

질의응답 시스템을 위한 백과사전 기반 지식베이스와 온톨로지*

울산대학교 컴퓨터정보통신공학과,¹ 춘해대학 컴퓨터정보처리과,²
한국전자통신연구원 음성/언어연구정보센터³
최호섭¹ · 옥철영¹ · 김창환² · 왕지현³ · 장명길^{3†}

Encyclopedia-Based Knowledge Base and Ontology for Question Answering System

Ho-Seop Choe,¹ Cheol-Young Ock,¹ Chang-Hwan Kim,² Ji-Hyun Wang,³ Myoung-Gil Jang³

Department of Computer Engineering and Information Technology,¹ University of Ulsan, Ulsan, Korea

Department of Computer Information Technology,² Choonhae College, Ulsan, Korea

Speech/Language Technology Research Center,³ ETRI, Daejeon, Korea, Korea

요 약

기존의 정보검색시스템이 사용자의 질의에 의해 키워드가 포함된 의미 있는 문서를 제공하는 시스템이라면, 질의응답시스템은 사용자 질의에 맞는 정답을 적절한 언어처리 기법을 통해 텍스트로부터 추출하여 제공하는 시스템이다. 이러한 언어처리 기법을 이용한 질의응답 시스템에서 시스템의 성능 향상에 도움을 줄 수 있는 것이, 실세계의 지식을 저장하고 있는 지식베이스라 할 수 있다. 지식베이스가 가지고 있는 실세계의 지식을 어떻게 효율적으로 활용하느냐에 따라 질의 처리 분석과 정답 정확률을 향상시킬 수 있는 것이다. 본 논문에서는 실세계의 지식을 어느 정도 체계적·의미적으로 반영하고 있는 것을 백과사전으로 판단하여, 백과사전의 '인물' 범주(category)를 중심으로 백과사전 지식베이스의 틀을 마련하고자 하였다. 또한 어휘의 계층적 구조를 중심으로 한 온톨로지를 백과사전 지식베이스와 유기적으로 연결시킴으로써 보다 의미 있는 지식베이스를 형성하는 방안을 모색하고자 하였다.

서 론

질의응답 시스템은 자연언어에 의한 대화형 시스템(interactive system)의 일종으로, 대화로서의 자연스러움보다도 적절한 답을 검색하는 기술 또는 내부의 데이터나 지식에서 답을 추론하는 기술에 중점을 두는 시스템을 말한다. 이러한 질의응답시스템의 성능 향상 여부는 형태소 분석, 구

문 분석 등과 같은 언어처리 기법을 어떻게 효율적으로 이용하는가에 달려 있다고 할 수 있다. 더 나아가 사람들이 가지는 지식을 체계화함과 동시에 일정한 기준에 의해 연관성을 지어줄 수 있는 의미 분석·추론 등의 방법을 사용함으로써 질의응답시스템의 효용성과 성능에 큰 영향을 줄 수 있다.

의미 분석이나 추론 단계에서 많이 활용되는 것이 지식베이스¹라고 할 수 있는데 시소러스, 의미망, 온톨로지와 같은 어휘 데이터베이스(lexical database)를 비롯해, 실세계의 지식을 저장하고 있는 사실 자료(fact database) 혹은 실세계 지식베이스(real-world knowledge base)가 많이

*본 논문은 ETRI의 "코어 온톨로지 구축을 위한 기초어휘 분류 및 수집"이라는 용역 과제 일부로 연구되었으며, 사용된 백과사전 데이터는 (주)두산의 「엔사이버 백과사전」이다.

E-mail : jungokim@plaza.snu.ac.kr

E-mail : hoseop@mail.ulsan.ac.kr

E-mail : okcy@mail.ulsan.ac.kr

E-mail : chkim@choonhae.ac.kr

E-mail : jhwang@etri.re.kr

†E-mail : mgjang@etri.re.kr

¹ 언어처리에 사용되는 지식베이스는 두 가지로 요약할 수 있다. 하나는 어휘들의 계층 구조를 중심으로 형성된 어휘 지식베이스(Lexical Knowledge Base)라 할 수 있고, 다른 하나는 실세계의 지식을 체계적으로 담고 있는 사실 자료(Fact Database) 또는 실세계 지식베이스(Real-World Knowledge Base)라 할 수 있다.

활용되고 있다. 이러한 지식베이스의 효율적인 구축은 지식의 전산학적 체계화라는 점에서 의미를 가질 수 있으며, 지식기반 시스템(knowledge-based system)에서도 적절히 이용될 수 있다.

그러나 인간이 가지는 지식을 어떻게 체계화하고, 체계화된 지식을 어떻게 저장하고, 저장된 지식을 어떻게 활용해야 하는가에 대한 논점은 아직까지 국내에서는 그 틀을 마련하지 못하고 있다. 또한 지식 체계화 작업은 학제적인 연구로서, 어느 특정 한 분야의 독자적인 연구·개발은 어려우며, 단순히 웹 문서나 텍스트(또는 말뭉치) 등에서 지식을 추출하는 것도 쉬운 작업이 아니다.

본 논문에서는 이러한 지식 체계화 작업의 한 방법으로, 백과사전(encyclopedia)을 기반으로 지식 처리의 한 부류인 질의응답시스템에 필요한 지식베이스를 구축함과 동시에 이러한 지식베이스와 유기적으로 연동되는 온톨로지 구축 방안을 모색한다. 이를 위해 웹 문서나 일반 말뭉치가 아닌 백과사전을 보편적인 지식의 집합으로 간주하여, 백과사전 범주(category) 중 '인물(person)' 범주의 표제어 항목들을 대상으로 지식베이스와 온톨로지를 구축한다. 이러한 연구는 다년간 진행된 연구의 틀을 마련함과 동시에 질의응답시스템, 지능형 시스템 등에 광범위하게 활용될 수 있는 지식베이스, 온톨로지의 기반을 다지는 작업이라 할 수 있다.

관련연구

지식베이스를 바탕으로 실세계의 지식을 체계적으로 데이터베이스화하고 이를 온톨로지의 개념구조와 연결시켜 질의응답시스템을 구축하는 것은, 추론을 근간으로 하는 시스템의 질적 향상뿐만 아니라 기존의 이용가능한 모든 자원을 결합하여 사용할 수 있게 한다.^{1,9)} 즉 어휘들의 의미적·개념적 연결 구조와 실세계 지식의 효율적·체계적 구조는 언어처리뿐만 아니라 정보검색, 기계번역 등에서도 기술적 향상에 큰 영향을 줄 수 있는 것이라 할 수 있다.

지식베이스는 정보검색이나 기계번역에서 많이 활용되고 있다.²⁾에서는 개념기반 정보검색의 성능 향상을 위해 개념 확장 알고리즘의 비교 실험과 여러 가지 지식베이스의 성능 비교 실험을 하고 있다. 지식베이스 성능 실험에 사용된 지식베이스는 문헌기반 지식베이스·시소러스기반 지식베이스·통합형 지식베이스·동어의 처리형 지식베이스로서, 이 중 특정 문헌 데이터베이스로부터 자동 구축된 의미망 구조의 문헌기반 지식베이스의 활용만으로도 높은 검색 성능을 보였다.⁴⁾에서는 의미기반 정보검색 기술 중 지식베이

스로 자동 구축된 정답문서 기반 검색 기술에서 지식베이스를 효율적으로 이용하는 방법을 제시하였다. 의미기반 정보검색에서 구축하는 지식베이스는, 명사의 의미론적 상하 관계를 중심으로 구축된 개념망과, 개념망의 개념어(concept word)가 웹상에서 어떤 주제에 대해서 사용되고 있는가에 따라 웹 문서를 분류해 놓은 정답문서집합으로 구성된다. 이 검색 기술에서 지식베이스의 개념망이 크게 확장되고 방대한 양의 정답문서를 구축된다면 높은 성능 향상이 기대된다 할 수 있다.

⁵⁾는 국내에서도 많이 알려져 있는 Mikrokosmos 온톨로지(5,000여개의 개념)를 중심으로 한 기계번역 방법을 제시하고 있으며, 여기에 사용된 온톨로지를 이용하여¹⁾에서는 온톨로지와 사실 자료(face database)로 이루어진 지식베이스를 이용한 질의응답시스템 구축의 실제와 방법을 제시하고 있다.¹⁾에서 중요한 점은 사실 자료가 추출된 정보를 적절한 온톨로지의 개념들의 인스턴스(instance)로 구현하여 연결시켜 구축된다는 점이다.²⁾ 지식베이스와 온톨로지의 유기적인 연결은 상호보완적인 역할을 담당할 뿐만 아니라 지식을 효과적으로 체계화시키고, 추론의 기능을 강화시켜 줄 수 있다는 점에서 중요하다고 할 수 있다.

이러한 논의들에서 간과할 수 있는 점이 지식 추출의 대상이라 할 수 있는데, 일반적으로 특정 분야의 말뭉치나 웹 문서를 사용하는 경우가 대부분이다.^{1,2,4,7,8)} 지식 추출의 대상은 질의응답 시스템 뿐만 아니라 지식기반 시스템을 구현할 때 중요한 역할을 담당함으로써, 어느 정도 정형화되어 있는 지식 데이터를 사용하여 지식베이스 구조의 틀을 마련하고 지식을 추출하여야 한다. 백과사전식 지식을 지닌 것으로 판단하는 웹 문서나 언어처리 효율성을 가진 말뭉치는, 특정 대상에 대한 보편적인 지식보다는 다소 개별적이고 산만한 체계성을 가지고 있으므로 지식 추출이 용이한 편이지 못하다. 그리하여 본 논문에서는 지식 추출 대상을, 다양한 분야의 보편적인 지식이 어느 정도 체계적으로 정리되어 있는 백과사전으로 하여 지식베이스와 온톨로지를 구축하는 방안을 모색한다.

백과사전의 구성과 '인물' 범주의 기술 방식

1. 백과사전의 구성

백과사전(encyclopedia)이란 학문, 예술, 사회, 경제 따위의 과학과 자연 및 인간의 활동에 관련된 모든 지식을 압축

2 ¹⁾에서는 개념(concept)과 인스턴스(instance)를 구분하여, '개념'이란 실제의 대상을 추상화하여 설정된 것인 반면 '인스턴스'는 이 개념의 현실세계에서 나타나는 실제의 대상이 된다고 정의하고 있다.

하여 부문별 또는 자모순으로 배열하고 풀이한 책을 말한다. 이것은 국어를 모아 일정한 순서로 배열하여 의미, 주석, 어원, 품사 따위를 밝히고 정리한 국어사전(dictionary)과는 구분된다. 지식은 사물에 대한 개개의 단편적인 실제적·경험적 인식·원리적·통일적으로 조직되어 객관적 타당성을 요구할 수 있는 판단의 체계라 할 수 있다. 이러한 지식을 일정한 기준에 의해 집대성해 놓은 일반적인 데이터 또는 언어자원을 백과사전으로 판단할 수 있다. 이러한 백과사전이 담고 있는 지식 정보를 추출하기 위해서는 백과사전의 일반적인 서술체계와 데이터 특성을 파악해야 하는데, <Fig. 1>는 백과사전의 일반적인 정보를 나타낸 것이다.

백과사전의 일반적인 서술체계는 백과사전 집필자가 필수적으로 담아야 하는 정보들이며, 백과사전의 데이터 특성은 웹이나 문서로 제공되는 백과사전의 일반적인 표현들이다. 본 논문에서 대상으로 하는 백과사전의 구성도 <Fig. 1>의 특성들을 가지고 있으므로, 백과사전의 지식 정보의 표현적 특성을 파악하는 것은 중요하다 할 수 있다.

2 백과사전 '인물' 범주의 기술 방식

본 논문에서 지식베이스와 온톨로지 구축을 위한 지식 추출 대상으로 하는 백과사전의 '인물' 범주는 최상위 13개 범주의 하나로서, 총 표제어 항목수는 31,086개로 구성되어 있으며, 백과사전 전체 표제어 항목의 약 10%를 차지하고, '인물' 범주에서 표제어 항목의 중복 데이터를 제거하면 24,804개 표제어 항목을 가진다. 또한 세부적으로 '인물' 범주는 25개의 하위 범주와 총 197개의 범주노드를 가지고 있는데, '인물'이라는 범주의 특성상 특정 도메인의 제약을 가지고 있지 않다(Table 1).

백과사전의 '인물' 범주의 지식베이스와 온톨로지를 구축하기 전에 선행되어야 하는 것이, 바로 '인물' 범주의 문서

기술 방식을 통한 지식 정보의 추출 가능성을 모색하는 것이다. <Fig. 2>는 '인물' 범주의 기술 방식을 정리한 것이다.

이를 통해 '인물' 범주에서 지식 정보를 자동으로 추출할 수 있는 지식과 반자동으로 추출할 수 있는 지식을 구분할 수 있는데, 백과사전 지식베이스(이하 '백과사전 KB'라 함) 구조를 형성하는 인물의 공통 속성(common property)과 개별 속성(individual property) 설정과 지식 정보 추출의 단서 어휘(clue word) 설정에 중요한 역할을 담당한다.

백과사전 KB 구조와 온톨로지 구축 방안

1. 백과사전 KB 구조 설계

본 논문에서의 백과사전 KB 구조는 1) 백과사전의 범주 체계, 2) 백과사전 표제어를 온톨로지와 백과사전 범주 체계를 유기적으로 연결시킨 표제어 정보, 3) 백과사전 텍스트로부터 범주별로 추출한 공통적·개별적 속성과 속성값(property_value)을 가진 백과사전 템플릿(template)으로 이루어진 구조를 말한다. 이러한 구조는 백과사전의 전체 내용을 효율적으로 지식베이스화하기 위한 구조라고 할 수 있다.

먼저 백과사전의 KB 구조는 백과사전에서 제시하는 일반적인 범주 체계를 가진다. 백과사전의 범주 체계는 계층적인 구조를 형성하면서 표제어 항목을 분류하는 중요한 정보임으로, 온톨로지 계층 정보와는 별개로 관리되어 표제어와 유기적인 연결 구조를 가져야 한다.

다음으로 백과사전 KB 구조는 백과사전의 표제어들과 온톨로지의 어휘(개념노드)와 유기적으로 연결된 구조를 가져야 한다. 백과사전의 표제어들은 온톨로지의 실세계 대상인 인스턴스(instance)가 되기 때문에 지식의 효율적인 체

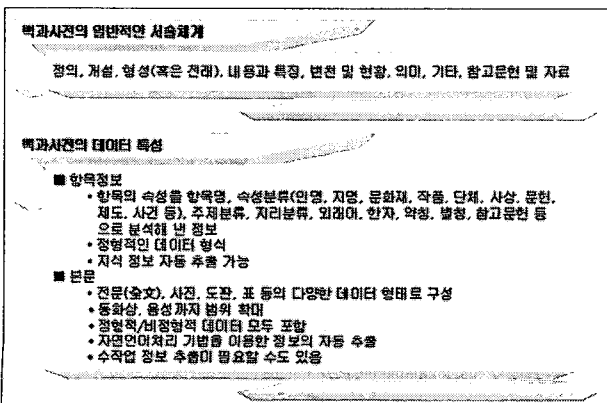


Fig. 1. 백과사전의 일반적인 서술체계와 데이터 특성.

Table 1. '인물' 범주의 25개의 하위 범주

| | |
|----|--|
| 인물 | 철학, 경제, 정치, 교육, 뉴스와 미디어, 의학, 음악, 영화 및 연극, 무용, 레저와 스포츠, 수상자, 디자인 및 패션, 종교, 사회, 법, 군사, 과학, 건축, 미술, 사진, 문학, 역사화 지리, 대중예술, 인간문화재, 기타 |
|----|--|

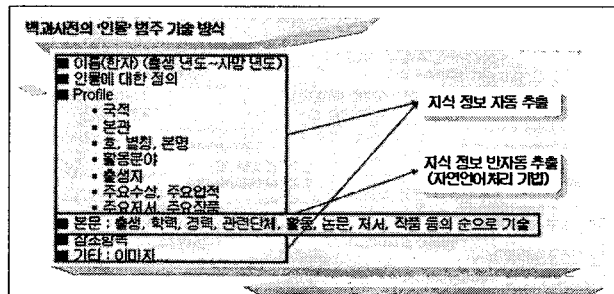


Fig. 2. 백과사전 '인물' 범주의 기술 방식.

계화와 질의응답시스템의 활용을 위해 유기적으로 연결될 필요가 있는 것이다.

마지막으로 백과사전 KB 구조는 범주별로 속성과 속성값을 가지는 백과사전 템플릿(template)을 가진다. 백과사전 KB 구조에서의 속성(property)은 특정 범주(또는 분야)의 표제어(사물)들이 가지고 있는 공통적인 성질을 가리키는데, 특정 범주에는 하위 범주들을 가지고 있기 때문에 속성의 상속(inheritance)이라는 측면에서 공통 속성과 개별 속성으로 구성된다. 즉 공통 속성은 특정 범주의 표제어들이 공통적으로 가지는 성질이며, 개별 속성은 특정 범주의 하위 범주의 표제어들이 개별적으로 가지는 성질이 된다.³

이와 같이 백과사전 KB 구조는 백과사전이 가지고 있는 지식 정보를 상호 유기적으로 연결시켜 질의응답시스템의 성능을 향상시킬 수 있는 구조를 가지고 있으며, 백과사전의 정보뿐만 아니라 온톨로지의 개념구조와 효율적으로 연결되어 있는 구조를 가진다. <Fig. 3>은 '인물' 범주를 중심으로 한 백과사전 KB 구조의 개념도를 나타낸 것이다.

2 코어 온톨로지 구축

본 논문에서의 온톨로지는 ETRI 명사 개념망을 수정하여 온톨로지화 변환하여 사용한다. ETRI 명사 개념망은 한국어 명사 어휘로 표현되는 개념을 정확하게 파악하기 위하여 개념들간의 다양한 관계를 연결시켜 놓은 어휘 데이터베이스로서, 현재 총 노드수는 47,000여개로 구성되어 있으며, 상하관계(hyponymy)를 중심으로 동의관계·유의관계·부분-전체관계 등 언어학적인 의미관계(semantic relation)를 이용하여 구축되어 있다. 이 개념망을 바탕으로

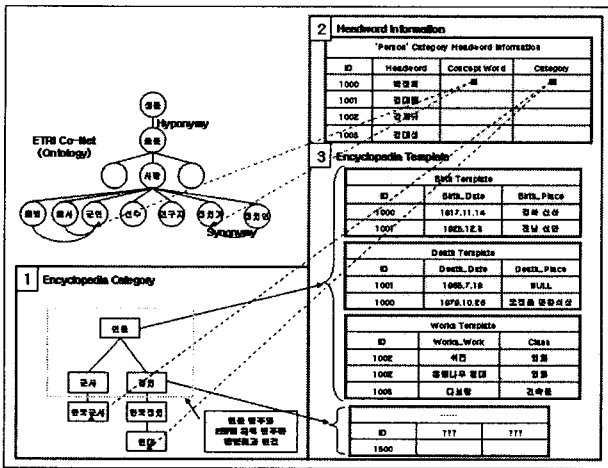


Fig. 3. 백과사전 KB 구조 개념도.

3 물론 특정 범주의 하위 범주가 또다른 하위 범주를 가질 경우에는 기존의 공통 속성과 개별 속성 모두가 또다른 하위 범주의 공통 속성이 된다.

질의응답시스템에 사용될 온톨로지를 구축하는데, ETRI 명사 개념망을 수정·추가하여 코어 온톨로지(core ontology)로 변환하고, 도메인 온톨로지(domain ontology)를 추가로 구축하여 코어 온톨로지와의 효율적으로 연결되는 구조를 가지는 것이다. 즉 도메인 온톨로지가 가지고 있지 못한 보편적 개념 체계를 코어 온톨로지가 가짐으로써 전체 도메인을 대상으로 한 질의응답시스템의 성능을 향상시킬 수 있기 때문이다.

'인물' 범주의 백과사전 KB와 온톨로지 구축

1. '인물' 범주의 백과사전 KB 구조 구축

4장에서 기술한 백과사전 KB를 바탕으로, 본 논문에서 구축 대상으로 삼은 '인물' 범주의 KB 구조 구축 방법을 제시한다. <Fig. 4>는 '인물' 범주의 백과사전 텍스트에서의 정보 추출 방법을 나타낸 것이다.

1) '범주 체계' 와 '표제어 정보' 의 자동 구축

백과사전의 '인물' 범주는 총 197개의 범주 노드를 가지고 있다. 이들을 백과사전 텍스트에서 자동으로 추출하여 계층적 연결 구조를 구축하였다. 이 '인물' 범주 체계는 온톨로지와는 별개로 관리되며 백과사전 KB에 포함되는 요소이다.

백과사전의 '인물' 범주의 표제어 개수는 24,804개이며 중복 표제어 7,002개를 포함하였을 경우에는 31,806개이며, 중복 표제어는 한 표제어가 두 개 이상의 범주에 포함된 경우를 가리킨다. 중복 데이터의 예로는, '박정희'의 경우 '군사'의 하위 범주인 '한국군사'와 '정치'의 하위 범주인 '한국정치'현대에 각각 포함되어 있다. 이러한 정보를 바탕으로 '인물' 범주의 표제어와 표제어ID를 자동으로 추출하여 '인물' 범주 체계와 자동으로 구축하였다. 표제어 정보의 온톨로지 개념노드와의 연결은 5.2절에서 다루기로 한다.

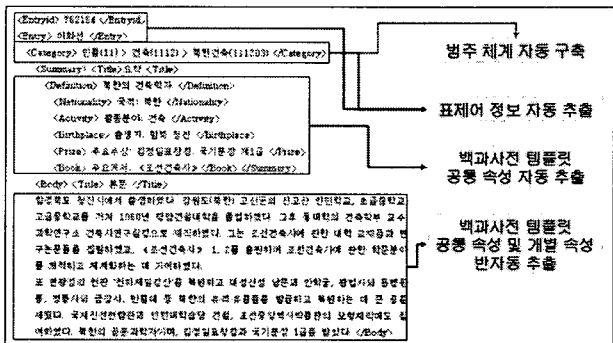


Fig. 4. '인물' 범주의 백과사전 텍스트에서의 정보 추출.

Table 3. '인물' 범주의 백과사전 템플릿 구조(공통 속성을 중심으로)

| Category | Template | Property |
|-------------------------------|---------------------|-------------------------------|
| 인물 (Person) | 출생 (Birth) | 출생장소(Birth_Place) |
| | | 출생일(Birth_Date) |
| | | 국적(Birth_Nationality) |
| | 사망 (Death) | 본관(Birth_AncestralHome) |
| | | 사망장소(Death_Place) |
| | | 사망일(Death_Date) |
| | 명칭 (Name) | 사망원인(Death_Cause) |
| | | 별칭(Name_Alias) |
| | 활동 (Activity) | 세분류(Name_Alias_Subdivision) |
| | | 활동분야(Activity_Sphere) |
| | 학력 (Schooling) | 졸업학교(Schooling_School) |
| | | 졸업일(Schooling_School_Date) |
| | | 단체(Background_Group) |
| | 경력 (Background) | 직위(Background_Group_Position) |
| | | 소속일(Background_Group_Date) |
| | 저서 (Books) | 저서명(Books_Book) |
| | | 세분류(Books_Book_Subdivision) |
| | 작품 (Works) | 작품명(Works_Work) |
| | | 세분류(Works_Work_Subdivision) |
| | 업적 (Achievement) | 주요업적(Achievement_Work) |
| 수상 (Winning a prize) | | 수상명(WinningAPrize_Prize) |
| 수상일(WinningAPrize_Prize_Date) | | |

2. '인물' 범주 관련 온톨로지 구축

'인물' 범주와 관련된 온톨로지 구축은, 앞서 4장에서 언급한 바와 같이 ETRI 명사 개념망을 수정하여 코어 온톨로지를 구축한다. 특히 '인물' 범주는 백과사전의 모든 범주를 포괄하고 있는 범주이므로, 몇몇 도메인 온톨로지는 '인물' 범주 전체를 포괄적으로 다루기 힘들다. 여기에서 중요한 점은 백과사전을 기반으로 한 온톨로지 구축에 있어서 백과사전의 표제어만이 온톨로지의 개념노드와 연결되는 것이 아니라는 점이다. 온톨로지의 개념구조는 '인물' 범주의 표제어를 인스턴스로 가지고 그것을 백과사전 KB 구조의 표제어 정보로 구축하는 것뿐만 아니라 '인물' 범주의 백과사전 텍스트에 등장하는 일반적인 기초 어휘들과 필수적인 어휘들을 온톨로지에 포함해야 한다. 즉 '인물' 범주와 관련된 온톨로지를 구축하는 것은 코어 온톨로지와 연관성이 높다고 할 수 있는 것이다.

먼저, '인물' 범주의 백과사전 텍스트에서 자동으로 추출한 온톨로지 정보는 '인물' 범주의 표제어 요약 정보에 나타나는 뜻풀이(definition)의 '인물'의 직업 관련 정보이다. 이것은 '인물' 범주에서 나타나는 '~건축학자, ~음악가' 등의 표현을 통해 80%이상 자동으로 추출할 수 있었다. 이 직

업 관련 정보는 백과사전 KB 구조의 표제어 정보와 직접적인 관련을 가지고 있기 때문에 중요한 온톨로지 정보라 할 수 있다.

다음으로, '인물' 범주의 백과사전 템플릿 구조에 사용된 공통 속성, 개별 속성, 속성값을 뽑기 위해 사용된 단서 어휘들을 온톨로지에 포함시켜야 한다. 이것은 질의응답시스템의 질의 처리에서도 필요한 어휘적 정보이며, 온톨로지에 반드시 포함되어야 하는 개념적인 어휘들이라 할 수 있다. 현재 약 10,000여개의 어휘가 ETRI 명사 개념망에 포함되어 있으며, 추가로 약 10,000여개의 어휘가 포함될 예정이다, 현재 50% 이상의 진행률을 보이고 있다.

마지막으로, 위에서 언급되었던 온톨로지 포함 어휘를 제외한 '인물' 범주 백과사전 텍스트의 본문 정보에 나타나는 어휘가 7,000여개 추가될 예정이다.

온톨로지는 완전 자동 구축이 어려운 만큼 효율적인 개념 계층 구조와 개념들간의 의미 관계를 설정하는 것은 어느 정도의 시간이 요구된다고 할 수 있으나, ETRI 명사 개념망이 수정·보완된다면 단시간 내에 많은 정보를 구축할 수 있으리라 판단된다.

결론 및 향후 연구 방향

본 논문에서는 실세계의 지식을 어느 정도 체계적·의미적으로 반영하고 있는 것을 백과사전으로 판단하여, 백과사전의 '인물' 범주(category)를 중심으로 백과사전 KB 구조의 틀을 마련하고자 하였다. 또한 어휘의 계층적 구조를 중심으로 한 온톨로지를 백과사전 지식베이스와 유기적으로 연결시킴으로써 보다 의미 있는 지식베이스를 형성하는 방안을 모색하고자 하였다. 그리하여 질의응답시스템의 성능을 향상시킬 수 있는 백과사전 KB 구조와 온톨로지의 구축과 이 둘 사이의 유기적인 연결 관계를 설정하였다.

정형화되어 있지 않은 자료에서 실세계의 지식을 효율적으로 지식베이스화하는 일은 쉬운 일이 아니며, 전문가들의 전문적인 지식을 활용한 어느 정도의 수작업이 요구된다. 본 논문에서 사용된 백과사전 역시 정형화되지 않은 정보들은 일부 수작업으로 추출하지만, 백과사전이라는 특성상 상당 부분 정형화된 형식을 취하고 있어 지식 정보의 자동 추출과 반자동 추출을 적절히 이용할 수 있었다.

향후 개체명 인식 방법, 백과사전 범주 체계와 온톨로지를 이용한 지식 정보 추출 방법, 백과사전에 나타난 고유명사의 효율적인 처리 방안 등을 계속적으로 연구한다면 계속적으로 좋은 결과가 나올 수 있을 것이라 예상된다.

REFERENCES

- 1) 신호필 (2002) : “지식기반(Knowledge-based) 질의응답시스템 : 사실 자료(Fact Database) 구축을 중심으로”, 인지과학 제 13 권 제 1 호, 한국인지과학회
- 2) 노영희 · 정영미 (2000) : “의미망 지식베이스를 이용한 개념기반 정보검색기법의 성능 연구”, 정보관리학회지 제 17 권 제 3 호, 한국정보관리학회
- 3) 김창화 · 백두권 (1989) : “지식베이스 구축을 위한 지식정의 언어와 지식생성”, 한국경영과학회지 제 14 권 제 2 호, 한국경영과학회
- 4) 장명길 외 (2001) : “의미기반 정보검색”, 정보과학회지 제 19 권 제 10 호
- 5) Mahesh Kavi (1996) : *Ontology Development for Machine Translation : Ideology and Methodology*, Technical Report MCCS-96-292, Computing Research Laboratory, New Mexico State University
- 6) Mike Uschold, Micheal Gruninger (1996) : *ONTOLOGIES : Principles, Methods and Applications*, Knowledge Engineering Review, Volume 11 Number 2
- 7) Atsushi Fujii, Tetsuya Ishikawa (2001) : “Question Answering Using Encyclopedic Knowledge Generated from the Web”, Proceedings of the ACL-EACL 2001 Workshop on Open-Domain Question Answering
- 8) Edward Loper (2000) : *Applying semantic relation extraction to information retrieval*, Master's thesis, Massachusetts Institute of Technology
- 9) John Prager, Jennifer Chu-Carroll (2001) : “Use of WordNet Hypernyms for Answering What-Is Questions”, TREC 10