

개인화 된 웹 네비게이션을 위한 온톨로지 기반 추천 에이전트

정현섭⁰ 양재영 최중민
한양대학교 컴퓨터공학과
{hsjung, jyyang, jmchoi}@cse.hanyang.ac.kr

An Ontology-based Recommendation Agent for Personalized Web Navigation

Hyunsup Jung⁰ Jaeyoung Yang Joongmin Choi
Dept. of Computer Science and Engineering, Hanyang University

요 약

본 논문에서는 온톨로지를 이용하여 웹 문서의 분류와 사용자의 정보 요구에 대한 개인화 된 정보를 제공하는 에이전트를 제안한다. 에이전트는 웹 문서들이 가지는 의미 구조를 표현한 개념 계층 즉, 온톨로지를 바탕으로 웹 문서를 분류하게 되며 온톨로지를 이용하여 사용자의 정보 요구를 정확히 파악하고 사용자의 브라우저를 돕게 된다. 온톨로지는 개념에 대한 특징, 개념간의 관계 그리고 문서 분류를 위한 제약조건으로 이루어진다. 사용자의 현 위치에서의 선행 탐색을 통하여 문서를 획득하게 되며 구축된 온톨로지를 이용하여 분류한다. 에이전트는 분류된 문서에 대한 사용자의 관심분야를 파악하여 프로파일을 유지하게 되며 최종 문서의 추천은 프로파일을 바탕으로 이루어지게 된다.

1. 서론

급속도로 발전하는 인터넷이 가져온 정보의 과부하(information overload)로 인해 사용자는 양적 풍부함을 누리고 있지만 원하는 정보를 찾는 데 있어서 많은 노력과 시간을 대가로 치러야만 하는 등의 불편함을 감수 해야만 한다. 특히 웹 사이트에 대한 특별한 사전 지식이 없는 사용자는 원하는 정보를 찾는 데 상당한 어려움을 겪게 되었으며 이런 여러 문제들을 해결하기 위한 방법들이 관심사로 대두 되었다. 하지만 고안된 대부분의 방법들은 웹 문서들이 가지는 의미론적 관계를 고려하지 않고 단순히 단어에 기반하는 방법을 사용하였다.

따라서 본 연구에서는 이러한 문제를 온톨로지를 이용한 지식 기반 시스템을 이용하여 해결하고자 한다. 뉴스 사이트에 대해 구축된 온톨로지를 이용하여 웹 문서(기사)들이 가지는 의미론적 내용과 관계를 식별할 수 있고 브라우징 시 사용자의 정보 요구를 좀더 정확하게 파악할 수 있다. 그리고 이를 통해 유도된 사용자 프로파일(profile)을 유지함으로써 좀더 사용자의 요구를 잘 반영한 개인화 된 뉴스 기사를 제공하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서는 관련연구를 통하여 여러 개인화 된 정보 추천 기법을 알아보고 3장에서는 본 논문에서의 웹 문서 분류 기법을 소개하며 4장에서는 본 논문에서 제안하는 에이전트

의 전반적인 구조를 그리고 5장에서는 이를 이용한 정보 추천 방법에 대해 알아보고 마지막으로 6장에는 결론 및 향후 연구 방향을 제시한다.

2. 관련연구

오래전부터 사용자의 브라우저를 도와주는 방법에 대한 연구가 있어왔다. 다중의 사용자로부터의 경험을 통해 학습과 강화 학습(reinforcement learning)을 통한 정보 추천 능력을 향상시키는 WebWatcher[1]와 사용자의 관심 사항을 기록한 프로파일을 여러 기계학습 방법을 통해 학습하여 웹 문서에 대한 흥미 여부를 식별하는 Syskill & Webert[2] 라는 시스템, 그리고 사용자의 행동을 추적하고 자발적으로 링크를 선행 탐색하여 사용자의 다음 브라우징 행위를 추측, 흥미가 있을 만한 아이템들을 추천하는 인터페이스 에이전트인 Letizia[3]가 있다. 그리고 WebMate [4]는 코퍼스(corpus)와 키워드 확장 방법을 이용하여 프로파일을 확장하여 개인화 된 정보를 제공한다. 또한 [5, 6]에서는 좀더 효과적인 개인화 된 정보 제공을 위해 웹 사용 마이닝(usage mining) 기법을 이용 하였으며 이 결과로 생기는 사용 프로파일과 사용자의 정보 요구 상향을 기반으로 웹 문서를 추천한다.

하지만 대부분의 시스템은 웹 문서가 가지는 의미 구조를 고려하지 않았으며 사용자의 정보 요구가 고정적이라는 것을 가정으로 하고 있다. 또한 사용자의 명시적인 관심여

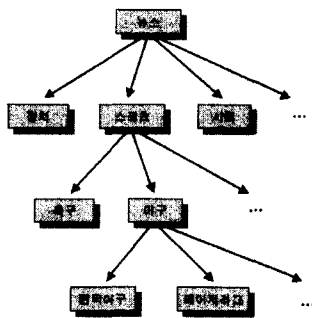
부의 입력이 필요하므로 사용자에게 부담으로 작용하게 되며 극히 주관적이라 할 수 있다. 따라서 명시적 입력이 없는 경우 시스템의 성능이 향상되지 않는다. 또한 웹 사용 마이닝 기법은 웹 사용 데이터로부터 필요한 정확한 정보의 추출이 어렵다는 것과 빈번히 변화하는 사이트에서는 효과를 발휘하지 못한다는 단점을 가진다.

따라서 본 논문에서는 웹 문서들 사이의 의미구조가 명확하고 웹 사이트의 변화가 빈번히 발생하는 뉴스 사이트를 대상으로 개인화 된 정보를 제공하여 사용자의 브라우저를 돕는 에이전트를 제안한다.

3. 웹 문서 분류

온톨로지란 기계판독이 가능한 용어(term)와 용어들 사이의 개념화(conceptualization)된 명세(specification)나 혹은 온톨로지와 논리(logic)의 관계를 의미하는 등 다양하고 폭넓은 의미를 지니고 있다[9]. 하지만 본 논문에서는 온톨로지를 " 흥미있는 도메인(domain)에 대한 공유된 개념화의 기술(description)" 이라는 의미로 정의한다. 이런 의미에서의 온톨로지는 도메인에 관련된 개념(concept)과 개념간의 관계(relation) 그리고 개념과 관계에 대한 공리(axiom)로 기술될 수 있다.

웹 문서를 분류하는 과정은 개념 계층상의 노드에 맞게 문서에 레이블링(labeling)하는 과정을 말한다. 본 논문에서는 뉴스 사이트에서의 문서를 대상으로 하고있으며 계층적 개념 구조는 [그림 1]과 같이 표현될 수 있다.



[그림 1] 뉴스 사이트에서의 계층적 개념 구조

따라서 이 개념 구조에 맞게 온톨로지를 구성하게 되며 각 노드는 개념명과 특징(feature), 관계(relation) 그리고 제약조건(constraint)로 이루어진다. 온톨로지는 XML로 표현되며 특징은 개념을 표현할 수 있는 단어나 구의 집합으로 이루어지고 관계는 노드간의 관계를 표현하며 {isA, partOf, hasPart}으로 정의된다. 그리고 제약조건은 특징들로 표현 가능한 조건들을 기술하며 {isRelatedTo, followedBy}으로 표현되며 웹 문서 분류를 위해 중요한 의미를 가진다.

모든 문서는 루트(root)노드에서부터 시작하여 가장 유사하다고 판단된 하나의 노드에 할당되게 된다. 즉, 이전 레벨의 노드가 만족되는 하위 노드들만 분류의 대상이 된다. 이 방법을 통해 적은 노드의 지식으로 개념 구조를 표현할 수 있으며 계층적인 분류를 통해 좀더 정확한 분류가 가능하게 된다. 분류를 위한 유사도 계산은 [수식 1]과 같으며 문서는 최고의 유사도를 가지는 하나의 노드에 할당되므로 한 문서는 최종 하나의 클래스(class)로 분류되게 된다.

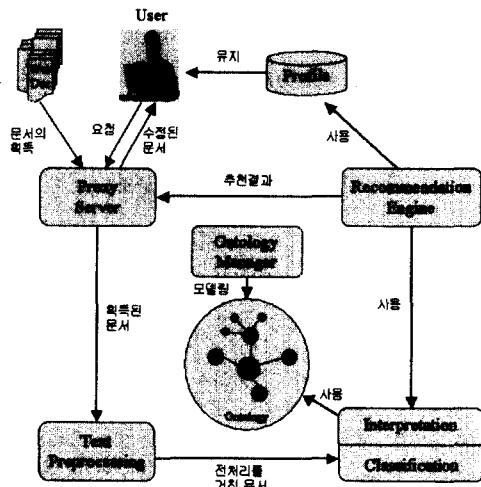
$$Sim(Node, d) = \frac{\sum_{i=0}^N freq_{i,d} / \max_{i,d}}{N} \times \frac{V_i}{V} \quad [수식 1]$$

여기서 N은 한 노드에서의 총 특징의 수이며 freq_{i,d}는 문서의 단어와 매칭되는 각 특징의 빈도수를 말하고 V는 제약조건의 수를 그리고 V_i는 문서에 의해서 만족되는 제약조건의 수를 나타낸다.

문서 분류과정에서 관계의 사용은 다음과 같다. 웹 문서 분류 결과, 문서가 어느 노드에 할당되었고 이 노드가 또 다른 어느 노드와 관련이 있을 경우 관련된 노드에 이르는 경로까지 포함해서 프로파일에 동시에 유지하게 된다.

4. 시스템 구조

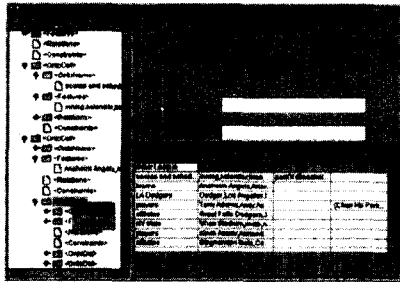
본 논문이 제안하는 온톨로지 기반의 추천 에이전트의 구조는 [그림 2]와 같다. 사용자는 일반 웹 브라우저를 통해 문서를 요구하게 되고 이 요구는 프락시(proxy) 서버를 통해 시스템에 전달되게 되며 요청한 문서는 전처리(preprocessing) 과정을 거치게 된다. 현재 온톨로지는 수동으로 구축되며 관리의 온톨로지 에디터[그림 3]를 통해서 가능하다.



[그림 2] 온톨로지 기반 추천 시스템 구조

구축된 온톨로지를 바탕으로 온톨로지 해석과 문서에 대한 분류작업을 행하게 되며 이 결과에 대한 사용자의 행위 즉, 문서의 선택과 관련한 정보를 프로파일로 유지하게 된다.

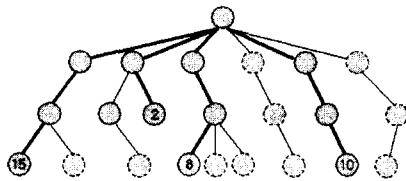
그리고 최종적으로 사용자의 관심사항이 학습된 프로파일의 내용을 바탕으로 추천엔진은 웹 문서의 추천 여부를 결정하게 되며 최종 추천결과가 추가된 수정된 웹 문서가 프락시 서버를 통해 다시 웹 브라우저로 보내지게 된다.



[그림 3] 온톨로지 에디터(Editor) 및 뷰어(Viewer)

5. 개인화 된 정보 추천

개인화 된 정보 추천을 위해 본 논문은 사용자에 대한 관심사항이 반영된 프로파일을 이용한다. 사용자에 대한 프로파일은 [그림 4]와 같이 온톨로지의 경로(path)와 해당 경로에 분류된 문서에 대한 요청 횟수의 쌍으로 이루어 진다.



[그림 4] 사용자 프로파일 구조

굵은 선으로 표현된 경로는 온톨로지 상에서 사용자의 문서 요청에 대한 경로가 존재함을 의미하며 숫자는 경로에 해당하는 문서 요청 횟수로써 요청한 문서에 대한 경로가 존재할 경우 증가하게 된다.

추천 대상이 되는 문서들은 사용자가 요청한 문서에서의 선행 탐색방법을 사용하여 획득한 문서들이 된다. 문서에 대한 추천 점수(recommendation score)는 계산 경로에 대한 사용자의 방문 횟수를 최대 방문횟수로 정규화 한 점수와 사전 탐색 시 문서의 존재 위치를 고려해서 최종 점수를 산출하게 된다. 이 경우 사용자의 현 위치에 가까이 존재하는 문서가 더 적합하다는 것을 가정하여 높은 점수를 부여한다.

마지막으로 사용자는 시스템이 관련 있다고 판단

된 문서들에 대한 링크와 관련 정도를 나타내는 점수를 추가한 수정된 웹 문서를 일반 브라우저를 통해서 최종 결과로 받게 된다.

6. 결론 및 향후연구

본 논문에서는 온톨로지를 이용하여 웹 문서들이 가지는 의미적 관계를 개념 구조로 표현하고 이를 이용하여 웹 문서를 분류하고 사용자 정보요구를 효과적으로 파악하며 그리고 이를 바탕으로 개인화 된 정보를 추천하는 에이전트를 구현하였다.

향후 연구는 크게 네 가지로 나누어 볼 수 있다. 첫째 온톨로지 구축의 자동화 혹은 반 자동화를 위하여 웹 문서에서부터 의미적 개념이나 관계를 자동으로 추출하는 방법에 대한 연구와 이에 관련된 인터페이스의 개발을 들 수 있다. 둘째로는 현재의 단일 클래스 문서 분류를 확장하여 다중 클래스 분류를 가능하게 하는 연구이다. 셋째로는 온톨로지의 기본 목적인 정보 공유의 차원에서의 협동(collaborative) 시스템의 구축을 들 수 있으며 마지막으로 온톨로지와 사용자 프로파일을 이용한 개인화 된 문서 필터링(filtering) 시스템에 대한 연구가 있을 수 있다.

7. 참고문헌

[1] H. Lieberman. Letizia: An Agent that Assists Web Browsing. In *IJCAI-95*, pp. 475-480, 1995.
 [2] Joachims, T. Freitag, D. and Mitchell, T. Webwatcher: A Tour Guide for The World Wide Web. In *IJCAI-97*, pp. 770-777, 1997.
 [3] M. Pazzani, J. Muramatsu and D. Billsus. Syskill & Webert: Identifying Interesting Web Sites. In *Proc. of 13th Natl. Conf. on Artificial Intelligence*, pp. 54-61, 1996.
 [4] Liren Chen and Katya Sycara. Webmate : A Personal Agent for Browsing and Searching. In *Proc. of 2nd Intl. Conf. on Autonomous Agents*, pp. 132-139, 1998.
 [5] B. Mobasher, R. Cooley, and J. Srivastave. Automatic Personalization Based on Web Usage Mining. Technical Report TR99010, Dept. of Computer Science, DePaul University, 1999.
 [6] B. Mobasher, H. Dai, T. Luo, Y. Sung, and J. Zhu. Integrating Web Usage and Content Mining for More Effective Personalization. In *Proc. of First Intl. Conf. on E-Commerce and Web Technologies (ECWeb2000)*, pp. 165-176, 2000.
 [7] J. Heflin, J. Hendler, and S. Luke. SHOE: A Knowledge Representation Language for Internet Applications. Technical Report CS-TR-4078 (UMIACS TR-99-71), Dept. of Computer Science, University of Maryland at College Park. 1999.
 [8] S. Luke, L. Spector, D. Rager, and J. Hendler. Ontology-based Web Agents. In *Proc. of First Intl. Conf. On Autonomous Agents*, pp. 59-66, 1997.
 [9] J. Hendler. Agents and the Semantic Web. *IEEE Intelligent Systems*, 16(2), pp. 30-37, 2001.