

시맨틱 웹 포털에서의 검색과 시각화 방법 연구

이명진¹, 이기준¹, 박상언², 홍준석², 김우주¹

1. 연세대학교 정보산업공학과
2. 경기대학교 경영정보학과

Search and Visualization Method on the Semantic Web Portal

Myungjin Lee, and Ki Jun Lee, Sangun Park, June Seok Hong, Wooju Kim

요 약

웹에서의 정보가 지속적으로 늘어남에 따라 현재의 웹은 더욱 더 많은 한계를 드러내고 있다. 정보검색의 측면에서 본다면 웹 페이지는 사람이 이해하기 위한 표현 정보만을 담고 있기 때문에 사용자는 단순히 키워드의 포함 여부에 따라 많은 문서를 검색 결과로 제공받게 되며, 이들 사이에 필요한 정보를 발췌하는데 많은 시간을 소비하게 된다. 이러한 문제를 해결하기 위한 노력의 일환으로 W3C에 의해 시맨틱 웹이 제안되었다. 시맨틱 웹은 자원과 자원간의 관계 정보로 이루어진 온톨로지를 기반으로 하고 있으며, 따라서 사용자는 온톨로지 기반의 검색을 통해 의미 있는 정보를 제공할 수 있다. 온톨로지는 의미적인 정보를 담고 있기 때문에 사용자에게 적절한 정보를 제공해 줄 수 있는 새로운 검색 방법과 이를 사용자에게 보여줄 수 있는 시각화 방법을 요구하게 된다.

본 연구에서는 온톨로지를 기반으로 한 의미적 정보의 연관관계를 찾아나가는 검색 방법과 이를 사용자에게 제공해 주기 위한 시각화 방법을 제안하고자 한다.

논문에서 제안하고자 하는 바는 온톨로지로부터 사용자의 질의와 관련된 자원과 다른 자원 사이의 관계를 검색해서 제공하고자 하는 것이다. 이를 통해 사용자는 단순히 질의어가 포함된 결과가 아닌 질의어와 다른 자원간의 관계를 파악할 수 있으며, 자원 간의 관계를

기반으로 탐색해 나갈 수 있다.

Abstract

As the information of web dramatically increase, the existing web reveals more and more limitations in information search because web pages are designed only for human consumption by mixing content with presentation. In order to improve this situation, the Semantic Web comes on the stage by W3C. Semantic web is based on ontology that defines relationships between resources and it is enough to bring a significant advancement in web search. But to do this, the Semantic Web must provide a novel search and visualization methods which can make users instantly and intuitively understand why and how the results are retrieved because ontology has formal explicit descriptions of meaning. In this paper, we propose a semantic association-based search methodology that consists of how to find relevant information for a given user's query in the ontology, that is, a semantic network of resources and properties and how to provide proper visualization and navigation methods on the results. From this work, users can search the semantically associated resources for their query and also navigate such associations between resources.

1. 서론

웹이 등장한 이후 인터넷에서는 많은 양의 정보들이 디지털화 되었으며 엄청난 속도로 정보의 양이 증가하고 있다. 하지만 이러한 현상은 사용자가 원하는 정보를 검색하는데 있어 이를 더욱 더 어렵게 만들고 있다. 웹에서의 정보 검색에 있어서 이러한 문제점을 해결하기 위한 노력의 일환으로

시맨틱 웹이 W3C에 의해 제안되었으며, 시맨틱 웹은 기계가 이해할 수 있는 정형화된 의미 기반의 기술[18]을 통해 사용자에게 보다 유용한 정보를 제공해 줄 수 있다. 시맨틱 웹 포털은 시맨틱 웹 기술을 이용하여 구축된, 보다 진보된 형태의 웹 포털로써, 특정 도메인에 대한 정보와 서비스를 제공하는 웹 어플리케이션이다.

이러한 시맨틱 웹 포털은 정형화된 기술 논리를 담고 있는 온톨로지[17]를 이용해서 특정 도메인에 대한 자원을 기술한다. 이러한 시맨틱 웹 포털의 특징 때문에 기존과는 다른 정보 검색 방법이 필요하다. 기존의 검색 방법인 키워드 기반의 검색 방법은 단순히 페이지에 사용자의 질의어가 포함되어 있는지의 여부를 가지고 결과를 제공해 주지만, 온톨로지를 기반으로 하는 의미적 검색에서는 자원간의 관계를 기반으로 검색을 수행해 나가게 된다. 따라서 의미적 검색 방법에서는 자원과 자원을 연결해 주는 관계가 중요한 요소라 할 수 있다. 본 논문에서는 자원과 속성으로 이루어진 시맨틱 네트워크를 어떻게 검색하고 확장해 나갈 것인가에 대해 논하고자 한다. 더불어 검색 결과를 보여주는 것 또한 검색어를 기반으로 하는 검색 방법과 다르기 때문에 중요한 요소 중 하나이다. 검색어 기반의 검색 방법은 단지 문헌과 검색어 간의 유사도를 기반으로 검색 결과를 순위화 하였지만, 의미 기반 검색에서는 주어진 질의어와 각 자원간의 관계를 기반으로 순위화를 해야만 한다.

본 논문에서는 자원과 속성으로 이루어진 온톨로지를 기반으로 사용자의 질의어와 관련된 정보를 찾아나가는 의미 기반 검색 방법과 검색된 결과를 사용자에게 시각화하는 방법을 제시하고자 한다. 이를 통해 사용자는 단순히 질의어를 포함하고 있는 결과가

아닌 질의어와 다른 자원 간의 관계를 파악할 수 있게 될 것이다. 본 논문은 다음과 같은 구성을 가진다. 2장에서는 이와 관련된 다양한 관련 연구를 기술하고, 3장에서는 의미 기반 검색 방법을 제시할 것이다. 4장에서는 제시된 검색 방법을 구현한 프로토타입 형태의 시맨틱 웹 포털에 대해 논하고자 한다.

2. 관련연구

본 논문의 주된 논점은 기존 키워드 검색과는 구별되는 사용자에게 최적의 정보를 제공할 수 있는 시맨틱 검색과 시각화 방법을 제안하는 것과 그것을 이용한 특정 분야의 시맨틱 웹 포털을 구현하는 것이다. 이러한 관점에서 기존에 구현된 시맨틱 웹 포털과 프로토타입 시스템들을 소개한다. SEAL(Semantic web portAL)[12]은 일반적인 시맨틱 포털 구현에 대한 내용을 제안한다. 포털에서의 의미적인 정보 제공과 접근 방법, 또한 정보 유지 방법 등에 대한 내용을 소개하고 있다. 다음으로, OntoWeb Portal[14]은 시맨틱 웹 관련 학계와 업계의 기관들끼리 정보를 공유할 수 있는 커뮤니티 포털이라 할 수 있다. ODESeW[2]는 각 분야의 온톨로지들로부터의 지식을 자동 생성하는 것에 초점을 두고 있는 시스템이다. 이 시스템은 지식 포털을 위한 다양한 기능들을 제공하고 있다. 마지막으로, OntoWebber[8]는 지식 포털을 구현하고

유지하기 위한 온톨로지 기반의 도구이다. 지식 포털을 모델링하고 지식 포털의 구현과 유지에 관한 다양한 정보를 제공한다.

또한 시맨틱 웹에서의 온톨로지 시각화와 관련하여 다양한 시각화 방법과 기술들을 다룬 논문[1]을 소개하고자 한다. 이 논문에서는 온톨로지의 시각화 방법론을 다섯 가지 분류로 나누어 소개하고 있다. 첫 번째로 Indented List는 온톨로지를 윈도우 탐색기와 같은 트리 형태로 제공하는 방법으로써, Protégé[15], OntoEdit[13], 그리고 Kaon[9]이 이러한 형태의 시각화 방법을 사용하고 있다. 두 번째는 Node-Link와 Tree 형태의 방법으로써, 이는 노드 사이의 연결을 이용하여 표현하는 방법이다. 세 번째는 상위 개념 안에 하위 개념을 포함하는 방법을 이용하는 Zoomable 방법이다. 이는 상위 개념으로부터 화면을 확대하는 방법을 이용해서 그에 속한 하위 개념을 확인할 수 있다. 온톨로지 시각화를 위한 네 번째 방법은 Space filling이다. 이 방법은 상위 개념의 화면 공간에 하위 개념들을 채움으로써 온톨로지를 표현하는 방법이다. 이러한 방법에서 각 개념들을 나타내는 영역의 크기는 클래스에 포함된 인스턴스의 수나 자원이 가지고 있는 속성의 개수에 따라 달라지기 때문에 이와 관련된 정보를 직관적으로 확인할 수 있다. 마지막 방법은 Context + Focus와 Distortion으로써, 이는 마인드 맵과 같은 시각화 방법을 제공한다. 각 노

드는 상위 노드를 중심으로 주변에 확장된 형태로 표현되며, 확장된 노드를 줄여서 시각화 범위의 크기를 줄일 수도 있다.

마지막으로 시맨틱 웹에서 각 자원의 관계로부터 중요도를 결정하기 위한 방법을 연구한 논문이 있다[10]. 저자는 자원과 자원 사이의 정보량에 따라 특이치를 산출하였다. 본 논문에서는 이를 확장하여 보다 더 가치 있는 관계를 발견하고 이를 중심으로 검색을 확장해 나가는 방법을 사용할 것이다.

3. 의미 기반 검색과 탐색 방법

본 연구의 목적은 시맨틱 웹 포털에서 온톨로지를 기반으로 시맨틱 검색과 시각화 방법을 제안하는 것이다. 본 논문에서는 이해를 돕기 위해 전자 분야 도메인의 정보에 대한 의미적 연관 관계로 이루어진 온톨로지를 구축하였다. 본 논문에 사용된 온톨로지를 구축하기 위한 첫 번째 단계로써 전자신문 사이트와 특허 사이트, 그리고 논문 사이트로부터 뉴스, 인물, 논문, 제품, 특허, 그리고 기술 7개 분야의 정보를 수집하였다. 일반적으로 각 정보의 출처 별로 정보에 대한 형식과 접근 방법이 다르기 때문에 서로 다른 기관들간에 매끄럽고 효율적인 정보의 교환과 공유가 불가능하다. 전자 부품을 생산하는 회사의 사람들 또는 전자 부품 산업과 관련된 일을 하는 사람들은 새로운 기술, 동향, 전자 부품 생산에 민감하기 때문에,

항상 전자 관련 신문을 읽으며, 논문, 기술, 특히 관련 정보를 얻기 위해 하루에 최소 5~7개의 사이트를 방문한다. 이들은 의미적으로는 연결되어 있지만 정보의 출처가 서로 다르기 때문에 필요한 정보를 검색하는데 많은 시간을 소비한다. 예를 들어, 특허 관련 정보는 특허를 제출한 회사나 특허의 기술과 같은 정보와 의미적으로 연결되어 있다. 이러한 정보의 의미적 연관 관계는 의미 기반 검색의 필요성을 입증하는 하나의 예이다.

시맨틱 웹 포털에서의 온톨로지는 도메인 지식에 대한 논리적 기술을 담고 있으며, RDF(Resource Description Framework)[5]를 기반으로 한 OWL(Web Ontology Language)[4]로 기술된다. RDF의 데이터 모델은 웹에 존재하는 자원을 주어-술어-목적어인 트리플(triple) 형태로 기술한다. 모든 자원이 이와 같은 형태에 따라 의미적으로 연결되어 있기 때문에 사용자는 자원들 사이의 의미적인 정보를 추출할 수 있으며, 자원 사이의 의미적 연결 관계를 결과로 얻을 수 있다.

본 논문에서 제안하는 의미 검색의 처리 과정은 질의 단계(querying)와 탐색 단계(navigation)로 이루어진다. 질의 단계는 질의어를 이용해서 온톨로지로부터 관련된 자원을 검색해 나아가는 단계이며, 두 번째 단계인 탐색 단계는 질의 단계를 통해 얻어진 결과를 사용자에게 시각화하기 위한 방

법을 제공한다.

3.1 질의 단계

온톨로지는 정형화되고 공유가 가능한 형태로 도메인 지식을 기술하기 위한 방법을 제공하며, 이를 위해 클래스(class)들 간의 계층 구조와 클래스 사이의 관계(relation), 클래스에 속한 인스턴스(instance)를 이용해서 기술한다. 질의 단계의 목적은 온톨로지로부터 사용자의 질의어와 관련 있는 클래스와 인스턴스를 찾아나가는 것이다. 질의 단계는 질의어와 유사한 자원을 찾는 시드 검색과 시드 인스턴스들로부터 검색을 확장해 나가는 확장 단계로 이루어진다. 첫 번째 단계인 시드 검색은 전통적인 질의어 검색과 유사하며, 시드 검색의 결과는 질의어를 포함하고 있는 클래스와 인스턴스들이다. 두 번째 단계인 확장 단계는 시드 검색의 결과로부터 그와 연결된 관계를 이용해서 의미적 관계를 확장해 나가는 단계이며, 이를 위해서는 확장의 여부를 결정하기 위한 새로운 방법이 필요하다.

우선 온톨로지는 다음과 같은 세 개의 원소의 집합으로 이루어진다[11].

$$KB = (I, C, P)$$

C 는 클래스들의 집합을 나타내며, P 는 속성의 집합을 나타낸다. I 는 모든 클래스에 속한 인스턴스들의 집합을 나타낸다. 그리고 $c_i \in C$ 는 하나의 클래스를 나타내며, $p \in P$ 는 속성, $i_j \in I$ 는 클래스 c_i 에 속

한 하나의 인스턴스를 의미한다.

질의 단계에서는 온톨로지에서부터 사용자가 입력한 질의어와 유사한 시드 인스턴스들을 검색하게 되며, 여기서 검색된 시드 인스턴스들은 확장 단계에서의 후보들이 된다. 시드 인스턴스의 집합은 다음과 같이 정의된다.

$$Seed_q = I_q \cup \{i \mid i \in C_q\}$$

$Seed_q$ 는 시드 인스턴스들의 집합을 나타내며, q 는 사용자가 입력한 질의어를 나타낸다. I_q 는 $I_q = \{i_j \mid \text{name of } i_j \text{ contains } q\}$, C_q 는 $C_q = \{c_i \mid \text{name of } c_i \text{ contains } q\}$, 마지막으로 $\leq C$ 는 클래스 C 의 하위 클래스들을 의미한다.

사용자가 의미 검색을 하기 위해 *BLU* (*Back Light Unit*)를 입력하면, 온톨로지에서부터 *BLU* 질의어를 포함하는 모든 인스턴스와 클래스(하위 클래스를 포함)에 속하는 인스턴스들을 후보로 선택하게 된다.

시드 인스턴스를 검색한 후 다음 단계는 시드 인스턴스와 연결된 개념을 찾아나가는 단계이다. 기존의 키워드 검색과 다른 점은 자원과 자원 사이에 연결된 의미 네트워크를 따라 검색을 수행한다는 것이다. 전자 온톨로지의 모든 개념은 그래프 구조와 같은 형태로 서로 연결되어 있다. 그림 1은 자원 사이의 의미적 관계를 탐색해나가는 과정을 보여주고 있다.

```

Calculate Path Decision Threshold  $\theta(p)$  for all property  $p$ 
FOR all instance  $s_i$  in  $Seed_q$  do
  Set  $r$  with seed instance  $s_i$ 
  WHILE depth of path < maximum depth
    FOR all  $triple(r, p, t)$  in Semantic Graph do
      IF  $\theta(r, p) < \theta(p)$  THEN
        Expand path into Semantic Graph
      ENDIF
    ENDFOR
    Set  $r$  with the last resource of the path
  ENDWHILE
ENDFOR
    
```

그림 1. 의미 검색 알고리즘

그림 1의 알고리즘은 주어진 자원으로부터 연결된 관계의 특이치를 산출하여 확장의 여부를 결정하게 된다. 의미 검색을 수행하기 위해 각 시드 인스턴스와 연결된 속성의 한계치(Path Decision Threshold)를 산출하고, 이에 따라 검색의 확장 여부를 결정하게 된다. 그림 1의 알고리즘에서 depth of path는 시드 인스턴스와 연결된 노드의 수를 의미하며, 한계치 θ 는 시드 인스턴스와 연결된 속성의 확장 여부를 결정하기 위한 기준값을 나타낸다. 본 연구에서는 연결된 노드의 최대 길이를 3으로 제한하였다. 즉, 의미 검색의 결과는 최대 3개의 연결까지를 검색 결과로 제시해 준다. 검색되는 노드의 길이는 검색 결과에 많은 영향을 미치게 되며, 이에 대해서는 추후 연구로 남겨둔다.

그림 1 알고리즘의 가장 중요한 요점은 검색의 확장 여부를 결정하는 것이다. 만약 어떤 인스턴스와 연결된 온톨로지 상의 모든 속성들로 연결을 확장하게 되면 그 관계

는 너무나 많고 쓸모 없는 검색 결과가 될 것이다. 정보 이론에 따르면, 정보의 양은 이벤트가 발생하는 확률의 음의 로그 값으로 계산된다[3]. 본 논문의 검색 방법은 많은 사람이 알고 있는 정보보다는 적은 사람이 알고 있는 정보에 더 많은 가중치를 부여한다. 정보 이론에 따라, 적은 확률로 나타나는 연결의 확장을 해 나가기 위해 모든 속성에 한계치를 할당하고, 이 값은 다음 자원으로의 확장 여부를 결정하는 기준이 된다.

Kemafor의 연구[10]에서 저자는 정보의 양을 측정하기 위해 θ -특이성을 사용하였다. 이는 해당 속성의 도메인(domain)과 레인지(range)에 해당하는 클래스의 인스턴스들이 포함된 집합에서 유일성(uniqueness)을 계산하는 방법이다. 다시 말해, θ -특이성은 속성이 사용될 수 있는 모든 경우의 수와 실제 속성이 사용된 경우의 수의 비율을 의미하며, 이는 자원에 사용된 속성의 빈도로 설명될 수 있다.

본 연구에서는 한계치를 설정하기 위하여 앞의 연구[10]에서 제시된 θ -특이성을 사용한다. 이는 다음과 같이 정의된다.

$$\theta(p) = \frac{|(*, p, *)|}{|(x, *, y)|}$$

p 는 특정 속성을 의미하며, x 는 $x \in p.domain$ 을, y 는 $y \in p.range$ 를 나타낸다. 이것은 많이 사용된 속성보다 적게 사

용된 속성이 더 가치 있음을 의미한다. 하지만 이 공식은 단지 속성에 대한 정보량만이 고려되어 있다. 따라서 본 연구에서는 주어진 자원과 속성이 함께 반영된 다음과 같은 θ -특이성을 제안한다.

$$\theta(r, p) = \frac{|(r, p, *)| + |(*, p, r)|}{|(r, *, y)| + |(x, *, r)|}$$

r 은 주어진 자원을 나타내며, p 는 속성을 의미한다. 위 공식은 자원 r 을 포함하고 있는 연결의 수와 자원 r 과 속성 p 를 포함하고 있는 연결의 비율을 나타내며, 이는 주어진 자원 r 에 대한 속성 p 의 사용빈도로 설명될 수 있다.

만약 r 과 p 에 대한 θ -특이성이 한계치인 $\theta(p)$ 보다 적다면, 속성 p 로 연결된 자원으로 검색을 확장해 나간다. 이것은 자원 r 과 속성 p 의 사용빈도가 모든 자원에 대한 속성 p 의 사용빈도보다 적다면 더 가치가 있음을 의미한다.

3.2 탐색 단계

탐색 단계는 앞서 검색된 결과를 정렬하고 검색된 결과 중 사용자가 원하는 결과를 보여주기 위한 방법을 정의한다. 검색 단계의 결과는 시드 인스턴스로부터 다른 자원과 연결된 의미 관계로 제공된다. 여기에는 두 가지의 논제가 존재한다. 첫 번째로 검색 단계의 결과는 시드 인스턴스로부터 연결된 자원이 서로 다른 경로를 통해 연결되

어 있으며, 이러한 자원들 사이의 관계는 사용자에게 아주 중요한 요소 중 하나이다. 또한 앞서 검색된 경로는 시드 인스턴스와 경로의 마지막 자원이 같을 수 있기 때문에 어떻게 이러한 경로를 사용자에게 직관적으로 보여줄 것인가 하는 것이 하나의 문제이다. 두 번째는 검색 결과에 대한 분류와 순위화 정책이다. 일반적으로 의미 검색에 있어서의 검색 결과는 질의어와 연결된 인스턴스와 클래스가 많기 때문에 질의어 검색에 비해 그 결과의 수가 많다. 따라서 어떻게 검색 결과를 분류하고 순위화 하여 제공해 줄 것인가를 결정하는 방법 또한 중요한 문제 중 하나이다.

검색의 결과를 사용자에게 시각화하기 위해 탐색 단계는 세 단계로 이루어진다. 첫 번째 단계는 검색 결과를 정보의 출처에 따라 회사, 뉴스, 사람, 논문, 제품, 특허, 그리고 기술 7 개의 주제별로 분류하는 것이다. 이로써 사용자는 자신의 관심 영역에 대한 정보를 보다 쉽고 빠르게 찾아볼 수 있게 된다.

두 번째 단계는 각 영역에 속한 검색 결과를 내부적으로 공통의 자원을 갖지 않는 경로[19] 단위로 분류하는 것이다. 예를 들어, 두 개의 경로 $e_1, P_1, e_2, P_2, e_3, \dots, e_{n-1}, P_{n-1}, e_n$ 와 $f_1, Q_1, f_2, Q_2, f_3, \dots, f_{n-1}, Q_{n-1}, f_n$ 로 e_1 과 e_n , f_1 과 f_n 이 연결되어 있을 때, e_1 과 f_1 이 동일한 자원이고 e_n 과 f_n 만이 동일한 자원이면 이 두 경로는 내부적

으로 공통의 자원을 갖지 않는 경로로써 같은 그룹으로 분류된다. 이러한 분류로부터 사용자는 시드 인스턴스와 다른 자원 간의 다양한 경로를 초기 검색 결과로 파악할 수 있다.

탐색 단계의 마지막은 순위화이다. 순위화는 그룹 간의 순위화와 그룹 안의 의미 경로 간의 순위화로 이루어진다. 순위화는 다음과 같은 순위화 방식에 따라 이루어진다.

$$rank(sim(q, e_1), distance(e_1, e_n))$$

q 는 사용자가 입력한 질의어를 나타내며, e_1 은 시드 인스턴스를 나타내고, e_n 은 의미 경로에서 마지막 노드를 의미한다. 순위화 알고리즘에서 $distance(e_1, e_n)$ 은 e_1 과 e_n 사이의 거리를 나타내며, $sim(q, e_1)$ 은 시드 인스턴스와 질의어간의 단어 유사도를 의미한다. 순위화의 첫 번째 고려 대상은 시드 인스턴스와 질의어간의 유사도이다. 본 논문에서는 유사도를 측정하기 위해 문헌에서의 용어 사용 빈도를 측정하는 단어 빈도(term frequency)[7] 기법을 사용하였다. 순위화의 다음 고려 대상은 결과로 제시된 경로에서 시드 인스턴스와 마지막 노드 간의 거리이다. 만약 길이의 값이 0 이라면, 이것은 마지막 노드가 시드 인스턴스를 의미하며, 결과들 중에 첫 번째로 순위화가 된다. 다음은 시드 인스턴스에 직접적으로 연결된 자원을 가진 경로가 위치하

게 된다. 만약에 단어간 유사도 값이 동일하다면 길이에 따라 순위가 결정되게 된다. 즉, 검색 결과의 순위는 $sim(q, e_i)$ 에 의해 순위화가 이루어진 후 $distance(e_i, e_n)$ 에 의해 최종 순위가 결정된다.

4. 의미 검색을 위한 애플리케이션

본 절에서는 사용자가 입력한 질의어를 포함하고 있는 시드 인스턴스와 연결된 검색 결과를 사용자에게 보여주기 위한 방법을 제시하고자 한다. 사용자가 검색된 결과를 보다 쉽게 이해하고, 질의어와 연관된 정보를 파악하는 방법에 중점을 두고자 하며, 이를 위해 몇 가지의 시각화 방법을 제시하고자 한다. 그림 2는 한국 전자 상거래 협회(KOEB)의 지원하에 만들어진 한국 전자 산업 온톨로지를 기반으로 하여 본 논문에서 제시하고 있는 시맨틱 웹 포털의 화면 구조를 보여주고 있다. 시맨틱 웹 포털의 사용자 인터페이스는 네 개의 부분으로 이루어져 있다.



그림 2. 시맨틱 웹 포털의 인터페이스

화면의 상단은 사용자의 질의어를 입력하기 위한 입력 상자이며, 왼쪽은 검색의 시작점이 되는 시드 인스턴스를 인스턴스가 속한 클래스 별로 분류하여 보여주는 화면이다. 왼쪽에 있는 특정 시드 인스턴스를 클릭하면, 사용자는 해당 시드 인스턴스와 관련된 검색 결과만을 확인할 수 있다. 인터페이스의 가운데 화면은 뉴스, 인물, 논문, 제품, 특허, 기술 영역에 따라 분류된 검색 결과를 보여준다. 각 검색 경로는 경로의 마지막 노드가 속한 클래스에 따라 분류되며, 각 결과는 앞서 논한 순위화 방법에 따라 정렬되어 출력된다. 또한 사용자는 각 검색 결과의 맨 뒤에 위치한 (+) 버튼을 클릭하여, 시작 노드와 종료 노드가 같지만 다른 경로를 가지고 있는 다양한 검색 결과를 확인할 수 있다. 마지막으로 오른쪽 화면은 사용자가 지금까지 검색한 질의어의 검색 리스트를 보여준다.

그림 3은 LED_BLU의 검색 결과 중 제품 분류에 대한 검색 결과를 보여주고 있다.



그림 3. 제품 분류의 LED_BLU 검색 결과

그림 3 검색 결과의 첫 번째 줄은 가장 상위에 순위화된 LED_BLU (시드 인스턴스)와 LED_12 (마지막 노드) 사이의 의미 경로를 보여주고 있다. 경로의 마지막 노드를 처음에 보여주는 이유는 사용자가 시드 인스턴스보다는 검색 결과에 더 많은 관심을 갖기 때문이다. 첫 번째 검색 결과를 보다 자세히 살펴보면 이는 주어-술어-목적어 형태의 두 개의 트리플로 이루어져 있다. 즉, LED_12의 타입이 LED라는 (LED_12 type LED) 트리플, 그리고 (LED consistOf(-) LED_BLU) 트리플로 이루어져 있다. 여기서 consistOf 술어에 있는 (-) 기호는 consistOf의 역 관계(inverse property)를 의미하며, 따라서 LED는 LED_BLU의 구성 요소임을 의미한다. 이와 같은 결과를 통해 사용자는 질의어가 다른 자원과 어떤 연결을 가지고 있는지를 파악할 수 있다.

다음의 그림 4는 첫 번째 결과를 W3C의 RDF 유효성 검사기(RDF Validator)[16]를

이용해서 출력한 데이터 모델이다.

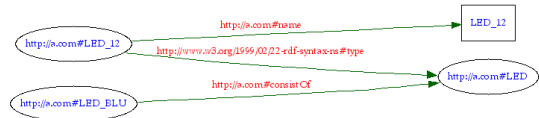


그림 4. 첫 번째 결과의 RDF 그래프

첫 번째 검색 결과의 마지막에 있는 (+) 버튼은 LED_12와 LED_BLU 사이의 다른 경로가 존재함을 의미한다. 사용자가 (+) 버튼을 클릭하게 되면 그림 5와 같이 LED_BLU와 LED_12 사이의 모든 연결을 보여주게 되며, 이를 통해 사용자는 두 자원이 서로 어떤 관계를 가지고 있는지 다양한 경로를 통해 알 수 있다.



그림 5. 그룹에서의 다양한 검색 결과

본 논문에서 제시한 의미 검색 시스템은 두 가지의 추가적인 기능을 지원한다. 첫 번째는 AND 연산이다. 만약 사용자가 두

단어와 관련된 개념을 찾고자 할 경우 AND 연산자를 이용해서 두 질의어 사이의 관계를 파악할 수 있다. 기존의 전통적인 검색 시스템에서는 두 개의 질의어를 입력하면 단순히 문헌에 두 질의어의 포함 여부에 따라 결과를 제시하여 주었다. 하지만 본 의미 검색 시스템에서는 각 질의어를 포함하고 있는 시드 인스턴스 간의 의미 관계를 파악할 수 있다.

두 질의어 사이의 AND 검색 결과는 두 질의어 사이에 자원과 속성으로 연결된 의미 경로이다. 즉, 두 질의어를 포함하고 있는 자원으로부터 만들어진 독립적인 결과 사이의 교집합과 같다고 볼 수 있다. 그림 6은 *삼성*과 *BLU* 질의어 사이의 AND 검색 결과를 보여주고 있다.



그림 6. AND 연산자의 검색 결과

그림 6의 첫 번째 줄은 삼성전자가 특허

의 소유주이며, 이 특허는 *hasKeyword*와 *consistOf*의 역관계로 *BLU*와 연관되어 있음을 보여주고 있다. 이러한 결과로부터 사용자는 *삼성*과 *BLU*에 연관 있는 개념이나 인스턴스들 사이의 관계 정보를 파악할 수 있다.

두 번째 기능은 AND 검색의 확장인 중심 주제어 검색이다. 만약 사용자가 하나의 주제어와 관련된 추가 결과를 얻고 싶다면, 그 주제를 중심주제어로 정하여 검색을 진행할 수 있다. 일단, 사용자가 하나의 키워드를 중심주제어로 등록하면 알고리즘은 검색 결과와 중심주제어 사이의 관계, 즉 AND 검색 결과를 보여주는 방식으로 진행된다.

5. 결론

시맨틱 웹의 시각화는 기존의 질의어 검색과 그 검색 방법과 순위화 방법이 다르기 때문에 시맨틱 웹 기술에 있어서 아주 중요한 연구 과제 중 하나이다. 또한 이는 의미 검색 절차와 검색 결과의 특성이 반영되어 사용자의 검색 편의성을 높일 수 있는 형태로 제공되어야 한다.

본 논문에서는 전자 분야에 있어 시맨틱 시각화를 위한 의미 검색 방법과 탐색 방법을 제안하였다. 이는 질의 단계와 탐색 단계 두 단계로 이루어지며, 질의 단계에서는 온톨로지로부터 질의어를 포함하고 있는 자원과 그와 속성에 따라 연결되어 있는 다른 자원 사이의 관계를 검색해 나가는 방법을

제안하였다. 탐색 단계에서는 질의 단계에서 도출된 결과를 정보의 출처에 따라 7개의 분류로 나누고, 유사도와 자원 사이의 거리를 측정하는 방법에 따라 각 검색 결과를 순위화하는 방법을 제안하였다. 여기에 더불어 본 논문에서 제안하고 있는 시맨틱 시각화 방법을 보여주기 위하여, 의미 검색 서비스를 제공하는 시맨틱 웹 포털을 구축

하였다.

향후, 시맨틱 그래프에서 두 자원 사이의 의미 관계를 측정하기 위한 다양한 방법들을 정의하고, 이러한 연구 방법을 Freebase[6]와 같은 소셜 네트워크(social network)에 적용해 볼 예정이다.

참고문헌

- [1] Akrivi K., Constantin H., George L., Costas V. and Eugenia G. Ontology Visualization Methods – A Survey, ACM Computing Surveys (2007).
- [2] Corcho, Ó., Gómez-Pérez, A., López-Cima, A., López-García, V. and María del, C., S. ODESeW: Automatic Generation of Knowledge Portals for Intranets and Extranets. Proceedings of the 2nd International Semantic Web Conference, Springer (2003), 802–817.
- [3] David, J.C.M. Information Theory, Inference, and Learning Algorithms. Cambridge University Press, 2003.
- [4] Deborah, L.M. and Frank, v.H. OWL Web Ontology Language Overview. W3C Recommendation (Feb. 2004).
- [5] Frank, M. and Eric, M. RDF Primer. W3C Recommendation (Feb. 2004).
- [6] Freebase. <http://www.freebase.com/>.
- [7] Gerard, S. and Michael, J.M. Introduction to Modern Information Retrieval. McGraw-Hill, New York, USA, 1983.
- [8] Jin, Y., Decker, S. and Wiederhold, G. OntoWebber: Model-Driven Ontology-Based Web Site Management, Proceedings of the 1st Semantic Web Working Symposium 2001, 529–547.
- [9] Kaon. <http://kaon.semanticweb.org/>.
- [10] Kemafor, A., Angela, M. and Amit, S. SemRank: Ranking Complex Relationship Search Results on the Semantic Web. The 14th international conference on World Wide Web, ACM (2005), 117–127.
- [11] Liang, B., Tang, J. and Li, J., Association search in semantic web: search + inference. The 14th international conference on World Wide Web, ACM (2005), 992–993.
- [12] Maedche, A., Staab, S., Studer, R., Sure, Y. and Volz, R. SEAL-Tying Up Information Integration and Web Site Management by Ontologies. IEEE Data Engineering Bulletin 25 (2002), 10–17,.
- [13] OntoEdit. <http://www.ontoknowledge.org/tools/ontoedit.shtml/>.

- [14] Peter, S., Daniel, O., Raphael, V., Jijuan, Z., Mustafa, J., York, S., Rudi, S. and Robert, M. OntoWeb – A Semantic Web Community Portal. Proceedings of the 4th International Conference on Practical Aspects of Knowledge Management, ACM (2002), 189–200
- [15] Protégé. <http://protege.stanford.edu/>.
- [16] RDF Validator. <http://www.w3.org/RDF/Validator/>.
- [17] Thomas, G. A translation approach to portable ontology specifications. Knowledge Acquisition 1993, 199–220.
- [18] Tim, B.L., James, H. and Ora, L. The Semantic Web. Scientific American Magazine, 2001.
- [19] West, D.B. Introduction to Graph Theory 2/E. Prentice Hall, 2001.

저자소개

이명진(e-mail: xml@yonsei.ac.kr)은 현재 연세대학교 정보산업공학과 지능정보시스템 연구실 박사 과정에 재학 중에 있으며, 다수의 XML 및 XML Web Services 서적을 집필하였다. 관심분야는 시맨틱 웹과 의미기반 검색, 시맨틱 웹 포털, 시맨틱 웹 서비스이다.

이기준(e-mail: backdale@msn.com)은 2008년 연세대학교 컴퓨터산업공학과 학사를 취득하고, 2008년부터 현재까지 연세대학교 대학원 정보산업공학과 석사과정에 재학 중이다. 관심분야는 웹 서비스, 유비쿼터스 웹 서비스, 검색 엔진, 시맨틱 웹 서비스이다.

박상연(e-mail: supark@kgu.ac.kr)은 1992년 KAIST 전산학과 학사를 취득하고, 1999년 KAIST 테크노경영대학원 경영공학과 석사를 취득하였으며, 2006년 KAIST 테크노경영대학원 경영공학과 박사를 취득하였다. 2007년부터 현재까지 경기대학교 교수로 재직 중이다. 관심분야는 전자상거래, e-Marketing, 시맨틱 웹, 지능정보시스템, 온톨로지 추론이다.

홍준석(e-mail: junehong@kyonggi.ac.kr)은 1989년 서울대학교 경영학과 학사를 취득하고, 1991년 KAIST 경영과학과 석사를 취득하였으며, 1997년 KAIST 테크노경영대학원 박사를 취득하였다. 1999년부터 2002년까지 인제대학교 경영학과 교수로 재직하였으며, 2003년부터 현재까지 경기대학교 경영정보학과 교수로 재직 중이다. 관심분야는 지능정보시스템, 에이전트, 자동협상, 시맨틱 웹 서비스 등이다.

김우주(e-mail: wkim@yonsei.ac.kr)은 1987년 연세대학교 BBA 과정 학사 학위를 취득하고, 1994년 KAIST 경영과학 박사를 취득하였으며, 현재 연세대학교 정보산업공학과 교수로 재직 중이다. 관심분야는 시맨틱 웹, 시맨틱 웹 환경의 의사결정지원 시스템, 시맨틱 웹 마이닝, 지식 관리 및 인공지능 웹 서비스이다.