

Event 온톨로지 기반의 의미 정보 검색

한용진⁰¹, 박세영¹, 이영화¹, 김권양²

¹경북대학교 컴퓨터 공학과

{yjhan,sypark,yhlee}@sejong.knu.ac.kr

²경일대학교 컴퓨터 공학과

kykim@kiu.ac.kr

Semantic Search based on Event Ontology

Yong jin Han⁰¹, Se Young Park¹, Young-Hwa Lee¹, Kwon-Yang Kim²

¹Dept. of Computer Engineering, Kyungpook National University

²Dept. of Computer Engineering, Kyungil University

인물 정보에 대한 관심이 높아짐에 따라 검색 포털에서 인물 정보 검색에 대한 메뉴를 따로 구성하는 예가 많아지고 있다. 이러한 포털의 움직임은 인터넷 사용자들이 검색을 요구하는 정보의 종류에 따라 특별한 출력을 원하는 추세이기 때문이다 한 개인에 대한 인물 정보는 데이터베이스처럼 잘 정리된 형태로 집약되어 나타나기도 하지만 여러 정보 자원을 통해 서로 다른 내용의 정보가 흩어져서 나타나는 것이 일반적이다 이러한 인물 정보는 자주 업데이트된다 따라서, 사용자가 필요한 정보를 찾고 필요에 따라 정보를 재구성하기까지 많은 시간과 노력이 소요된다 정보 검색 시스템은 정확한 정보를 찾는 것은 물론, 사용자의 요구에 맞게 재구성된 정보를 제공할 필요가 있다

현재, 상용화된 검색 시스템들은 문서와 질의를 키워드 기반으로 처리하고 있다 질의 키워드가 출현하는 빈도수가 높은 문서를 검색하기 때문에 관련된 문서를 다수 찾아 주고 있지만 검색된 문서가 실제로 사용자의 요구에 맞는지는 보장할 수 없다 예를 들어, "이재용의 이력사항은?"이라고 질의하면, '이재용', '이력사항'이 포함된 문서를 검색하지만 이력사항과 관련된 '졸업', '경력', '인적사항' 등의 내용은 포함하지 않을 수도 있다 '이력사항'을 의미적으로 해석할 수 없기 때문이다

본 논문에서는 IT-People Event Ontology(ITPEO)¹⁾의 스키마를 이용해서 서로 다른 소스로부터 추출된 인물정보를 하나의 온톨로지로 인스턴스화 한다 의미 정보 검색 시스템은 온톨로지에 있는 정보를 검색 대상으로 하며 사용자의 요구에 맞게 재구성된 결과를 보여 준다

인물 정보는 시간에 따라 변화는 정보와 변하지 않는 정보로 나눌 수 있다 생일, 출생지, 성별은 변하지 않는 정보로 이름이 같은 사람을 구분할 수 있는 근거가 된다 직위, 소속과 같이 과거부터 현재까지

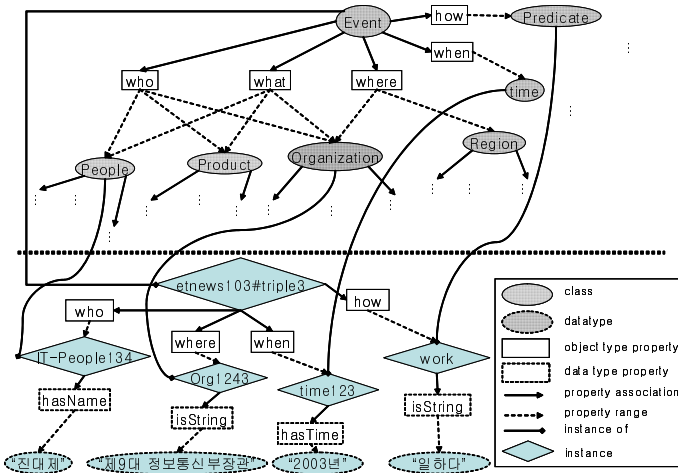


그림 1 IT-People 온톨로지의 인스턴스 예

변해 온 사건 정보는 그 자체로 개인에 대한 이력 정보를 제공해 준다 따라서, 정보를 최신화하는 것도 중요하지만 과거 정보를 유지할 필요도 있다 ITPEO는 시간 속성을 정의함으로써 온톨로지의 스키마를 유지하면서 새로운 사건 정보를 추가할 수 있도록 설계되었다. 그림 1은 ITPEO의 스키마와 Event의 인스턴스 예를 보여주고 있다 '진대제'라는 인물에 대한 직위 정보는 '2003년'이라는 시간 요소가 정의됨으로써 다른 시기에 인물의 직위 정보와 구분될 수 있다 온톨로지를 이용한 의미정보 검색에 관한

1) ITPEO는 정통부 및 정보통신연구진흥원의 정보통신선도기반기술개발사업의 연구결과로 구축되었습니다

기존 연구들이 있다. Wallace [1]는 온톨로지에 정의된 개념을 이용해 질의를 확장하는 방법을 제안했다. 질의어의 각 단어를 개념적으로 해석하지만 각 개념들 사이에 관계를 독립적인 것으로 보기 때문에 단어 사이의 의미 관계를 반영하지는 못한다. 질의 단어들 간의 관계를 통해 질의어에 대한 가중치를 반영한 연구도 있다 [2]. 하지만, 이러한 연구는 검색 대상을 문서로 하기 때문에 결국사용자는 필요한 정보를 직접 재구성해야한다. SEAL[3]은 온톨로지 자체를 검색 대상으로 하는 검색 서비스를 제공한다. 질의어는 온톨로지에 있는 개념 관계에 따라 해석되고 사용자 요구에 맞는 정보를 제공한다. 하지만, 인물에 대해 변하는 정보를 활용하고 소스 정보를 재구성하는 검색 모델을 제시한 예는 없다.

본 논문에서 제안하는 의미정보 검색 시스템은 사용자의 요구에 맞는 정보를 ITPEO에서 검색하고, 그 결과를 재구성해서 보여준다. 그림 2는 온톨로지에 기반한 의미 정보 검색 시스템의 구성을 보여 주고 있다. 데이터 베이스, 웹페이지 혹은 사람의 지식을 온톨로지화 하고 구축된 온톨로지 자체를 검색 대상으로 한다. ITPEO는 사람의 지식을 직접 입력한 정보와 비구조 문서로부터 패턴에 근거해 추출된 정보를 인스턴스화 한다. 본 논문에서는 HTML 테이블 형태의 인물 정보를 추출해서 기존의 ITPEO에 대량으로 인스턴스화 하였다. 현재, 자동 추출 방법은 연구 중에 있다[4].

제안된 시스템은 온톨로지에 있는 정보 자체를 검색할 뿐만 아니라, 기존의 소스 정보에서 얻을 수 없었던 추론된 정보 또한 사용자에게 제공할 수 있다. 예를 들어, “삼성전자 이재용의 이력사항은?”이라고 질의하면, 먼저 삼성전자 소속의 이재용을 찾고, 해당하는 이재용이라는 인물에 대한 이력사항을 검색하게 된다.

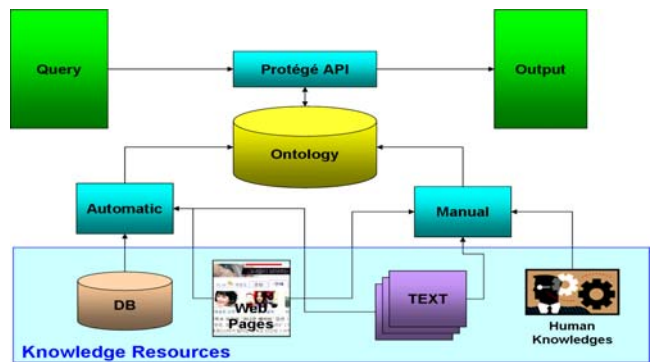


그림 2 온톨로지 기반 의미 정보 검색 시스템

키워드 기반 검색은 질의어를 포함하는 관련 문서를 검색해서 보여주지만 자연어 질의의 요구를 정확하게 반영하지는 못한다. 데이터베이스를 기반으로 한 검색은 신뢰할 수 있는 정보를 제공해 주지만 실시간으로 방대한 정보가 쏟아져 나오는 웹상의 정보를 반영해 주지는 못한다.

온톨로지는 정보를 개념과 개념 간의 관계로 표현한다. 따라서, 질의어를 의미적으로 처리함으로써 질의 요구를 반영한 검색을 가능하게 한다. 웹 혹은 데이터베이스의 다양한 정보를 자동으로 온톨로지화함으로써 인물과 관련된 변화하는 사건 정보도 개념 단위로 반영할 수 있다.

또한, 개념 단위의 정보들은 각각을 하나의 객체로 다룰 수 있기 때문에 다양한 형태의 표현이 가능하다. 온톨로지서 얻은 사건 정보를 자연어 형태로 생성하거나 인물과 연관된 정보를 그래프 형태로 표현함으로써 사용자가 접근하기 편한 형태로 결과를 보일 수 있다.

참고 문헌

- [1]Wallace A. P., and Ana M. C., "An Ontology Based-Approach for Semantic Search in Portals," *In Proceedings of the 15th International Workshop on Database and Expert Systems Applications*, pp127-131, 2004.
- [2]Jose A. R., Eduardo M., Jorge B., and Arantza I., "Seraching the Web: From Keywords to Semantic Queries," *In Proceedings of the Tird International Conference on Information Technology and Applications*, pp244-249, 2005
- [3]Maedche A., Staab S., Stojanovic N., Studer R., Sure Y., "SEmantic portAL: The SEAL Approach", *In Fensel D., Hendler J. A., Lieberman, H., Wahlster W. (eds.), Spinning the Semantic Web. MIT Press*, pp317-359, 2003.
- [4]경북대학교, 온톨로지 검증 및 온톨로지 기반 인스턴스 생성에 관한 연구최종 보고서, pp52-65, 2006.