

의미 추상화와 키워드 검색 기반의 시맨틱 네트워크 연관 검색 시스템

오근현⁰ 김용준 조성배
연세대학교 컴퓨터과학과

{ocworld, yjkim}@sclab.yonsei.ac.kr, sbcho@cs.yonsei.ac.kr

An Associative Search System for Semantic Networks based on Keyword Search and Semantic Abstraction A Mobile

Keunhyun Oh⁰ Yongjun Kim Sung-Bae Cho
Dept. of Computer Science, Yonsei University

1. 서 론

데이터의 중요성이 부각됨에 따라 어떤 데이터를 어떻게 획득할 수 있는지에 대한 문제와 더불어 데이터를 어떤 구조로 저장할 것이며 어떻게 저장된 데이터를 검색할 것인지에 대한 방법에 대한 연구가 이루어지고 있다. 효과적인 데이터의 저장과 검색을 위한 방식 중 하나로 시맨틱 네트워크(Semantic Network)로 데이터를 저장하고 연관검색을 통하여 데이터를 찾아내는 방법이 있다[1]. 시맨틱 네트워크 상의 노드가 많아지고 그 관계들이 복잡해지면 시각적으로 원하는 정보를 찾기가 어려워진다. 따라서 의미 추상화를 통하여 데이터를 손실시키지 않으면서 시맨틱 네트워크를 간략화하는 과정이 필요하다. 또한 연관검색을 적용시키기 위해 하나의 데이터를 선택하면 그에 연관된 데이터를 보여주는 선택 연관검색을 적용시킬 수 있다[2]. 그리고 선택 연관검색에서 발생하는 불필요한 검색 과정을 개선하기 위해 정확히 알고 있는 키워드를 입력하면 그 키워드를 중심으로 연관된 데이터와 그 사이의 관계를 보여주는 키워드 연관검색이 필요하다. 본 논문에서는 시맨틱 네트워크의 시각화 기반의 검색 시스템을 제안한다. 복잡한 네트워크를 간략하게 만들기 위해 [3]에서 제안된 의미 추상화(semantic abstraction)를 시맨틱 네트워크에 적합하게 적용하였고 적절하게 검색을 수행할 수 있도록 선택 연관검색과 키워드 연관검색을 제안한다.

2. 본 론

제안하는 검색 시스템은 시맨틱 네트워크 데이터를 이용하여 이를 시각화하고 그 구조와 어떤 타입의 노드와 어떤 관계가 중요한지를 보여주는 온톨로지 그래프를 구성하고 시각화한다. 그리고 이 시각화된 시맨틱 네트워크는 온톨로지 그래프 상에서 선택된 타입의 노드들과 그 노드들 사이의 관계들을 통해서만 보여지는 유도된 그래프를 구성함으로써 간략화 될 수 있다. 시각화된 시맨틱 네트워크 또는 간략화된 유도된 그래프 상에서 선택 연관검색과 키워드 연관검색을 통해 데이터를 검색할 수 있다.

의미 추상화는 데이터를 손실하지 않고 네트워크를 간략화하여 표현할 수 있는 개념이다. 온톨로지 그래프를 통하여 전체적인 네트워크의 구조를 알 수 있으며 또 어떤 노드와 어떤 관계가 중요한지 파악할 수 있다. 또한 사용자가 보고 싶은 타입의 노드들과 그 사이의 관계만을 볼 수 있는 유도된 그래프를 생성한다. 따라서 본 논문에서는 온톨로지 그래프의 타입노드의 크기와 타입엣지의 두께, 그리고 색을 일반적인 시맨틱 네트워크에 적합하게 새롭게 정의하였다.

온톨로지 그래프는 $OG=(\mathcal{T}_n, \mathcal{T}_e)$ 로 구성이 된다. \mathcal{T}_n 는 온톨로지 그래프에서 노드에 해당하며 시맨틱 네트워크 상의 노드의 타입을 의미하며 노드의 슈퍼셋 (superset)이다. \mathcal{T}_e 는 온톨로지 그래프에서 엣지에 해당하고 시맨틱 네트워크 상의 엣지의 타입을 의미하며 엣지의 슈퍼셋이다[3]. 이후로는 시맨틱 네트워크상의 노드와 혼란을 피하기 위하여 \mathcal{T}_n 를 타입노드라고 명명하고 \mathcal{T}_e 를 타입엣지라고 명명하겠다. 이렇게 구성된 온톨로지 그래프 상에서 타입노드를 선택하면 시맨틱 네트워크에서 해당 타입의 노드를 삭제할 수 있다. 그리고 삭제된 타입노드를 다시 선택하면 그 타입의 노드들은 다시 나타나게 된다. 복잡한 그래프에서 검색에 불필요한 타입의 노드들을 제거하는 의미 추상화의 과정을 통하여 그래프를 간략화 시켜 유도된 그래프를 생성한다. 이 유도된 그래프를 통하여 사용자는 데이터를 분석하고 찾기가 용이해진다.

검색시스템의 검색은 연관 검색을 바탕으로 한다. [2]에서 제안된 특정 노드로부터 단계적으로 검색을 수행해나가는 방식인 선택 연관검색을 적용하였다. 관련 데이터를 찾기 위해 데이터 노드를 선택하면 엣지로 1레벨 연결된 노드들이 보여지는 방식이다. 검색시스템에서는 키워드 연관검색을

지원한다. 확실하게 알고 있는 데이터가 있으면 그 데이터를 키워드로 검색을 수행하고 그 데이터와 연관이 있는 데이터만으로 네트워크를 구성하여 사용자에게 보여준다. 이러한 과정은 선택 연관검색을 수행시 최초 하나의 노드로부터 순차적으로 검색을 수행하게 되므로 비록 확실한 키워드를 알고 있음에도 불필요한 선택 연관검색을 추가적으로 더 수행해야 하는 불편함을 줄여준다.

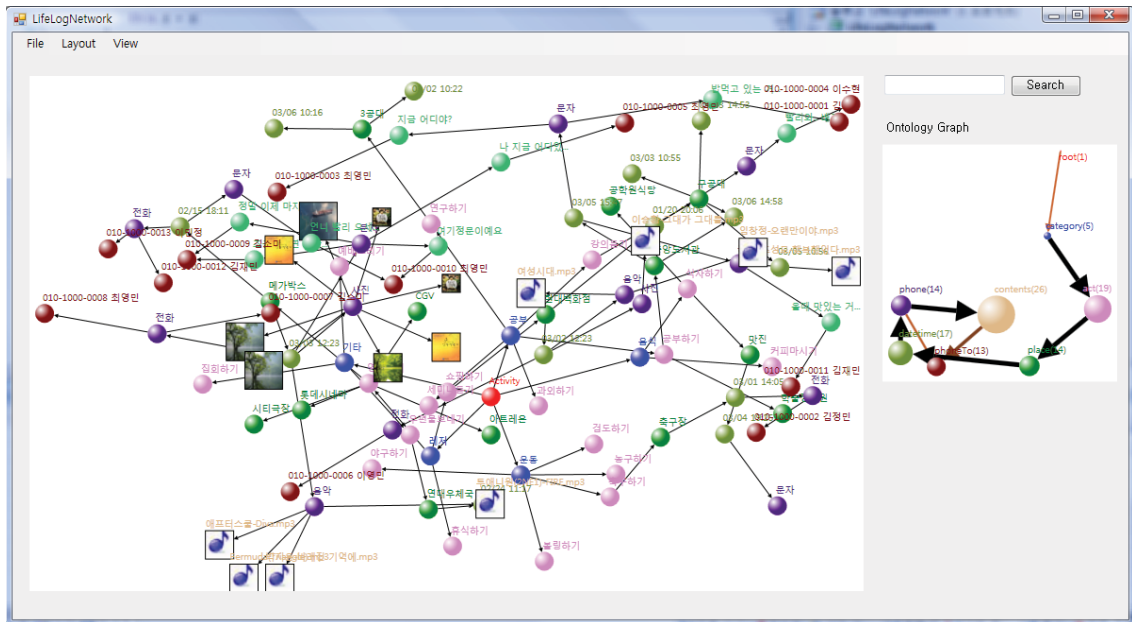


그림 1. 시각화된 모바일 라이프로그 시맨틱 네트워크

이 시스템은 닷넷프레임워크 3.5 기반의 C#으로 구현되었다. 그래프 시각화를 위해서 대량의 그래프를 시각화하기에 용이한 마이크로소프트사의 NodeXL 라이브러리를 사용하였다. [2]에서 제안된 모바일 라이프로그 시맨틱 네트워크를 사용하였다. 모바일 라이프로그 시맨틱 네트워크의 전체 시각화된 모습은 그림 1과 같다. 시맨틱 네트워크를 간략화하고 구조과약을 용이하게 하기 위해 의미 추상화를 시맨틱 네트워크에 적합하게 적용하였고 연관검색을 위해 선택 연관검색을 적용하고 불필요한 검색과정이 발생하는 것을 보완하기 위해 키워드 연관검색을 제안하였다. 유용성을 입증하기 위하여 모바일 라이프로그 시맨틱 네트워크 데이터를 이용하여 실험을 수행하였다.

3. 결론

본 논문에서는 시맨틱 네트워크 연관검색 시스템을 제안하였다. 시맨틱 네트워크를 시각화하여 데이터뿐 아니라 관계까지도 알 수 있도록 하였다. 시맨틱 네트워크를 간략화하고 구조과약을 용이하게 하기 위해 의미 추상화를 시맨틱 네트워크에 적합하게 적용하였고 연관검색을 위해 선택 연관검색을 적용하고 불필요한 검색과정이 발생하는 것을 보완하기 위해 키워드 연관검색을 제안하였다. 유용성을 입증하기 위하여 모바일 라이프로그 시맨틱 네트워크 데이터를 이용하여 실험을 수행하였다.

감사의 글

본 연구는 2009년도 정부(교육과학기술부)의 재원으로 한국연구재단-신기술융합형 성장동력사업(No. 2009-0093676)과 지식경제부 및 정보통신산업진흥원의 대학 IT연구센터 지원사업(NIPA-2009-(C1090-0902-0046))의 연구결과로 수행되었음.

4. 참고문헌

- [1] J. F. Sowa, "Semantic networks revised and extended for the second edition," *Encyclopedia of Artificial Intelligence*, 1992.
- [2] K. Oh, Y. Kim, and S.-B. Cho, "A mobile life-log search interface by associative traverse of semantic network," *Proc. of KCC*, vol. 36, no. 1(B), pp. 193-197, 2009.
- [3] Z. Shen, K.-I. Ma, and T. E.-R, "Visual analysis of large heterogeneous social networks by semantic and structural abstraction," *IEEE Trans. on Visualization and Computer Graphics*, vol. 12, no. 6, pp. 1427-1439, 2006.