

도서 정보 및 본문 텍스트 통합 마이닝 기반 사용자 맞춤형 도서 큐레이션 시스템

안희정* · 김기원** · 김승훈***

Personalized Book Curation System based on Integrated Mining of Book Details and Body Texts

Hee-Jeong Ahn* · Kee-Won Kim** · Seung-Hoon Kim***

Abstract

The content curation service through big data analysis is receiving great attention in various content fields, such as film, game, music, and book. This service recommends personalized contents to the corresponding user based on user's preferences. The existing book curation systems recommended books to users by using bibliographic citation, user profile or user log data. However, these systems are difficult to recommend books related to character names or spatio-temporal information in text contents. Therefore, in this paper, we suggest a personalized book curation system based on integrated mining of a book. The proposed system consists of mining system, recommendation system, and visualization system. The mining system analyzes book text, user information or profile, and SNS data. The recommendation system recommends personalized books for users based on the analysed data in the mining system. This system can recommend related books using based on book keywords even if there is no user information like new customer. The visualization system visualizes book bibliographic information, mining data such as keyword, characters, character relations, and book recommendation results. In addition, this paper also includes the design and implementation of the proposed mining and recommendation module in the system. The proposed system is expected to broaden users' selection of books and encourage balanced consumption of book contents.

Keywords : Personalized Book Curation, Big Data, Text Mining, Recommendation, SNS

Received : 2017. 03. 03. Revised : 2017. 03. 20. Final Acceptance : 2017. 03. 20.

※ This research is supported by Ministry of Culture, Sports and Tourism (MCST) and Korea Creative Content Agency (KOCCA) in the Culture Technology (CT) Research & Development Program 2016.

* Department of Computer Science, Dankook University, e-mail : dreaminghee90@gmail.com

** Assistant Professor, Department of Applied Computer Engineering, Dankook University, e-mail : nirkim@dankook.ac.kr

*** Corresponding Author, Professor, Department of Applied Computer Engineering, Dankook University, 152 Jukjeon-ro, Suji-gu, Yongin-si, Gyeonggi-do, 16890, Korea, Tel : +82-31-8005-3681, e-mail : edina@dankook.ac.kr

1. 서 론

IT 발전 및 4차 산업 혁명과 함께 데이터양이 기하급수적으로 증가하고 데이터 유형이 다양화되면서 데이터 수집, 축적, 분석으로 새로운 가치를 만들어 내는 빅데이터가 정보통신 분야 및 산업 전 분야에 새로운 이슈로 부상하고 있다. 빅데이터 분석은 영화/영상, 게임, 음악, 전자책 등 다양한 콘텐츠 분야에서도 주목받는 기술로 개인의 취향과 성향에 맞춤형된 콘텐츠를 제공하고, 사용자에게 더 많은 인기를 얻을 수 있는 매력적인 콘텐츠 기획 및 개발에 활용되고 있다[KOCCA, 2014]. 이러한 맞춤형된 콘텐츠를 추천해주는 콘텐츠 큐레이션 서비스는 도서 분야에서도 중요한 역할을 한다. 기존 도서 큐레이션 서비스는 도서 서지 정보, 도서 서평, SNS 정보 등을 기반으로 도서를 추천하며, 이 서비스는 베스트셀러 위주의 추천이 아닌 시장에 노출되기 어렵던 소외 장르 혹은 무명의 우수 콘텐츠들이 빛을 볼 수 있는 창구가 마련된다는 장점이 있다[Na et al., 2014]. 하지만 기존 큐레이션 서비스는 도서정가제로 인해 오프라인 서점 대비 가격 경쟁력이 약해진 온라인 서점의 경우 사용자의 취향이나 본문 내용 분석과 같은 고차원적이고 구매의욕을 불러일으키는 도서 분석 및 추천에는 한계가 있다. 기존 큐레이션 서비스에서 사용되는 데이터만을 이용한다면 도서 텍스트 내에서 나타나는 도서 주제어 기반의 추천 또는 동일한 시대를 기반으로 하는 도서 추천과 같은 구체적인 기준의 맞춤형 추천을 해주기는 어렵기 때문이다.

본 논문에서는 기존의 단순한 도서 서지 정보와 사용자의 선호도 및 구매 이력 기반의 추천 시스템의 한계를 개선하기 위해 도서 본문 텍스트 마이닝을 통한 내용 중심의 사용자 맞춤형 도서 큐레이션 시스템을 제안한다. 제안한 도서 큐레이션 시스템은 도서 서지 정보, SNS 데이터,

도서 본문 텍스트 등을 분석하는 마이닝 시스템과 분석된 정보를 바탕으로 사용자 맞춤형 도서를 추천해주는 추천 시스템, 마지막으로 도서 정보와 마이닝 시스템에서 추출된 분석 정보 및 추천 도서를 사용자에게 직관적으로 보여주기 위한 시각화 시스템으로 구성된다. 제안된 시스템은 추후 온라인 서점과 국/공립 도서관 및 초, 중, 고등학교 도서관 서비스에 적용하여 일반 사용자에게 차별화된 가치를 제공하는 맞춤형 도서 콘텐츠를 추천할 수 있을 것이며, 개인화된 맞춤형 도서 콘텐츠 추천을 통해 기존의 베스트셀러 중심에서 양서 중심으로 도서 선택의 폭을 넓히고 도서 콘텐츠의 균형 잡힌 소비를 유도할 수 있을 것이다.

본 논문의 구성은 다음과 같다. 제 2장은 관련 연구로 콘텐츠 큐레이션 서비스와 도서 추천 큐레이션 서비스의 특징 및 현황에 대하여 기술한다. 제 3장에서는 본 논문에서 제안하는 사용자 맞춤형 도서 추천 큐레이션 시스템에 대하여 자세하게 기술하며, 제 4장에서는 본 논문에서 제안한 시스템의 설계 및 구현을 기술한다. 마지막으로 제 5장은 결론에 대하여 기술한다.

2. 관련 연구

2.1 콘텐츠 큐레이션 서비스 특징과 현황

콘텐츠 큐레이션 서비스란 수많은 데이터 혹은 콘텐츠들 중 개인의 주관이나 관점에 따라 관련된 정보들을 수집, 정리하고 편집하여 사용자와 관련이 있거나 좋아할 만한 콘텐츠를 제공하는 서비스를 말한다[Ji, 2012]. 콘텐츠 큐레이션은 필터링 주제에 따라 '데이터 큐레이션'과 '소셜 큐레이션'으로 나뉘는데, 서비스 제공자가 정보, 콘텐츠의 속성, 사용자의 선호도 정보를 활용하여 콘텐츠 큐레이션 서비스를 제공하는 것을 '데이

터 큐레이션'이라고 한다. 반면, 소셜 큐레이션은 정보 혹은 콘텐츠의 필터링 주체가 사용자가 된다. 사용자는 다른 사람이 필터링한 콘텐츠를 구독하거나 자신의 콘텐츠를 타인과 공유할 수도 있다. 즉, 본 시스템에서 사용되는 콘텐츠 큐레이션은 데이터 큐레이션을 의미하며, 이는 특정 분야에 관심이 있는 사람들에게 해당 주제를 갖는 정보들을 수집하고 보다 편리하고 쉬운 방식으로 서비스를 제공하는 일종의 추천 서비스라고 정의할 수 있다[Ovadia, 2013]. 현재 콘텐츠 큐레이션은 다양한 분야에서 적용되고 있다. Lee and Lee[2016]는 사용자 프로파일과 피드백을 기반으로 한 맞춤형 스포츠 융합 콘텐츠 큐레이션 시스템을 제안하였다. 또한 카카오토픽은 개인화 및 소셜 필터링이 반영된 알고리즘으로 맞춤형 콘텐츠를 제공하고, 아마존의 경우 추천 글 기능, 주제별, 사용자 선호도별 등 다양한 방식으로 연결된 도서 콘텐츠 자동 추천 서비스를 제공한다.

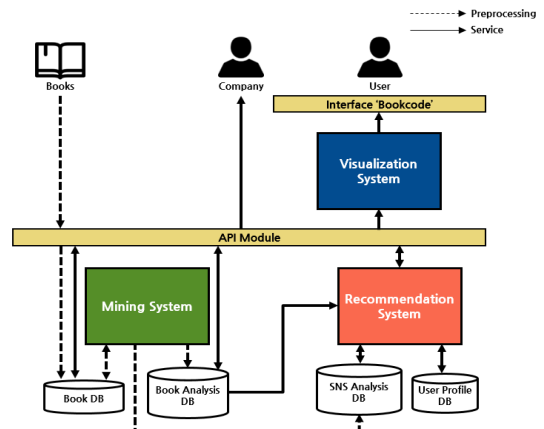
2.2 도서 추천 큐레이션 서비스

도서 분야에서도 추천 큐레이션 서비스의 필요성이 대두되고 있다. Noh[2014]는 클라우드 서비스, 빅데이터, 상황 인식 기술, 이용자 맞춤형 서비스, 도서 추천 서비스, 페이스북을 활용한 신규 도서 추천 서비스 등과 같은 차세대 디지털 도서관 서비스를 도출하였다. 이에 따라 Park, [2016]은 중, 고등학교 학생 개인의 상황과 필요에 적합한 도서를 추천하는 학교도서관 북 큐레이션 서비스를 제안하였고, Pho et al.[2015] 또한 도서관 이용자의 도서 열람, 대출 편의를 위한 맞춤형 추천 서비스인 디지털 큐레이터 서비스를 제안하였다. 대표적인 온라인 도서 추천 서비스에는 아마존의 야시브(Yasiv)와 국내의 북맥(Bookmac), 리더북스(Ridibooks) 등이 있다. 이 서비스들은 개인 SNS 또는 블로그 등에 남긴 서

평이나 추천 및 공유 기능을 기반으로 하여 도서를 추천해주는 것이 특징이다. 이런 서비스를 통해 거대 출판사 마케팅에 의한 베스트셀러 위주의 추천이 아닌 시장에 노출되기 어렵던 소외 장르 혹은 무명의 우수 콘텐츠들이 빛을 볼 수 있는 창구가 마련되었다[Na et al., 2014]. 기존 도서 추천 큐레이션은 단순한 도서 서지 정보와 사용자 선호도 및 구매이력 기반의 추천 시스템으로 구성되어있다. 하지만 도서정가제로 인해 오프라인 서점 대비 가격 경쟁력이 약해진 온라인 서점의 경우 사용자의 취향이나 본문 내용 분석과 같은 고차원적이고 구매의욕을 불러일으키는 도서 분석 및 추천 서비스가 필요하다.

3. 사용자 맞춤형 도서 추천 큐레이션 시스템

본 장에서는 도서 정보 및 본문 텍스트 통합 마이닝 기반의 사용자 맞춤형 도서 추천 큐레이션 시스템을 제안하고 각 기능에 대한 설명과 연구 현황을 작성한다. 본 시스템은 <Figure 1>과 같은 구조로 마이닝 시스템(Mining System), 추천 시스템(Recommendation System), 시각화 시스템(Visualization System)으로 이루어져 있다.



<Figure 1> The Proposed System Architecture

점선 화살표는 전처리를 나타내며, 실선 화살표는 서비스를 나타낸다. ‘마이닝 시스템’은 서지 정보 및 도서 본문 텍스트, SNS 데이터를 분석하여 추천에 필요한 정보를 추출한다. ‘추천 시스템’은 사용자의 프로파일 관리와 ‘마이닝 시스템’에서 추출된 분석 정보를 이용하여 사용자의 성향을 추출하며 추출된 성향에 따른 맞춤형 도서를 추출한다. ‘마이닝 시스템’과 ‘추천 시스템’에서 분석 및 추출된 정보는 ‘API 모듈(API Module)’을 통해 사용자 혹은 ‘시각화 시스템’에 전달되어 사용자가 직관적으로 이해할 수 있도록 시각화한다. ‘Bookcode’는 사용자에게 서비스를 제공하는 인터페이스(Interface)를 말한다.

3.1 마이닝 시스템(Mining System)

‘마이닝 시스템(Mining System)’은 사용자에게 도서를 추천하기 위해 다양한 정보를 분석하는 시스템으로 <Figure 2>와 같이 ‘도서 전처리(Book Preprocessing) 모듈’과 ‘도서 분석(Book Analysis) 모듈’, ‘SNS 분석(SNS Analysis) 모듈’로 구성된다. ‘도서 전처리 모듈’은 서지 정보 및 도서 본문 텍스트가 들어오면 분석을 위한 포

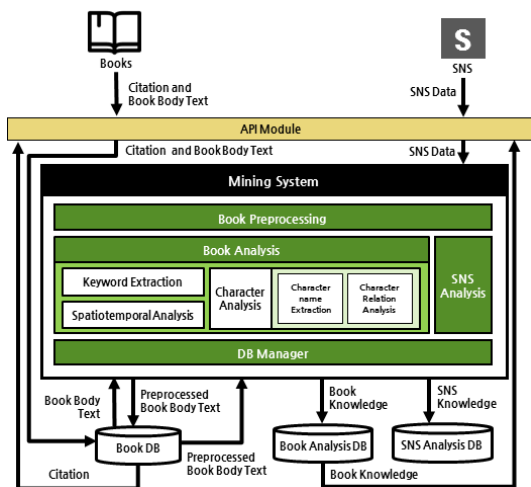
맷으로 일치시키는 역할을 하고, 전 처리된 도서는 ‘도서 데이터베이스(Book DB)’에 저장된다. ‘도서 분석 모듈’에서는 전 처리된 도서 본문 텍스트를 추출 목적에 따라 분석하며 분석 결과는 ‘도서 분석 데이터베이스(Book Analysis DB)’에 저장된다. 마지막으로 ‘SNS 분석 모듈’은 사용자의 SNS로부터 들어온 데이터를 이용하여 사용자의 성향을 추출하는데 필요한 정보를 분석하며 분석된 결과는 ‘SNS 분석 데이터베이스(SNS Analysis DB)’에 저장된다.

3.1.1 도서 분석(Book Analysis) 모듈

사용자 맞춤형 도서를 추천해주기 위해서는 도서 서지 정보와 같은 기본적인 자료 외에 본문 분석을 통해 구체적이고 다양한 내용의 정보를 추출할 필요가 있다. 따라서 도서 본문을 분석하는 ‘도서 분석 모듈’은 ‘키워드 추출(Keyword Extraction) 모듈’과 ‘시공간 분석(Spatiotemporal Analysis) 모듈’, ‘인물 관련 분석(Character Analysis) 모듈’ 세 가지 서브 모듈로 구성되어있고, 분석된 결과는 ‘도서 분석 데이터베이스(Book Analysis DB)’에 저장된다. 저장된 마이닝 결과는 ‘추천 시스템’에서 사용되며, ‘API 모듈’ 또는 ‘시각화 시스템’을 통해 사용자에게 제공된다.

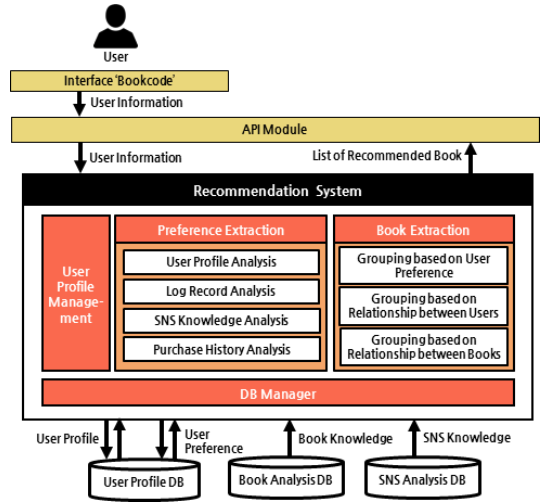
‘키워드 추출 모듈’은 도서 본문 내에서 핵심이 되는 키워드를 추출하는 모듈이다. 키워드는 문서의 의미 또는 내용을 요약하거나 표현할 수 있고, 더 나아가 문서 검색 및 문서 분류 등 많은 분야에서 활용될 수 있다.

‘시공간 분석 모듈’은 도서의 스토리가 진행되는 배경인 시간과 공간 정보를 추출한다. 도서에서 시간 및 공간적 배경은 도서 본문에서 인물이 활동하는 시대 및 장소를 의미하기 때문에 사용자가 도서 내용을 파악하는데 도움이 되며, 사용자가 선호하는 시간, 공간적 배경을 가진 도서를 추천하기 위한 요소로 사용할 수 있다.



(Figure 2) Mining System

‘인물 관련 분석 모듈’에서는 도서 내에 등장하는 주요 인물들을 추출하는 ‘인물명 추출(Character Extraction) 모듈’과 추출된 인물간의 관계를 파악하는 ‘인물 간 연관관계 분석(Character Relation Analysis) 모듈’로 구성된다. 도서에서 등장인물은 스토리를 이끌어나가는 중요한 역할을 하며, 또한 주요 등장인물과 그들 간의 설정된 관계에 따라 스토리 전개를 파악할 수 있다. 도서에서 추출된 주요 인물명이 ‘홍길동’인 경우, ‘홍길동’이 추출된 다른 도서와의 연관성 혹은 인물관계의 복잡성은 도서를 추천할 수 있는 요소로 사용될 수 있다.



<Figure 3> Recommendation System

3.1.2 SNS 분석(SNS Analysis) 모듈

‘SNS 분석(SNS Analysis) 모듈’은 사용자의 SNS 텍스트를 분석하는 모듈이다. 오늘날, 트위터나 페이스북과 같은 SNS(Social Network Service)의 사용이 증가하고 있으며, 사용자들은 SNS를 이용하여 자신의 생각을 표현한다. 본 모듈에서 분석한 SNS 정보는 특정 상품 또는 콘텐츠에 대한 선호여부를 확인할 수 있기 때문에 도서 구입 내역이 없는 신규 사용자에 대한 성향 파악에 도움을 줄 수 있다.

3.2 추천 시스템(Recommendation System)

‘추천 시스템(Recommendation System)’은 사용자의 성향에 적합한 도서를 추천하는 시스템으로 <Figure 3>과 같이 ‘사용자 프로필 관리(User Profile Management) 모듈’, ‘사용자 성향 추출(Preference Extraction) 모듈’, ‘도서 추출(Book Extraction) 모듈’로 구성되어 있다. ‘사용자 프로필 관리 모듈’에서는 사용자의 정적 프로필 정보(생년, 성별, 직업 등)와 동적 프로필 정보(로그기록, 구매내역 등)를 생성 및 관리한다. ‘성향 추출 모듈’에서는 사용자의 프로필, 로그기

록, SNS 분석 정보, 구매내역을 통해 사용자의 성향을 추출하고 ‘사용자 프로필 데이터베이스(User Profile DB)’에 저장한다. ‘도서 추출 모듈’은 분석된 사용자 성향, 사용자간 관계, 도서간의 관계를 기반으로 유사 도서들을 추출한다. 예를 들면, 사용자가 판타지 소설 카테고리에 속한 도서들에 대해 SNS 공유 및 구입 등의 행위를 하면 ‘성향 추출 모듈’에서는 해당 사용자가 도서 카테고리 중에서 판타지 소설에 대한 선호도가 높다는 것을 파악하며, ‘사용자 프로필 관리 모듈’을 통해 ‘사용자 프로필 데이터베이스’에 판타지 소설에 선호도가 높다는 사용자의 성향이 저장된다. 이를 기반으로 도서 추천에 있어 판타지 소설에 속한 도서를 다른 분야의 도서보다 높은 가중치를 가지게 되며, ‘도서 추출 모듈’을 통해서 추천 도서 목록을 추출한다. 추출된 추천 도서 목록은 ‘API 모듈’ 또는 ‘시각화 시스템’을 통해 사용자에게 제공된다.

3.2.1 사용자 프로필 관리(User Profile Management) 모듈

사용자 프로필은 정적 프로필과 동적 프로필로 구성된다. 정적 프로필은 생년, 성별,

직업 등 사용자의 기본 정보를 의미하며, 동적 프로파일은 사용자의 로그기록, 구매내역, 성향과 같이 변화하는 정보를 의미한다. 본 시스템에서는 사용자의 변화하는 성향을 파악하기 위해 프로파일을 동적으로 생성, 삭제, 업데이트 한다. 사용자로부터 정보를 입력받아 프로파일을 생성하고, 사용자의 성향이 바뀌면 프로파일 정보를 업데이트하도록 제안한다.

3.2.2 사용자 성향 추출(Preference Extraction) 모듈

‘성향 추출(Preference Extraction) 모듈’은 사용자 프로파일, 로그기록, SNS 분석 정보, 구매이력을 기반으로 사용자의 성향을 추출한다. 사용자의 성향이란 사용자가 도서를 구입할 때 고려하는 도서의 특징(주제어, 시공간, 인물 간 연관관계 등)을 통합한 복합적인 요소를 의미한다. 대부분의 추천시스템에서는 정보가 없는 신규 사용자에게 콘텐츠 혹은 상품을 추천해줄 때 어려움을 겪는다. 본 시스템에서는 구매기록이 없는 신규 사용자에게도 적합한 도서를 추천해 주기 위하여 SNS 분석 정보와 사용자의 로그기록까지 사용한 성향 추출을 제안한다.

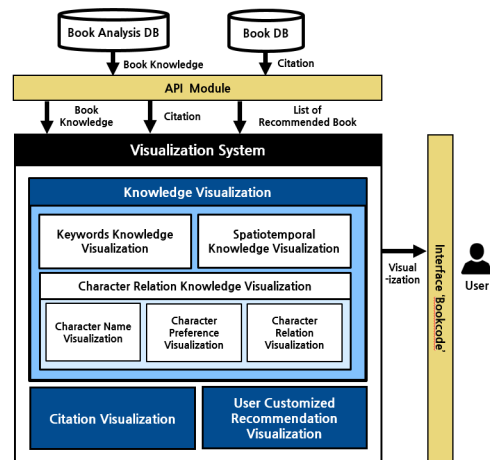
3.2.3 도서 추출(Book Extraction) 모듈

‘도서 추출(Book Extraction) 모듈’에서는 ‘성향 추출(Preference Extraction) 모듈’에서 분석된 사용자 성향, 유사 성향을 가진 사용자들의 구매목록, 도서간의 관계를 기반으로 유사 도서들을 군집하고 각 요소의 우선순위에 따라 도서 추천 리스트를 추출한다. 추출된 추천 도서 목록은 ‘API 모듈’ 또는 ‘시각화 시스템’을 통해 사용자에게 제공된다.

3.3 시각화 시스템(Visualization System)

‘시각화 시스템(Visualization System)’은 서지 정보 및 ‘마이닝 시스템(Mining System)’과 ‘추천

시스템(Recommendation System)’의 결과를 사용자에게 직관적으로 제공하는 시스템이다. 이 시스템은 <Figure 4>와 같이 ‘서지 정보 시각화(Citation Visualization) 모듈’과 ‘사용자 맞춤형 추천 결과 시각화(User Customized Recommendation Visualization) 모듈’, ‘도서 분석 정보 시각화(Knowledge Visualization) 모듈’로 구성되어 있다. ‘서지 정보 시각화 모듈’에서는 ‘도서 데이터베이스’에 저장된 서지 정보를 시각화하며, ‘도서 분석 정보 시각화 모듈’에서는 ‘분석 데이터베이스’에 저장된 키워드, 시공간, 인물 관련 분석 정보들을 시각화한다. ‘사용자 맞춤형 추천 결과 시각화 모듈’은 ‘추천 시스템’에서 사용자의 성향에 적합한 추천 도서 목록을 제공받아 시각화한다.



<Figure 4> Visualization System

3.3.1 서지 정보 시각화(Citation Visualization) 모듈

‘서지 정보 시각화(Citation Visualization) 모듈’에서는 출판사에서 제공해주는 도서의 기본 정보(도서명, 저자명, 출판사, 장르 등)를 시각화한다.

3.3.2 도서 분석 정보 시각화 모듈

‘도서 분석 정보 시각화(Knowledge Visualization) 모듈’에서는 ‘마이닝 시스템’에서 분석된 결

과를 사용자에게 직관적으로 제공하기 위해 시각화해준다. 이 모듈은 ‘키워드 정보 시각화(Keywords Knowledge Visualization) 모듈’, ‘시공간 정보 시각화(Spatiotemporal Knowledge Visualization) 모듈’, ‘인물 관련 정보 시각화(Character Relation Knowledge Visualization) 모듈’로 구성되어 있다.

‘키워드 정보 시각화(Keywords Knowledge Visualization) 모듈’은 ‘키워드 추출 모듈’에서 추출한 도서 본문 내 핵심 단어를 시각화한다.

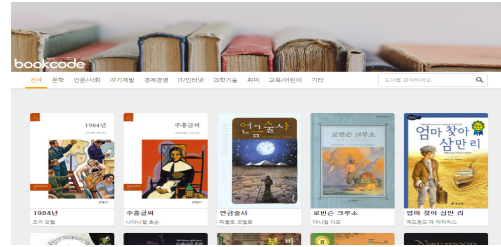
‘시공간 정보 시각화(Spatiotemporal Knowledge Visualization) 모듈’에서는 ‘시공간 분석 모듈’에서 추출한 도서의 시간적 배경과 공간적 배경을 시각화한다. 본 모듈에서는 도서 내 시간 흐름에 따른 공간 배경의 변화를 나타낼 수 있으며, 특정 시간의 주요 등장인물이나 키워드 등을 부가적으로 제공해 줄 수 있다.

‘인물 관련 정보 시각화(Character Relation Knowledge Visualization) 모듈’에서는 ‘인물 관련 분석 모듈’에서 분석한 결과를 시각화한다. ‘인물명 시각화(Character Name Visualization) 모듈’은 ‘인물명 추출 모듈’에서 추출된 인물 리스트를 시각화한다. ‘인물 간 연관관계 시각화(Character Relation Visualization) 모듈’에서는 ‘인물 간 연관관계 분석 모듈’에서 분석된 인물 간 연관관계 정보를 사용자에게 제공하기 위해 시각화한다.

4. 시스템 설계 및 구현

본 논문에서 제안한 시스템은 ‘bookcode’라는 이름으로 구현하였으며, 본 장에서는 시스템 내 모듈 중에서 인물 간 연관관계 모듈과 본문 및 사용자 행동을 기반으로 하는 추천 모듈에 대한 자세한 설계와 구현을 기술한다. <Figure 5>은 본 논문에서 제안한 시스템을 구체화한 서비스인 시범 사이트 ‘bookcode’의 서비스 화면이다.

(a)는 메인 화면이며, (b)는 해당 도서에 대한 서지 정보, 장르 분석, 인물명, 인물 관계 등의 도서 정보를 제공하는 서비스 화면이다.



(a) Main Page of ‘Bookcode’



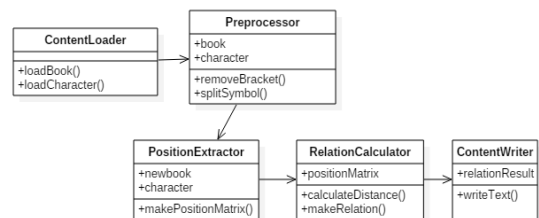
(b) Book Information Page of ‘Bookcode’

<Figure 5> ‘Bookcode’ Service

4.1 인물 간 연관관계 추출 모듈

4.1.1 인물 간 연관관계 추출 설계

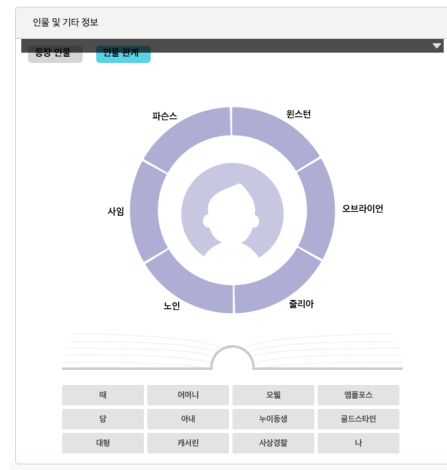
인물 간 연관관계는 ‘인물명 추출 모듈’에서 추출된 인물들의 도서 본문 내 위치한 거리를 이용한 사회구성망의 분석 방법을 사용한다[Park et al., 2013]. <Figure 6>은 인물 간 연관관계 추출을 위한 클래스 다이어그램이며, <Table 1>은 클래스 명세를 보여준다.



<Figure 6> Class Diagram of Character Relations Extraction

<Table 1> Class Specification of Character Relations Extraction

Class	Description
Contents Loader	Load book text and character list
Preprocessor	Reconstruct Book Text by separating the sentences based on the sentence delimiter
Position Extractor	Identify the appearance of the characters in the character list in the book text
Relation Calculator	Calculate the degree of relationship between two characters by using position value of each characters
Contents Writer	Store the calculated result of character relation in database

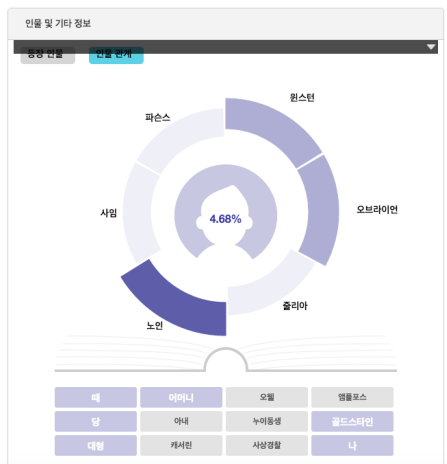


(a) Characters Information of Novel '1984'

4.1.2 시연 결과

<Figure 7>은 인물명과 인물 간 연관관계를 추출한 결과를 시각화하여 제공하는 본 서비스의 일부이다.

<Figure 7>은 소설 '1984'에 등장하는 인물과 인물 간의 관계 정도를 파악할 수 있다. (a)는 도서에 등장하는 주요인물과 그 외의 인물을 나타내고 있으며 중앙의 원 그래프에 위치한 인물이 주요인물이다. (b)는 해당 인물과 관련 있는 인물들을 나타내고 관련 인물을 클릭하면 해당 인물과의 연관 관계 정도 값을 나타낸다.



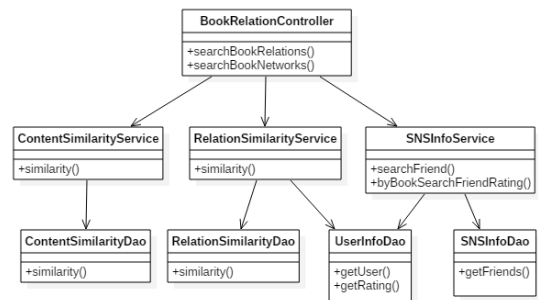
(b) Character Relations Information of '1984'

<Figure 7> Character and Character Relations Result of '1984'

4.2 본문 및 독자 행동 기반 추천 모듈

4.2.1 본문 및 독자 행동 기반 추천 설계

본 모듈에서는 도서 본문에서 추출된 키워드, 독자 프로파일 정보 및 SNS 정보를 기반으로 개인화된 연관 도서를 추출한다. 독자 프로파일은 성별, 연령, 선호도 점수 등이 포함되며, SNS 정보는 사용자의 친구와 도서 선호 정보 등이 포함된다. <Figure 8>은 해당 모듈의 클래스 다이어그램이며, <Table 2>는 클래스 명세를 보여준다.



<Figure 8> Class Diagram of Recommendation based on Book Text and User Information

<Table 2> Class Specification of Recommendation based on Book Text and User Information

Class	Description
Book Relation Controller	Web Service that visualize the recommended books
Content Similarity Service	Related Book Search Service based on keyword
Rating Similarity Service	Related Book Search Service based on used preference
SNS Info Service	Related Book Search Service based on SNS information
Content Similarity Dao	Manage Related Book Data based on keyword
Rating Similarity Dao	Manage Related Book Data based on used preference
User Info Dao	Manage User profile data
SNS Info Dao	Manage SNS information data

4.2.2 시연 결과

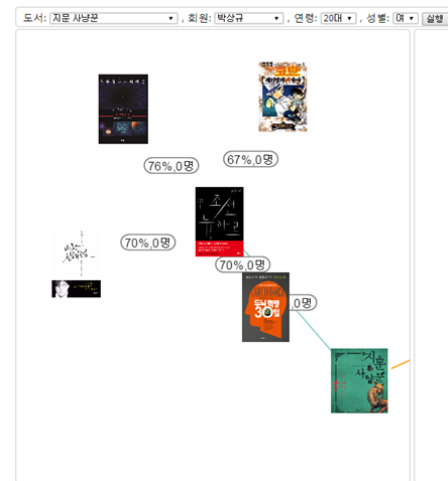
<Figure 9>는 'bookcode'에서 제공하는 연관 도서 추천 서비스의 시각화 일부이다. (a)는 회원가입을 하지 않은 사용자에게 대한 도서 추천 결과이다. 회원가입을 하지 않은 경우, 추출된 도서 키워드를 기반으로 하여 연관 도서를 추천해 준다. (b)는 회원가입을 한 사용자에게 대한 도서 추천 결과이다. 해당 추천에는 도서 키워드 뿐만 아니라 사용자 정보 및 SNS 정보를 기반으로 하여 도서 키워드, 성별, 연령별 등을 기준 요소로 구체적인 도서 추천을 가능하게 한다.

5. 결 론

정보 통신, 스마트 혁명과 함께 데이터양이 폭증함에 따라 사용자와 관련 있는 데이터를 분석하고 선호할만한 데이터를 사용자에게 추천해주는 데이터 큐레이션 서비스가 다양한 콘텐츠 분야에서 화두가 되고 있다. 기존 도서 큐레이션 시스템의 경우 기본적인 도서 서지 정보, 사용자



(a) Related Book Recommendation based on Book without User Data



(b) Related Book Recommendation based on Book, User Data and SNS data

<Figure 9> Results of Recommendation

프로파일 및 로그 기록을 기반으로 추천을 진행하기 때문에 시공간, 인물명과 같은 도서의 구체적인 내용을 기준으로 도서를 추천하기에는 어려움이 있다. 따라서 본 논문에서는 도서 서지 정보, SNS 데이터, 도서 본문 텍스트 등을 분석하여 사용자에게 도서 주제 키워드, 등장인물과 같은 구체적인 정보를 기반으로 도서를 추천해

주는 사용자 맞춤형 도서 큐레이션 시스템을 제안하고 각 모듈에 대한 기능 설명과 특정 모듈에 대한 설계 및 구현 내용을 작성하였다. 제안된 시스템은 추후 인터페이스 개발과 온라인 서점 및 도서관 서비스에 적용하여 사용자에게 차별화된 가치를 제공하는 맞춤형 도서 콘텐츠를 추천할 수 있을 것으로 기대한다. 이는 개인화된 맞춤형 도서 콘텐츠 추천을 통해 도서 선택의 폭을 넓히고 도서 콘텐츠의 균형 잡힌 소비를 유도하게 될 것이다.

References

- [1] Ji, S. J., "Contents Curation Service", *Internet and Security Issue, Korea Internet and Security Agency*, No. 12, 2012, pp. 25-30.
- [2] KOCCA, "Implications for the Big Data Market Conditions and Content Industries", *Kocca Focus, Korea Creative Content Agency*, Vol. 34, 2014, pp. 49-63.
- [3] Lee, H. H. and Lee, W. J., "A Study on the Design of Curation System of Customized Sport Convergence Contents for Activation of Sport for All", *Journal of Korea Multimedia Society*, Vol. 19, No. 2, 2016, pp. 396-404.
- [4] Na, Y. B., Kwon, H. Y., Lee, S. H., "A Study on the Conceptual Change in Audience and Reading Activity for the Development of E-book", *Humanities Contents*, Vol. 35, 2014, pp. 75-93.
- [5] Noh, Y. H., "A Study Suggesting the Development Direction of the Next Generation Digital Library", *Journal of the Korean Society for Information Management*, Vol. 31, No. 2, 2014, pp. 7-40.
- [6] Ovadia, S., "Digital Content Curation and Why It Matters to Librarians", *Behavioral and Social Sciences Librarian*, Vol. 32, No. 1, 2013, pp. 58-62.
- [7] Park, G. M., Kim, S. H., and Cho, H. G., "Analysis of Social Network According to The Distance of Characters Statements", *Journal of the Korea Contents Association*, Vol. 13, No. 3, 2013, pp. 427-439.
- [8] Park, Y. H., "Study on the Book Recommendation Standards of Book-Curation Service for School Library", *Journal of Korean Library and Information Science Society*, Vol. 47, No. 1, 2016, pp. 279-303.
- [9] Pho, S. H., Kim, Y. H., Kim, H. S., and Kim, W. J., "A Study on the Developing of Big Data Service in Public Library", *Korean Society for Information Management*, Vol. 32, No. 2, 2015, pp. 63-86.

■ 저자소개



Hee-Jeong Ahn

Hee-Jeong Ahn received Master's degree in Computer Science from graduate school of Dankook University on Feb. 2017. She was a graduate of

Multimedia Engineering at Dankook University in 2015. She is interested in Natural Language Processing, Text Mining, Data Analytics, etc.



Kee-Won Kim

Kee-Won Kim received his Ph.D. degree in Computer Engineering from Kyungpook National University in 2006, Republic of Korea. Currently, he

is an assistant professor of Dept. of Applied Computer Engineering, Dankook University, Korea. His current research interests are cryptography, VLSI, network security, data mining, etc.



Seung-Hoon Kim

Seung-Hoon Kim received his Ph.D. degree in Computer Science and Engineering from Pohang University of Science and Technology (POSTECH),

Korea in 1998. Dr. Kim is currently a professor of Dept. of Applied Computer Engineering, Dankook University, Korea since 2001. From 1989 to 1990 he was a member of technical staff in Electronics and Telecommunications Research Institute (ETRI), Taejon, Korea. From 1991 to 1993 he was a member of technical staff in POSDATA, Seoul, Korea. His current research interests include data computing and networking, IoT, distributed systems, etc.