

도서 정보· 본문텍스트 통합 마이닝 기반 사용자 참여형 시각화 및 추천 큐레이션 플랫폼 개발에 관한 연구

홍민하*, 최건희**, 박경훈***, 정광철****, 김승훈*****

*단국대학교 컴퓨터학과

**단국대학교 소프트웨어학과

***㈜스타네이션

****㈜유엘씨시스템즈

***단국대학교 응용컴퓨터공학과

e-mail : mhdjae66@hanmail.net*

rjsgmlgood@naver.com**

jamespark@starnation.co.kr***

jungkc@ulcsystems.com****

edina@dankook.ac.kr*****

A Study on Platform Development for User Participatory Visualization and Recommendation Curation based on Integrated Mining of Book Details and Body Texts

Min-Ha Hong*, Gun-Hee Choi**, Kyoung-Hoon Park***,

Kwang-Chul Jung****, Seung-Hoon Kim*****

*Dept. of Computer Science, Dankook University

**Dept. of Software Science, Dankook University

***Starnation Inc.

****ULCSystems Inc.

*****Dept. of Applied Computer Engineering, Dankook University

요 약

오늘날 인터넷의 발달과 전자 책(e-Book) 시장규모가 커짐에 따라 온라인을 통한 도서 정보 제공이 증가하고 있다. 하지만 현재 도서 정보나 도서 추천을 제공하는 온라인 사이트들은 기본 서지 정보만을 위주로 제공하고 있어 도서 본문을 활용한 정보 제공 및 추천 시스템의 필요성이 증가하고 있다. 따라서 본 논문에서는 도서 본문을 활용한 정보 제공 및 개인 맞춤형 추천을 위해 ‘도서 정보· 본문텍스트 통합 마이닝 기반 사용자 참여형 시각화 및 추천 큐레이션 플랫폼’을 제안하고, 이를 구축하였다. 제안한 서비스 플랫폼은 독자에게 다양한 방법으로 도서 정보를 제공하며, 독자는 적은 시간으로 많은 정보를 얻을 수 있도록 하여 사용자의 도서 선택의 폭을 넓혀줄 것이다.

1. 서론

오늘날 인터넷의 발달과 전자 책(e-Book) 시장규모가 커짐에 따라 도서에 대한 대량의 데이터들이 쏟아져 나오고 있으며, 이에 따라 도서 정보를 제공하는 웹사이트도 많아지고 있다. ‘문화유동복스’의 2014년 상반기 출판시장 현황보고서에 의하면 인터넷서점의 유통채널 시장점유율이 31.4%로 도매서점 25%, 대형서점 17%보다 월등히 앞서 전체 유통채널에서 점유율이 가장 높은 것으로 나타났다[1]. 그러나 높은 시장점유율에 비해 현재 국내외의 인터넷서점에서는 서명, 저자명, 출판사명 등 기본적인 서지정보를 중심으로 도서 정보를 제공하고 있다. 이러한 서지정보 중심의 정보 제공 방식은 사용자가 정보를 얻고자 하는 도서가 명확하지 않을 때에는 효율적이지 못하다.

따라서 사용자의 도서 선택에 도움이 되기 위해서는 도서의 주제어, 분위기 등 서지정보 이외의 정보 제공이 필요하다. 이를 위한 방법 중 하나가 도서 본문을 활용하는 것이다. 최근 인터넷서점과 포털 서비스를 중심으로 본문검색 서비스가 도입되고 있다. ‘본문검색’ 서비스란 인터넷상에서 도서의 내용을 키워드로 검색하여 편리하게 찾아보고 사용자가 종이책이나 전자 책을 구매하는데 판단 정보로 활용할 수 있도록 개발된 것이다[2]. 세계 최대 인터넷서점인 아마존닷컴은 최초의 본문 검색 서비스인 ‘Search Inside the Book’을 개시하였다. 인터넷 포털서비스 업체인 구글은 주요 도서관의 소장 도서들과 기타 문헌들을 스캔해 구글 사이트에서 검색할 수 있게 하는 ‘Google Print’ 도서관 프로젝트를 시작하였다.

이러한 환경과 실황을 기하여 본 논문에서는 본문을 활용한 정보 제공 및 개인 맞춤형 추천을 위해 ‘도서 정보· 본문텍스트 통합 마이닝 기반 사용자 참여형 시각화 및 추천 큐레이션 플랫폼’을 제안하고 이를 구축하였다.

2. 관련연구

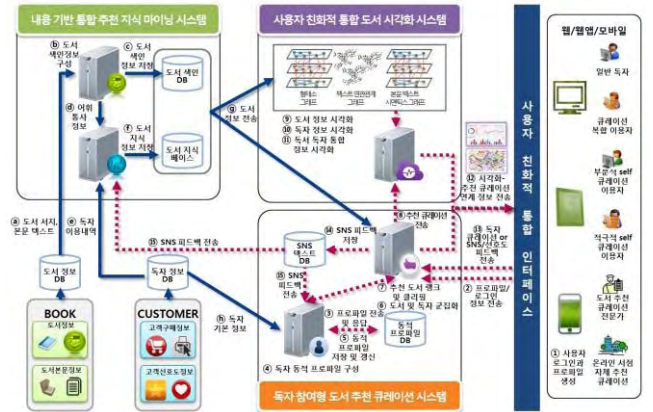
현재 제공되고 있는 도서 정보가 사용자의 도서 선택에 도움을 주기에 미흡하며 이에 따라 도서 정보를 효과적으로 제공하기 위한 연구가 활발히 진행되고 있다. [3]은 도서 정보 제공을 위한 기초 모델 구축하기 위해 효과적인 도서 현재 국내외에서 운영 중인 인터넷서점을 인터페이스, 정보원 구성, 기능적 측면을 중심으로 조사, 분석하였다. 정보원 구성측면에서 조사, 분석한 결과를 보면 대부분의 웹사이트에서는 일반적인 서지정보를 중심으로 정보를 제공하고 있어 기본 서지정보 이외에 서평, 본문 일부 발췌 등 도서에 대한 다양한 정보 제공의 필요성을 말하고 있다. 최근에는 사용자에게 제공할 수 있는 여러 정보들 중 본문을 활용한 도서 정보의 제공이 주목을 받고 있다. 실제로 도서 본문을 활용한 도서 정보 제공 서비스의 필요성에 따라 아마존의 ‘Search Inside the Book’, 구글의 ‘Google Book Search’, 네이버와 교보문고의 본문검색 등의 서비스가 도입되었다. [2]는 도서의 본문을 활용한 정보 제공이 기존에 콘텐츠명, 저자명, 출판사명 등을 통해서 찾던 콘텐츠 검색방식에 비해 자신이 원하는 정보를 훨씬 근접하게 찾아볼 수 있다는 장점이 있다고 한다.

또한 현재 제공하는 대부분의 추천 서비스는 이용 내역을 기반으로 추천이 이루어지고 있어 추천 방식 변화의 필요성에 대한 연구도 진행 중이다. [4]는 현재 인터넷서점에서 제공되는 추천 서비스는 특정 책을 구매한 사람이 구매한 또 다른 책을 추천하는 형식이 주를 이루고 있다고 한다. 때문에 [4]는 이러한 책 추천 방식이 상품 추천에 있어서 개별 사용자의 취향을 고려하기 어렵다는 한계점이 있음을 인식하고 인지적 요소, 니즈, 동기, 감성, 취향 등의 다양한 요소를 고려하여 유형을 구분하고 각 유형 별로 구매 시 중요하게 생각하는 특징 및 정보 탐색 방식을 연구하였다.

3. 도서 정보· 본문텍스트 통합 마이닝 기반 사용자 참여형 시각화 및 추천 큐레이션 플랫폼

본 절에서는 도서 정보· 본문텍스트 통합 마이닝 기반 사용자 참여형 시각화 및 추천 큐레이션 플랫폼 개발의 전체 시스템 구성에 대해 제안한다. 전체 시스템은 도서 분석을 위한 ‘내용 기반 통합 추천 지식 마이닝 시스템’, 사용자에게 정보를 효과적으로 제공하기 위한 ‘사용자 친화적 통합 도서 시각화 시스템’, 사용자에게 적합한 도서를 추천해주기 위한 ‘독자 참여형 도서 추천 큐레이션 시스템’의 3 가지로 구성되어 있다. 전처리 과정을 통해 도서 데이터를 분석하고

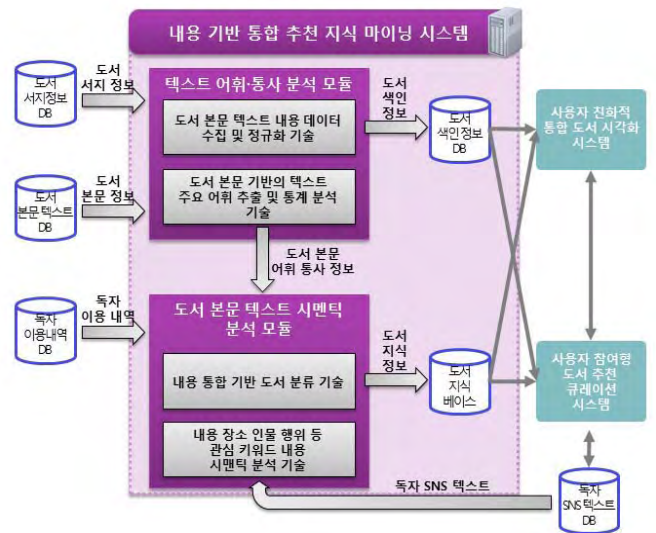
도서 선택에 있어 활용하는 정보를 기준으로 구분한 사용자 유형에 따라 개인 별 맞춤형 추천 서비스를 제공한다.



(그림 1) 전체 시스템 구성 및 서비스 흐름도

3.1 내용 기반 통합 추천 지식 마이닝 시스템

그림 2 는 제안한 내용 기반 통합 추천 마이닝 시스템이다. 본 시스템은 서비스 이용자들에게 도서를 추천하기 이전에 전처리 과정으로 도서를 분석하기 위한 시스템이다. 각각의 도서 본문에서 형태소 분석을 통하여 키워드를 추출하고 도서의 인물간의 관계, 도서 저자의 성향, 도서의 분위기 등 도서 본문에 내포되어있는 다양한 요소들을 파악하여 도서 정보 DB 에 저장한다. 또한 도서 데이터는 방대한 양의 빅데이터이므로 각 도서는 색인화하여 도서 색인 DB 에 저장한다.

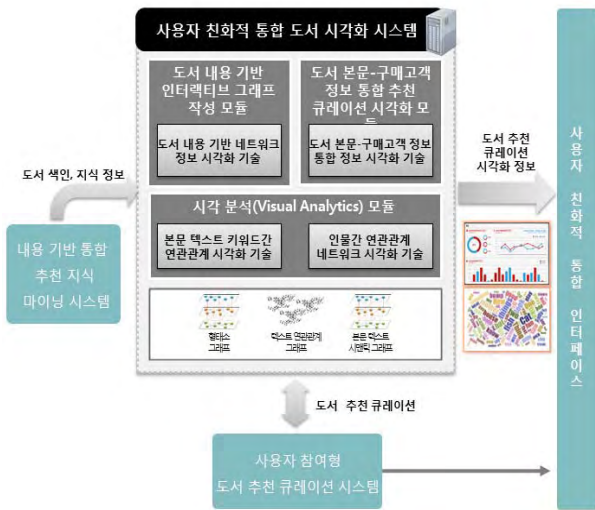


(그림 2) 마이닝 시스템 구성도

3.2 사용자 친화적 통합 도서 시각화 시스템

그림 3 은 제안한 사용자 친화적 통합 도서 시각화 시스템이다. 본 시스템은 다양하게 제공되는 도서 정보를 사용자가 한눈에 파악할 수 있도록 네트워크 형태, 그래프 형태 등 정보의 직관적 시각화

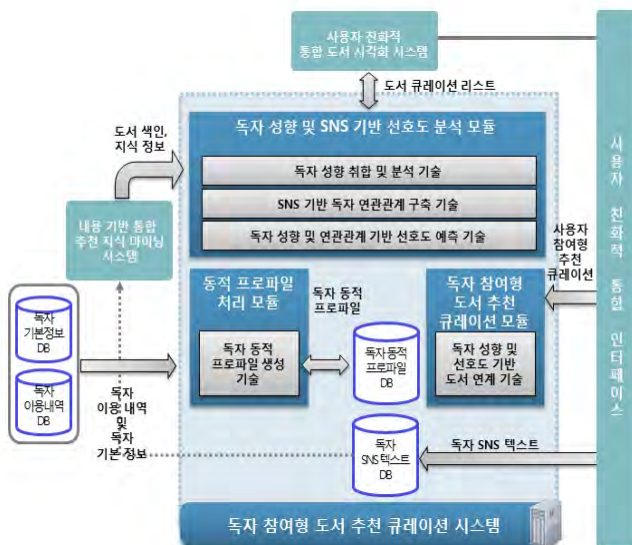
를 제공하는 역할을 한다. 시각화를 위한 데이터는 마이닝 시스템을 통해 분석된 도서에서 추출한 정보들과 본 논문의 시스템을 이용하는 사용자 정보를 토대로 한다.



(그림 3) 시각화 시스템 구성도

3.3 독자 참여형 도서 추천 큐레이션 시스템

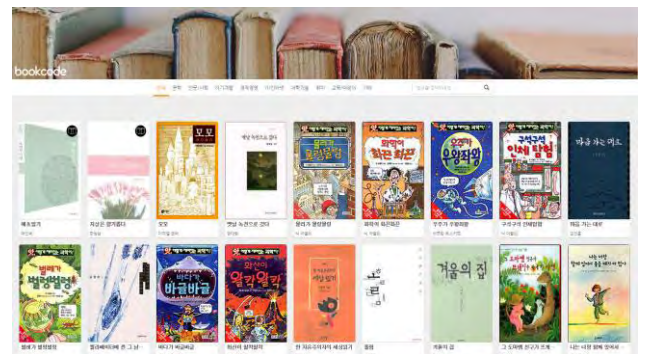
그림 4는 제안한 독자 참여형 도서 추천 큐레이션 시스템이다. 본 시스템은 수집된 정보를 이용하여 각 사용자에게 적합한 도서를 추천해 주는 것을 목표로 한다. 따라서 도서 추천에 이용되는 데이터는 사용자의 기본 정보 및 피드백 정보, 마이닝 시스템에서 분석된 도서 정보가 될 수 있다. 각 사용자들의 구매내역, 선택 도서, SNS 등을 통하여 사용자의 성향을 분석하고 사용자에게 각 성향에 따른 추천 도서를 제공한다. 기존 추천 시스템에서 제공하는 단순 사용자 기반의 추천이나 판매량 위주의 추천과는 달리 각 개인의 성향에 적합한 맞춤형 추천서비스를 제공한다.



(그림 4) 큐레이션 시스템 구성도

4. 서비스 플랫폼 구축

본 절에서는 제안한 서비스 플랫폼을 사용자에게 적용 해보고 사용자로부터 반응을 파악하기 위해 웹 서비스형태로 시범서비스(bookcode.org)를 구축하였다. 구축한 시범서비스는 약 3 만권의 본문 텍스트와 약 50 만권의 서지정보를 분석하여 장르 분석 정보, 연관 도서 추천 정보 등을 제공하고 있다. 향후에는 출판사, 작가와의 협약을 통해 도서 본문 데이터를 증가시키고 정교화된 알고리즘으로의 개선 및 개발을 통하여 양적, 질적으로 향상시킨 서비스모델을 구축해나갈 계획이다. 그림 5는 시범서비스에서 제공하고 있는 도서정보 서비스이다.



(그림 5) 시범서비스 메인 화면

시범서비스에서는 도서의 본문 텍스트를 분석하는 처리과정을 거쳐 나온 분석 결과를 기본 정보, 장르 요약 분석, 장르 상세 분석, 연관도서 추천 정보 4가지로 제공한다[5].

4.1 도서 기본 정보 제공

구축한 시범서비스는 도서 기본 정보인 도서명, 저자명, 출판사, 출판일, ISBN 등 기본 서지정보와 목차, 서평 등의 정보를 제공한다.

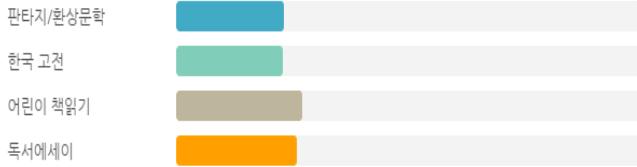


(그림 6) 도서 기본 정보

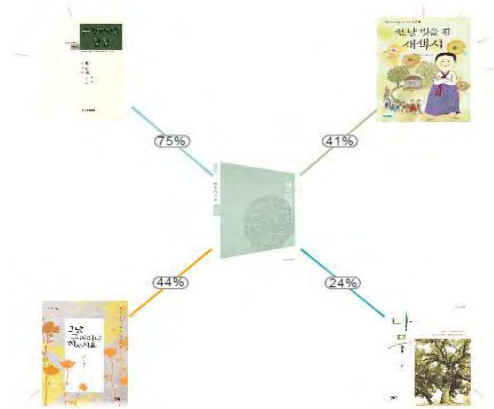
4.2 도서 장르 요약 분석 정보 제공

도서 장르 요약 분석은 도서 본문 전체를 분석한 정보를 이용하여 도서 본문에 사용된 단어의 빈도수를 기준으로 해당 도서가 어떤 장르 정보를 포함하는지를 간략하게 표현한 것이다. 도서는 어느 하나의

장르만으로 분류할 수 없다. 따라서 도서 본문에 포함되어있는 모든 장르를 파악한 후 그 중 포함률이 높은 대표 장르들을 추출하였다. 추출한 장르 정보는 장르의 포함 정도를 파악하기 쉽도록 막대 그래프를 이용하여 시각화하였다.



(그림 7) 도서 장르 요약 분석



(그림 9) 유사 도서 추천

4.3 도서 장르 상세 분석 정보 제공

도서 장르 상세 분석에서는 도서 본문의 내용을 통해 파악한 장르 정보를 Area 그래프를 이용하여 시계열적으로 표현한다. 도서 한 권을 x 개의 구간으로 나누어 각 구간에 포함된 장르와 그 포함률을 시각적으로 표현한다. 해당 정보를 통해 도서의 흐름에 따른 장르 파악과 각 장르의 구간 별 포함 정도의 변화를 알 수 있다. 따라서 사용자는 도서를 읽어보지 않은 상태에서 도서의 대략적 내용 파악이 가능하기 때문에 간접적으로 도서를 읽은 효과를 얻을 수 있다.



(그림 8) 도서 장르 상세 분석

4.4 연관 도서 추천 정보 제공

그림 9 은 유사 도서 추천 시각화를 나타낸다. 사용자가 정보를 얻고자 선택한 도서와 유사한 도서를 네트워크 형태로 시각화 한다. 도서 간 유사성을 판단하는 기준으로는 키워드, 장르, 작가 등 다양한 요소가 적용이 될 수 있으며 현재 시범서비스에는 그 첫 단계로 본문에 포함된 장르 별 키워드를 기준으로 유사성을 판단하여 적용하였다. 예를 들어, 그림 7,8 과 같이 선택된 기준 도서가 판타지/환상문학, 한국 고전 장르를 포함하고 있으면, 판타지/환상문학과 관련된 키워드를 많이 포함하고 있는 타 도서와 한국 고전 장르와 관련된 키워드를 많이 포함하고 있는 타 도서가 연관 도서로 추천된다. 연관 도서 추천 정보 제공은 사용자가 현재 선택한 도서 이외에 선호할 것으로 예측되는 타 도서를 추천해주는 선호도 예측 방법의 하나이며, 정보를 얻은 사용자는 타 도서에 대한 정보를 추가적으로 제공받음으로써 선택의 폭을 넓힐 수 있다.

4. 결론 및 향후 연구

본 논문에서는 도서 본문에 대한 필요성을 인식하고 도서 본문을 이용한 정보 제공 및 추천시스템을 구축하기 위해 ‘도서 정보· 본문텍스트 통합 마이닝 기반 사용자 참여형 시각화 및 추천 큐레이션 플랫폼 개발’이라는 주제로 연구를 진행하고 있다. 또한 연구 진행과 동시에 현재까지 진행된 연구를 시범서비스에 적용하여 구축하였다. 본 시스템은 해당 연구를 통해 판매량 위주의 기존 도서 추천 시스템을 벗어나 도서 본문을 활용한 추천시스템을 사용자에게 제공하는 것이 목적이다. 따라서 차후에는 출판사, 작가와의 협약을 통해 도서 본문 데이터를 증가시키고 정교화된 알고리즘으로의 개선 및 개발을 통하여 양적, 질적으로 향상시킨 서비스모델을 구축할 계획이다.

Acknowledge

본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2014 년도 문화기술연구개발지원사업의 연구결과로 수행되었음

참고문헌

- [1] 문화유통박스, “2014 년 상반기 출판시장 현황보고서,” 2014
- [2] 이용준, “도서본문검색 서비스의 현황과 발전방안 -네이버와 교보문고를 중심으로”, 2010, 동서언론 제 13 집
- [3] 이란주, “인터넷 서점의 효과적인 웹사이트 설계에 관한 연구”, 2001, 정보관리학회지 제 18 권 제 1 호
- [4] 이수진, 이상민, 박호열, 전해원, 박정아, “인터넷서점 이용자의 독서 유형에 따른 추천 서비스 모델”, 2012, 한국 HCI 학회 학술대회
- [5] 홍민하, 박경훈, 이원진, 김승훈, “도서 데이터와 본문 텍스트 통합 마이닝을 기반으로 한 도서 콘텐츠 장르 분석 및 시각화 시스템 구현”, 2015, 한국컴퓨터정보학회