
WeBook: 학습자를 위한 시맨틱 지식구조 기반의 정보 검색 서비스

WeBook: A Information Retrieval Service based-on Semantic Knowledge Structures For Learners

송민아, Mina Song*, 남궁현, Hyun Namgoong**, 정진욱, Jinuk Jung***,
김흥기, Hong-Gee Kim****, 윤주현, Juhyun Eune*****

요약 본 논문에서는 사용자 참여 서비스가 발달하면서 수 많은 사용자에 의해 산발적으로 생성되는 다양한 웹 콘텐츠에서 학습자에게 의미있는 웹 콘텐츠를 제공하기 위한 검색 서비스인 위북(WeBook)을 제안한다. 위북은 도메인에 대한 배경지식이 없는 초보학습자(도메인 입문자)에게 시맨틱 지식구조 기반의 검색 방법을 제공함으로써, 일반적인 키워드 검색으로는 특정 분야에 대한 전반적인 지식의 흐름이나 지식구조를 파악하기 어려운 일반적인 키워드 검색을 보완하였다. 또한 도메인에 대한 어느 정도 지식을 가진 학습자(도메인 익숙자)에게는 관련 개념 기반의 검색 방법을 제공함으로써, 학습자의 지식수준과 상관없이 동일한 검색 방법을 이용해야 했던 일반적인 키워드 검색방법을 보완하였다.

Abstract This paper proposes a WeBook service which provides an easy access on scattered web contents. The WeBook improves traditional keyword-based information searches by providing a semantic search method based on serial knowledge structure to beginners in specific information domains. It also provides a concept-based search method for the users who already have background knowledge on the domains. We believe this service will supplement keyword-based searches by use of structured knowledge, e.g., serial knowledge structure and concepts, for searching web contents.

핵심어: *Semantic Web, Semantic Web Search, Mash-up, E-learning, Knowledge Structure, Word-network*

본 논문은 문화체육관광부 및 한국콘텐츠진흥원의 2010년도 콘텐츠산업기술지원사업인 '디지털 문화콘텐츠 융복합 서비스를 위한 시맨틱 웹 매쉬업 플랫폼 기술' 연구 과제의 지원을 받아 수행되었음.

*주저자 : 서울대학교 치과대학 석사과정 e-mail: songmina@snu.ac.kr

**공동저자 : 서울대학교 치과대학 박사과정 e-mail: nghyun@snu.ac.kr

***공동저자 : 서울대학교 치과대학 연구원 e-mail: kool.uk@gmail.com

****교신저자 : 서울대학교 치과대학 교수 e-mail: hgkim@snu.ac.kr

*****공동저자 : 서울대학교 미술대학 교수 e-mail: jheune@snu.ac.kr

1. 서론

사람들은 지식을 습득하기 위해 흔히 책과 웹 검색을 병행하곤 한다. 그렇다면, 왜 사람들은 책이 있음에도 웹 검색을 통해 추가적인 지식을 획득하려고 하며, 웹 검색을 통한 지식습득이 가능함에도 책을 필요로 하는 것일까?

책이 있음에도 웹 검색을 통해 추가적인 지식을 획득하려는 이유로는 웹 콘텐츠의 다양함을, 웹 검색을 통한 지식습득이 가능함에도 책을 필요로 하는 이유로는 키워드 기반의 일반적인 웹 검색 방법을 선택할 수 있다.

웹 검색의 경우, 웹 사용자가 일반적인 키워드 검색을 통해서 특정 분야에 대한 전반적인 지식의 흐름이나 지식구조를 파악하는 것은 쉽지 않다. 특히 배경지식이 없는 특정 도메인에 대한 입문자(초보 학습자)의 경우, 웹 검색을 통한 지식습득에 더 어려움을 느끼게 된다.

특히 사용자 참여 서비스가 발달하면서 다양한 사용자에 의해 산발적으로 생성되는 웹 콘텐츠의 양이 폭발적으로 증가하면서, 도메인 입문자가 단순 키워드 기반 검색으로 웹 리소스 내에서 해당 도메인에 대한 일련의 지식의 흐름이나 지식구조를 파악하기란 쉽지 않다.

또한 해당 분야에 대한 배경지식이 없는 도메인 입문자와 어느 정도 이상의 지식을 가지고 있는 도메인 익숙자가 해당 분야에 대한 지식의 정도가 다름에도 불구하고 같은 검색 방법을 통해 지식을 습득해야 하는 것 역시 일반적인 키워드 검색의 보완되어야 할 점이다.

따라서 본 논문에서는 산발적으로 생성되는 다양하고 풍부한 웹 리소스에서 웹 사용자가 전반적이며 체계적인 지식구조에 맞추어 유관한 지식을 보다 쉽게 획득할 수 있도록 지원하기 위해, 일반적인 키워드 검색방법을 보완한 학습자의 지식수준에 따른 시맨틱 지식구조 기반의 정보 검색 서비스인 위북(WeBook)을 제안한다.

2. 시스템 설계 및 구현

2.1 지식구조

지식구조란 지식 구성요소간의 관계를 의미하며[1], 지식구조를 생성하기 위해서 본 연구에서는 목차와 인덱스를 활용하였다. 목차란 특정분야의 전문가가 해당 분야에 대해, 해당 분야의 용어들로 생성한 일련의 지식구조이다. Seeburg 는 목차를 저자가 특정 분야에 대해 구축한 의미적으로 연결된 개괄적인 구조라 정의하였으며 [2], 사람들은 이러한 목차를 통해 개념들간의 관계 및 흐름을 파악할 수 있다고 설명한다[2][3]. 인덱스 역시 특정분야의 전문가가 해당 분야에 대해, 해당 분야의 용어들 중 중요하다고 판단되는 용어들을 모아 놓은 용어집이다.

2.2 시스템 구현

본 논문에서 제안하는 시맨틱 지식기반 정보 검색 서비스인 위북(WeBook)의 프로세스는 그림 1 과 같다

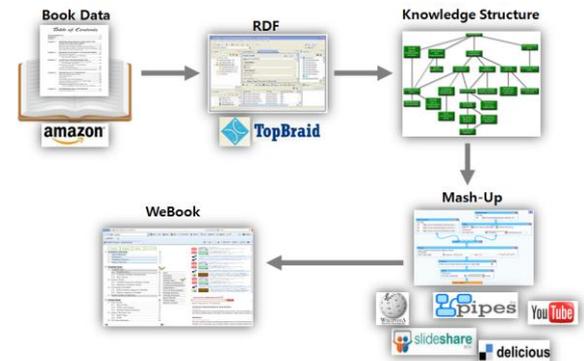


그림 1. 위북(WeBook) 프로세스

2.2.1 책 데이터 추출

본 연구를 위해 아마존 책 섹션에서 제공하는 책의 저자와 ISBN 과 같은 책의 기본적인 데이터와 지식구조를 생성하기 위한 목차와 인덱스 데이터를 추출하였다. 추출한 책 데이터는 RDF 스키마와 OWL 을 기반으로 도메인 온톨로지의 생성 및 관리를 지원하는 TopBraid 틀을 이용해 반자동으로 RDF 형식으로 변환하였다. 그림 2 와 그림 3 은 아마존에서 제공하는 목차의 일부를 RDF 형식으로 변환한 결과이다.

1	Probabilistic Informatics	3
1.1	What Is Informatics?	4
1.2	Bioinformatics	6
1.3	Probabilistic Informatics	7
1.4	Outline of This Book	8
2	Probability Basics	11
2.1	Probability Basics	11
2.1.1	Probability Spaces	12
2.1.2	Conditional Probability and Independence	14
2.1.3	Bayes' Theorem	17
2.2	Random Variables	18
2.2.1	Probability Distributions of Random Variables	18
2.2.2	Independence of Random Variables	23

그림 2. 아마존에서 제공하는 목차 일부

```
<wb:relatedIndex rdf:resource="#ISBN_0123704766_I_Ba
<wb:relatedIndex rdf:resource="#ISBN_0123704766_I_Mu
<wb:relatedIndex rdf:resource="#ISBN_0123704766_I_Th
<wb:hasDepth rdf:datatype="http://www.w3.org/2001/XM
<wb:hasStartPage rdf:datatype="http://www.w3.org/200
<wb:relatedIndex rdf:resource="#ISBN_0123704766_I_In
<wb:hasSuperChapter rdf:resource="#ISBN_0123704766_C
<wb:hasNextChapter rdf:resource="#ISBN_0123704766_C_
<rdf:type rdf:resource="webook/Chapter"/>
<dc:title rdf:datatype="http://www.w3.org/2001/XMLSc
<wb:relatedIndex rdf:resource="#ISBN_0123704766_I_La
<wb:hasConcept rdf:resource="#Conditional Independen
<wb:hasOrder rdf:datatype="http://www.w3.org/2001/XM
<wb:hasConcept rdf:resource="#Conditional Probabilit
<wb:isChapterOf rdf:resource="#ISBN_0123704766"/>
<wb:relatedIndex rdf:resource="#ISBN_0123704766_I_Mu
<wb:relatedIndex rdf:resource="#ISBN_0123704766_I_Co
<wb:hasPriorChapter rdf:resource="#ISBN_0123704766_C
<wb:hasEndPage rdf:datatype="http://www.w3.org/2001/
```

그림 3. RDF 로 변환한 목차 일부

책의 기본적인 데이터를 표현하기 위해 DC(DCMI, Dublin Core Metadata Initiative)와 SKOS(Simple Knowledge Organization System), 그리고 FOAF(Friend OF A Friend)를 상속받았으며, 목차와 인덱스 데이터와 관계 표현을 위한 일부 속성들을 생성하였다. 표 1 은 본 연구에서 제안하는 시맨틱 지식구조 기반 정보 검색 시스템인 위북(WeBook)에서 생성한 속성 일부들, 그림 4 는 위북(WeBook)시스템의 스키마를 나타낸다.

dc:title	해당 목차 원(原) 문장 (예, Conditional Probability and Independence)
wb:isChapterOf	목차가 소속된 책
wb:hasConcept	해당 목차가 가지는 컨셉 (예, 목차의 원(原) 문장이 Conditional Probability and Independence인 경우, 컨셉은 Conditional Probability와 Conditional Independence가 됨)
wb:relatedIndex	해당 목차와 연관되어 있는 인덱스
wb:hasSuperChapter	목차의 상위레벨 목차 (예, 1.1의 상위 목차는 1, 1.1.1의 상위 목차는 1.1)
wb:hasSubChapter	목차의 하위레벨 목차 (예, 1.1의 하위 목차는 1.1, 1.2, 1.3, 1.1의 하위 목차는 1.1.1, 1.1.2)
wb:hasPriorChapter	목차의 이전 목차 (예, 1.2.2의 이전 목차는 1.2.1, 1.2의 이전 목차는 1.1)

표 1. 목차 정보와 관계를 표현하기 위해 생성한 속성 일부

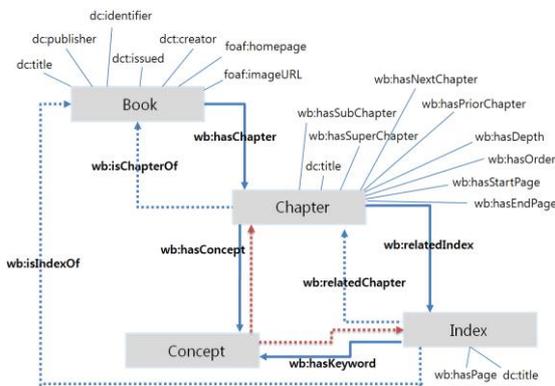


그림 4. 위북 스키마

2.2.2 매쉬업 사이트

위북 시스템은 학습자의 웹 정보 검색을 지원하기 위해 학습자자료로 활용가능한 콘텐츠들이 많이 업로드 되는 서비스 중 대표적인 사이트를 선정하여 매쉬업하여, 시맨틱 지식구조 기반 웹 검색 결과를 제공한다. 본 연구를 위해 선택한 매쉬업 대상 사이트로 슬라이드쉐어 (<http://www.slideshare.net>), 유튜브(<http://www.youtube.com>), 위키피디아(<http://www.wikipedia.org>), 딜리셔스(<http://www.delicious.com>)를 선택하였으며, 이들 사이트의 매쉬업을 위해 야후 파이프(<http://pipes.yahoo.com/pipes>)를 이용하였다.

2.3 사용자 인터페이스

위북은 도메인에 대한 배경지식이 없는 도메인 입문자에게 해당 도메인에 대한 전반적인 지식 구조를 제공하기 위한 시맨틱 지식구조 기반의 검색 화면(그림 5)과 도메인 익숙자의 검색을 돕기 위한 관련 개념간의 네트워크를 제공하는 워드 네트워크 기반의 검색 화면(그림 6)으로 구성되어 있다.

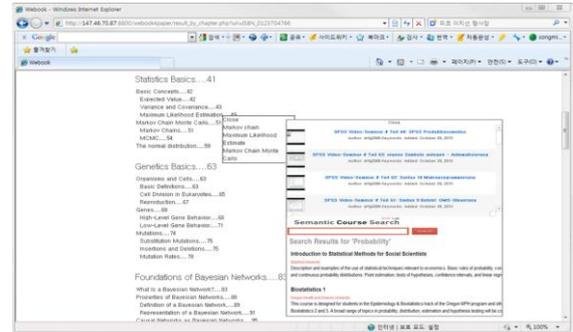


그림 5. 시맨틱 지식구조 기반 검색 화면

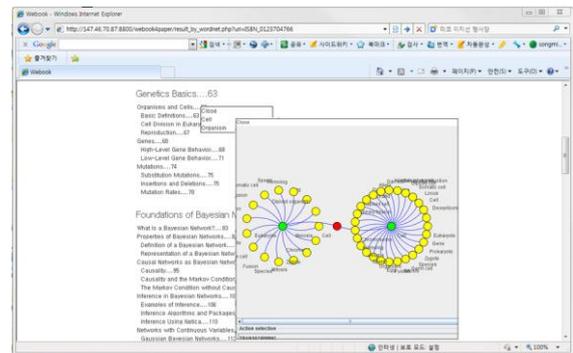


그림 6. 워드 네트워크 검색 화면

3. 결론

본 논문에서는 특정 도메인에 대한 학습자의 지식수준에 따른 검색을 지원하기 위해, 일반적인 키워드 기반의 웹 검색을 보완한 시맨틱 지식구조 기반의 검색 방법과 워드 네트워크 기반의 검색 방법을 제안하였다. 후속연구에서는 본 논문에서 제안하는 시스템에 대한 사용성 평가 및 전문가 평가가 요구된다.

참고 문헌

[1] M.K. Jones, Z. Li & M.D. Merrill, "Domain Knowledge Representation for Instructional Analysis" Educational Technology, Vol. 30, No. 10, pp. 7-32, 1990
 [2] C. Seeberg, et al. "Individual Tables of Contents in Web-based Learning Systems" Conference on Hypertext and Hypermedia, pp. 167-168, 1999
 [3] E. Dennis, et al. "SuperBook: an automatic tool for information exploration -hypertext?" Conference on Hypertext and Hypermedia, pp. 175-188, 1987