 웹에서 그래프 표현 형식을 제공하여 온톨로지 생성의 편의성을 높여야 한다.

4.3. 용어의 정의 및 표현 범위의 제한

본 논문에서 제시하는 온톨로지의 생성 방법은 블로그, 위키 등의 웹 2.0 기반 어플리케이션에서 사용자가 간단히 온톨로지를 제작하고 그 온톨로지를 온톨로지 저장소로 전송하여 신뢰성 있는 대중 온톨로지로 재 가공 하는 방법이다. 본 논문에서는 사용자가 제작한 온톨로지를 ‘원시 온톨로지’ 라고 표현하도록 하며 집단지성을 이용하여 재 가공되어 완성된 온톨로지를 ‘대중 온톨로지’ 라고 한다. 그리고 대중 온톨로지를 만들기 위해 매개 역할을 하는 온톨로지가 필요하게 되는데 이를 ‘매개 온톨로지’ 라 한다. 본 온톨로지 생성방법은 전문가가 온톨로지를 생성하는 방법이 아니므로 복잡한 관계 표현이 들어간 온톨로지는 배제하며 일반 사용자가 쉽게 제작 할 수 있도록 위 사용자 편의성에서 제시한 네 개의 클래스 관계표현 어휘만을 사용한다.

4.4. 시스템 아키텍처

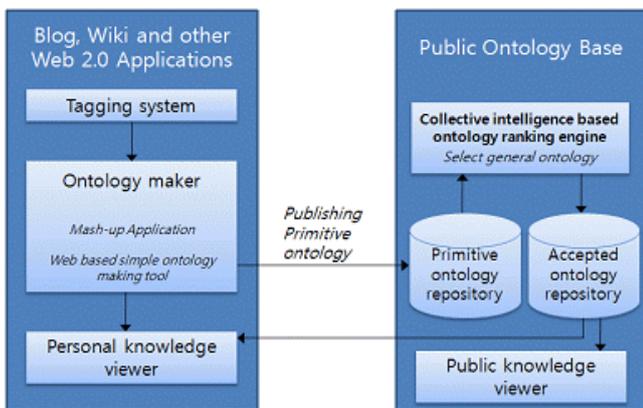


그림 1 - 전체 시스템

집단 지성을 이용한 대중 온톨로지를 생성하기 위하여 다음 두 가지의 시스템이 필요하다.

- 웹 기반 사용자 온톨로지 생성 도구
- 공공 온톨로지 생성 및 저장 서버

첫 번째는 사용자가 원시 온톨로지를 제작하기 위한 웹 2.0 기반 어플리케이션이며 두 번째는 사용자 어플리케이션에서 전송한 원시 온톨로지들을 모아 객체 관계표현이 모두 나열된 매개 온톨로지를 만든다.

그림 1은 본 시스템의 전체 시스템 아키텍처를 보인다. 그림 1에서 좌측부가 웹 기반 사용자 온톨로지 생성 도구이며 우측부가 공공 온톨로지 생성 및 저장 서버이다. 그림 1의 각 모듈들의 기능은 다음과 같다.

태깅시스템

좌측부에서 태깅시스템 (Tagging system)은 사용자들이 웹에서 문서를 쓸 때 태그를 달아 놓게 하는 시스템으로써 웹2.0 기반 어플리케이션에서 모두 지원하고 있는 것을 가정한다.

온톨로지생성자

좌측부 중앙에 위치한 온톨로지 생성자(Ontology maker)는 태깅시스템에서 저장된 태그들의 관계를 기술하는 원시 온톨로지 생성 도구이다. 본 원시 온톨로지 생성도구는 작성한 태그들에 대해 상·하위 관계 및 동치, 반의 관계를 설정해주고 작성된 원시 온톨로지는 그림 1의 우측부인 공공 온톨로지 저장소로 전송된다. 그리고 작성된 원시 온톨로지는 개인 웹사이트에서 개념간의 관계표현을 출력할 수 있도록 개인 지식 뷰어(Personal knowledge viewer)에 온톨로지를 제공한다.

지식 뷰어

지식 뷰어는 개인 지식 뷰어와 공공 지식 뷰어로 구분이 된다. 개인 지식 뷰어(Personal knowledge viewer)는 개인이 만든 온톨로지와 공공 온톨로지를 사용자 웹 페이지에 출력한다. 그리고 공공 지식 뷰어는 공공 온톨로지를 웹 페이지에 출력한다.

온톨로지 생성 엔진

온톨로지 생성 엔진(Ontology creation engine)은 집단 지성의 원리를 이용하여 사용자들이 전송한 온톨로지를 모아 대중 온톨로지를 작성한다.

온톨로지 저장소

온톨로지 저장소는 원시 온톨로지 저장소와 공공 온톨로지 저장소로 구분된다. 이 중 원시 온톨로지 저장소는 사용자들에게 전송받은 온톨로지를 저장하며 온톨로지 생성 엔진으로 원시 온톨로지를 전송하여 공공 온톨로지를 생성하게 한다. 그리고 공공 온톨로지 저장소는 온톨로지 생성 엔진에서 만들어진 공공 온톨로지를 저장한다. 사용자/대중 지식 뷰어에 온톨로지를

전송할 수 있도록 RSS[17] 등의 웹2.0 기반의 출판 매커니즘을 활용한다.

4.5. 집단지성을 이용한 온톨로지의 생성

집단지성에 대한 연구는 개미가 집단의 의사활동을 이용하여 최적의 경로를 탐색하고 가장 좋은 먹이감을 얻어내는 현상을 과학적으로 분석함으로써 시작되었다. 개미들은 페르몬을 이용하여 다른 개미들과 의사소통을 하며 페르몬의 밀도에 따라 긍정과 부정을 구분한다. 따라서 집단 지성을 컴퓨터 과학에서 활용하기 위하여 계산 모형을 수립하는데 이를 활용한다.[14] [14]에서는 웹 페이지에서 페르몬의 밀도를 계산하여 적용하고 밀도의 강도에 따라 탐색을 하는 방법을 취하고 있다. 본 논문에서는 [14]의 계산 모형을 따르되 페르몬 밀도를 계산하는 객체를 온톨로지 저장소에 모아진 원시 온톨로지에 적용하여 사용한다. 아래의 표 1은 [14]에서 제시한 모델과 본 논문에서 제시한 모델의 차이를 보인다.

표 1- 집단 지성의 적용 모델

대상	Lovelle et al.	본 논문
곤충 사회	웹 시스템	웹 시스템
개미	웹 클라이언트/서버	웹 2.0 유저
음식	웹에 산재한 정보	잘 정의된 온톨로지
음식 구하기	웹 서핑	온톨로지 취득
페르몬 분비	웹 서버의 로그 데이터	원시 온톨로지 생성
상호작용	클라이언트/서버간의 요청 및 응답	온톨로지 통합
페르몬 밀도	웹 페이지의 대중성 (Popularity) 및 중요성	온톨로지의 대중성

집단지성의 개념을 이용하여 온톨로지를 생성하기 위하여 본 논문에서는 웹2.0 기반 사용자 어플리케이션에서 온톨로지 저장소로 원시 온톨로지를 전송하여 대중 온톨로지를 제작하는 방법을 택하였다. 원시 온톨로지로부터 대중 온톨로지가 만들어지기 까지의 과정은 다음과 같다.

Step1) 원시 온톨로지의 생성

- 각 사용자들이 원시 온톨로지를 생성
- 온톨로지 저장소로 온톨로지를 전송

Step2) 원시 온톨로지의 취합

- 전송받은 원시 온톨로지를 취합
- 동일한 개념에 대한 온톨로지 군집화

Step3) 매개 온톨로지의 생성

- 동일 개념에 대한 빈도 계산 및 통합

Step4) 대중 온톨로지의 생성

- 매개 온톨로지의 RDF 트리플 중에서 높은

페르몬 값을 갖는 노드를 선택하여 대중 온톨로지를 생성

대중 온톨로지를 생성하기 위하여 요구되는 위 4개의 단계 중, Step1)에서는 사용자가 RDF/OWL 온톨로지를 생성한 후 전송하는데 이 때 사용되는 문법은 위 4.2.장에서 제시한 바와 같이 4개의 관계 표현만으로 제한한다. 사용자가 생성한 하나의 원시 온톨로지를 PO 라고 표현하며 PO 는 하나의 RDF 그래프를 의미한다. Step2)에서 원시 온톨로지를 온톨로지 저장소에서 취합할 때 모아진 원시 온톨로지 집합을 S 라고 하며 식 (3)과 같이 표현된다. PO 가 n 개 작성되었을 때 집합 S 는

$$S = \{PO_1, PO_2, \dots, PO_n\} \quad (3)$$

의 원소를 갖는다. 그리고 S 의 각 원소에 포함된 PO 는 *subject* s , *predicate* r , *object* o 를 갖는 RDF 트리플들의 집합이다. 본 논문에서는 식 (4)와 같이 s, r, o 를 갖는 하나의 RDF 트리플을 D 로 표현한다.

$$D = \{ (s, r, o) \mid (s, r, o) \in PO \} \quad (4)$$

수집한 원시 온톨로지 집합 S 에 포함된 특정한 RDF 트리플 D 에는 페르몬이 남겨지게 되는데 이 페르몬을 p 라고하며 식 (5)와 같이 표기한다.

$$(p : D \rightarrow V \times T) \quad (5)$$

식 (5)에서 V 는 웹 페르몬의 밀도를 말하며 범위는 $[0:1]$ 이다. 그리고 T 는 마지막으로 접근한 시간이며 범위는 $[1:2]$ 이다. 여기부터는 편의상 페르몬이 연계된 원시 온톨로지에서도 $P_v(d)$ 를 페르몬의 밀도값으로 표기하며 $P_t(d)$ 를 마지막 접근 시간으로 표기한다. $P_v(d)$ 는 식 (6)에서 보이는 바와 같이 DS_k 의 원소 개수의 합에서 D 의 총 개수를 나눈 값으로 계산하는데 DS_k 는 S 에 속한 어떤 *subject* s_k 를 갖는 D 들의 집합을 뜻한다.

$$p_v(d) := \frac{\text{number of } D}{\text{number of entities in } DS_k} \quad (6)$$

그리고 $P_t(d)$ 는 초기 값이 1이고 D 의 업데이트가 1회 일어나면 각 회당 0.1을 더하며 새로운 업데이트가 없는 경우 24시간에 한번씩 -0.1 만큼 감소하게된다. 따라서 $P_t(d)$ 와 $P_v(d)$ 에 의해 페르몬 p 는 최대 2, 최소 0의 값을 갖는다.

Step3)에서 매개 온톨로지를 생성하게 되는데 매개 온톨로지는 위에서 보인 $P_v(d)$ 와 $P_t(d)$ 를 이용하여 계산한 p 의 값을 모두 가지고있는 온톨로지를 뜻한다. 즉, 사용자들이 전송한 온톨로지를 모두 통합하여 객체 관계의 논리적인 모순 및 대중성을 고려하지 않은 온톨로지이다.

대중 온톨로지는 매개 온톨로지서 페르몬 p 를

이용하여 생성하는데 각 RDF 트리플에서 페르몬 p 가 0.6 이상인 것만을 선택하여 생성한다. 만약 같은 *subject*와 *object*에 대해 다른 *predicate*를 가지고있고 p 가 0.6 이상인 RDF 트리플이 두개 이상된다면 가장 큰 값을 가진 RDF 트리플을 선택한다. 선택된 RDF 트리플들의 집합은 하나의 온톨로지로 생성하는 것이 가능한데 이 온톨로지가 본 논문에서 제시하는 대중 온톨로지이다.

5. 예제 및 평가

본 장에서는 위 4장에서 제시한 온톨로지 제작 방법을 이용하여 온톨로지를 생성하는 예제 보이며 기존 온톨로지 제작 방법과의 비교를 통해 평가한다. 온톨로지를 생성하기 위하여 원시 온톨로지가 수집되어야하는데 본 논문에서는 그림 2와 같이 임의의 원시 온톨로지 3개를 제작하였다. 그림 2에서 좌측부터 PO_1, PO_2, PO_3 이라고 하고 PO_1, PO_2 의 최상위 노드 $\{A, i, C\}$ 를 D_1, D_2 로 표시하며 PO_3 의 최상위 노드 $\{A, j, C\}$ 를 D_3 라고 한다.

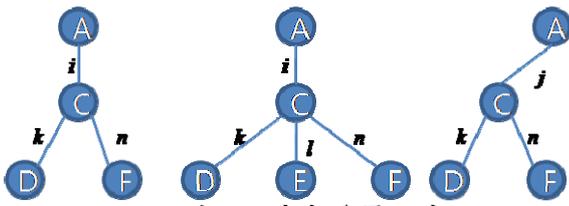


그림 2 - 원시 온톨로지

그리고 C 를 *subject*로 갖는 RDF 트리플의 집합을 DS_A 라고 할 때 $DS_A = \{D_1, D_2, D_3\}$ 이며 D_1, D_2 는 같으므로 노드 $\{A, i, C\}$ 의 $Pv(d) = 0.66$ 이 되며 D_3 의 $Pv(d)$ 는 0.33 이 된다. 그리고 그 외에 나머지 노드에 대해 모두 $Pv(d)$ 를 계산하면 다음 그림 3과 같이 매개 온톨로지가 생성된다. 그림 3에서 좌측의 그림은 $Pt(d)$ 를 1로 고정시켰을 경우의 결과이다.

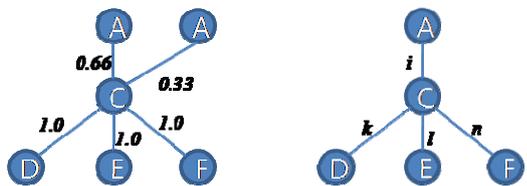


그림 3 - 매개 온톨로지(좌)와 대중 온톨로지(우)

그림 3의 좌측에 위치한 매개 온톨로지에서 페르몬 p 가 0.6 이상인 RDF 트리플만을 모아 새로운 온톨로지를 생성하면 그림 3의 우측에 위치한 바와 같이 대중 온톨로지를 생성하게 된다.

본 논문에서 제시한 온톨로지의 제작 방법은 비록 자동적인 온톨로지의 생성 방법은 아니지만 웹2.0 사용자들이 쉽게 원시 온톨로지를 생성하고 생성된 원시 온톨로지들을 모아 협업적인 방법을 통해 대중적으로 수렴 가능한 수준의 대중 온톨로지를

만드는데 그 특징이 있다. 본 논문에서 제시하는 온톨로지 제작 방법을 평가하기 위하여 온톨로지 제작 도구로써 많이 사용되는 Protege와의 성능을 다음 표 2와 같이 비교하였다.

표 2 - Protege와의 비교

항목	Protege	집단지성
생성 자동화	반자동	반자동
생성주체	전문가	비전문가
참여자	소수	불특정 다수
GUI	지원	지원
문법	OWL-DL	RDF/OWL-lite
제작난이도	중	하
추론	가능	불가능

본 논문에서 제시하는 온톨로지의 제작 방법 및 아키텍처는 온톨로지 생성 시에 많이 사용되고 있는 Protege에 비하여 불특정한 비전문가 다수에 의해 온톨로지를 생성할 수 있다는 장점을 지니고 있으며 상대적으로 온톨로지를 생성하는 난이도가 낮다. 반면 사용자들이 직접 원시 온톨로지를 생성해야하므로 사용할 수 있는 문법에 한계가 있으며 생성한 온톨로지를 통해 추론하는 것이 가능한지에 대한 검증은 거친바 없다는 한계를 지니고 있다. 그러나 이런 한계에도 불구하고 웹2.0 기반 웹 시스템에서 다양한 도메인의 대중성이 있는 온톨로지를 다수 사용자에게 의해 생성할 수 있는 장점을 가지고 있어 시맨틱 웹의 대중화에 기여할 수 있을 것으로 예상된다. 그리고 본 논문에서 제시한 방법으로 제작된 온톨로지들은 다양한 도메인의 지식을 표현하고 있으므로 어느정도의 가공 과정을 거쳐 각 도메인의 연구 및 개발에서 활용이 가능하다.

본 논문에서는 집단지성의 개념을 활용한 온톨로지 생성 시스템의 설계 및 방법을 보이고 있다. 향후 연구에서는 본 논문에서 보인 온톨로지 생성 방법의 아직 발견되지 않은 오류를 찾아 보완할 것이며 블로그나 위키와 같은 웹2.0 기반 어플리케이션에서 실제로 활용할 수 있도록 시스템을 구현할 것이다.

6. 결론

웹2.0의 대두를 통해 최근의 인터넷 환경은 분산 컴퓨팅 및 서비스 기반 인터넷으로 빠르게 변화하고 있다. 이 가운데, 웹2.0으로의 변화를 가속화 시키는 중요 기술로서 집단지성이 다양한 분야에서 활용이 되고 있다. 집단지성은 분산 환경하에서 사용자들의 지식 및 경험이 축적되어 대중적으로 활용 가능한

형태의 정보를 만들어 낼 수 있다는 것이 그 특징이다.

한편 2001년 시맨틱 웹이 제창된 이후 시맨틱 웹의 중요한 기술 요소인 온톨로지의 생성은 다양한 연구와 개발이 이루어져 왔음에도 불구하고 현재까지 대량의 온톨로지를 생성할 수 있는 방법은 제시되지 못하고 있다.

본 논문에서는 웹2.0의 중요 개념중에 한가지인 집단지성을 활용하여 시맨틱 웹 분야의 오랜 난제였던 온톨로지의 대량 생성을 가능케 하는 시스템을 제안하였으며 집단지성을 이용한 온톨로지의 생성 방법을 제시하였다. 본 논문에서 제시한 온톨로지의 생성 시스템 및 방법은 웹2.0 기반 사용자 어플리케이션에서 원시 온톨로지를 제작하고 제작된 원시 온톨로지는 온톨로지로서 전송되어 매개 온톨로지로서 통합된다. 매개 온톨로지는 전송된 원시 온톨로지들의 페르몬 값을 가지고 있어 이를 이용하여 대중 온톨로지를 생성한다. 이 방법을 활용하면 다양한 도메인의 온톨로지를 비록 전문가가 아니지만 다수의 사용자가 참여하여 대량으로 생성할 수 있다. 뿐만 아니라 작성된 온톨로지를 활용하여 지능형 웹 기반 서비스의 연구 및 개발의 디딤돌이 될 것으로 기대한다.

참조 문헌

- [1]Mark Greaves, "Semantic Web and Web 2.0" Web Semantics: Science, Services and Agents on the World Wide Web 6 (2008) 1-3
- [2]Luke Tredinnick, "Web 2.0 and Business : A pointer to the intranets of the future?", Business Information Review, 2006
- [3]T.Berners-Lee, J.Hendler, and O.Lasilla "The Semantic Web", Scientific American, May 2001.
- [4]Davies, John, Fensel, Dieter, Harmelen Frank Van. 2003. Towards the semantic web: ontology-driven knowledge management. [s. l.]: Wiley.
- [5]Dean Jones, Trevor Bench-Capon, Pepijin Visser. 1998. Methodologies For Ontology Development. In Proc.
- [6]Protege, <http://protege.stanford.edu>
- [7]Protege OWL API Programmer's Guide, <http://protege.stanford.edu/plugins/owl/api/guide.html>
- [8]Koji Zettus & Yasushi Kiyoki, "Towards Knowledge Mangement Based on Harnessing Collective Intelligence on the Web", LNAI 4248, 2006
- [9]James Surowiecki, "Wisdom Of Crowds", Random House, 2005
- [10]OWL : <http://w3g.org/TR/2004/REC-owl-features-20040210/>
- [11]Rubén Prieto Díaz. A Faceted Approach to Building Ontologies. the 2003 IEEE International Conference on Information Reuse and Integration, IRI - 2003, Las Vegas, NV, USA, October 27-29, 2003.
- [12]Vijayan, Sugumaran, Veda C. Storey. Ontologies for conceptual modeling: their creation, use, and management. Data & Knowledge Engineering Volume 42, Issue 3, September 2002, pp. 251 - 271
- [13]Thanh Tho Quan, Siu Cheung Hui, Alvis Cheuk M. Fong, Tru Hoang Cao. Automatic Generation of Ontology for Scholarly Semantic Web. ISWC 2004: Third International Semantic Web Conference, Hiroshima, Japan, pp 726-740, November 7-11, 2004.
- [14]J.M. Cueva Lovelle et al. Swarm Intelligent Surfing in the Web" ICWE 2003, LNCS 2722, pp. 431-440, 2003
- [15]James Surowiecki, "Wisdom Of Crowds", Random House, 2005
- [16]Szuba T.: Computational Collective Intelligence. Wiley and Sons NY. 2001, Wiley Book Series on Parallel and Distributed Computing.
- [17]RSS Advisory Board. Really Simple Syndication: RSS 2.0.1 Specification (revision 6). Retrieved March 2, 2006, <http://www.rssboard.org/rss-2-0-1-rv-6>.