

도서 추천을 위한 임의 저자 도서에 대한 시계열 분석 시각화

김서희[○], 정광철^{*}, 이원진^{**}, 김승훈^{***}

[○]단국대학교 멀티미디어공학과

^{*}(주)유엘씨시스템즈

^{**}단국대학교 소프트웨어학과

^{***}단국대학교 응용컴퓨터공학과

e-mail:rlatjgml38@naver.com[○], jungkc@ulcsystems.com^{*},

god7300@dankook.ac.kr^{**}, edina@dankook.ac.kr^{***}

Implementation of Time Series Analysis and Visualization about Author's Books for Book Recommendation

Seo-Hee Kim[○], Kwang-Chul Jung^{*}, Won-Jin Lee^{**}, Seung-Hoon Kim^{***}

[○]Dept. of Multimedia Engineering, Dankook University

^{*}ULCSYSTEMS Inc.

^{**}Dept. of Software Science, Dankook University

^{***}Dept. of Applied Computer Engineering, Dankook University

● 요약 ●

도서 정보 양이 급증하면서 사용자 성향과 선호도에 맞는 정보를 추천해주는 서비스의 중요성이 높아지고 있으며, 이와 관련하여 도서를 추천해주는 플랫폼 연구가 활발하게 진행되고 있다. 독자에게 성향과 선호도에 맞는 추천을 해주기 위해서는 사용자, 도서, 저자 등을 대상으로 하는 분석이 필요하며, 분석된 정보를 사용자에게 직관적으로 제공해주는 것이 필요하다. 따라서 본 논문에서는 저자에 대한 도서 정보를 시계열적으로 분석하고, 분석된 결과를 사용자에게 직관적으로 제공하는 시각화 방법을 제안한다. 제안한 방법은 저자의 도서를 시계열 방식으로 분석하고, 이를 시간 시각화와 레이더차트를 사용하여 도서정보를 제공한다. 또한 시간 시각화와 레이더 차트를 통해 두 저자의 도서 일대기와 분류의 변화를 직관적으로 확인할 수 있다.

키워드: 도서(book), 시각화(visualization), 시계열 분석(time series analysis), 추천(recommendation)

I. 서론

오늘날 정보의 양이 막대한 요즘에는 정보를 분류하고 유용한 정보를 골라내어 사용자의 성향과 선호도에 맞추어 차별화된 정보를 추천해주는 큐레이션 서비스가 중요해지고 있다. 그러나 국내 온라인 서점의 경우 사용자의 검색, 판매량, 신상품, 가격 등 단순한 방법으로만 정보를 제공하고 있다. 방대한 양의 도서 정보를 소비자에게 맞추어 추천해 주기 위해서는 사용자의 성향과 선호도에 대한 종합적인 이해를 바탕으로 도서 정보와 도서 본문 텍스트의 내용적 의미 분석을 통합한 맞춤형 도서 추천 시스템 전략이 미흡한 현실이다. 따라서 기존의 도서 서지 정보와 사용자 선호도 및 구매 이력에 기반한 판매량 위주의 추천 시스템을 탈피하여, 도서 데이터와 본문 텍스트 통합 마이닝을 기반으로 한 내용 중심의 추천 지식 마이닝 체계를 개발하고, 이를 기반으로 사용자 친화적 도서 정보 시각화 및 사용자 참여형 도서 추천 큐레이션 플랫폼 연구가 진행되고 있다.

기존에 사용자, 도서 한 권의 내용에 대한 분석과 그에 대한 시각화

연구는 있었으나, 저자에 대한 시계열에 따른 도서 분석은 찾아볼 수 없었다. 도서의 출판일자와 고객의 구매 일자, 검색 일자 등과 같은 시간 정보와 도서 정보를 사용자에게 전달하기 위해서는 사용자가 이해하기 쉽도록 구체적, 실제적, 직관적으로 데이터를 시각화하는 연구가 필요하다. 본 논문에서는 임의 저자의 도서 정보를 시계열적으로 분석하고, 분석된 결과를 사용자에게 직관적으로 제공하는 시각화 방법을 제안한다.

II. 관련 연구

1. 빅데이터 시각화

빅데이터 시각화는 많은 양의 데이터 중 유용한 정보를 추출하여 사용자가 한 눈에 파악할 수 있도록 하는 작업이다[1]. 빅데이터 시각화는 정보 시각화, 과학적 시각화, 인포그래픽스 등으로 나눌

수 있다. 정보 시각화는 빅데이터를 색, 도표, 그래프, 이미지 등으로 표현하는 것을 말하며, 과학적 시각화는 복잡한 데이터를 쉽게 탐색 가능하도록 3차원 그래픽 기술을 통해 시각화하는 것을 말한다. 인포그래픽스는 복잡한 수나 글로 되어있는 정보를 파트, 지도, 다이어그램 등을 활용한 디자인이다[2]. 이러한 빅데이터 시각화의 기법으로는 시간, 분포, 관계, 비교, 공간 시각화 등이 있다. 각 기법의 특징은 [표 1]에서 보는 바와 같다.

표 1. 시각화 기법의 종류와 특징[1]
Table 1. Type and Feature of Visualization Method

종류	특징
시간 시각화	시계열 데이터의 패턴을 쉽게 확인 가능
분포 시각화	데이터가 어떻게 분포되어 있는지 확인 가능
관계 시각화	데이터들 사이의 관계를 시각화
비교 시각화	다양한 변수의 특징을 비교하여 전체적인 정보 확인 가능
공간 시각화	지도를 이용한 시각화

특히, 본 논문에서는 작가의 도서에 대한 정보를 시계열적으로 분석하고, 분석된 결과를 사용자에게 직관적으로 제공하는 시각화 방법을 제안하고 있기 때문에 다양한 시각화 기법 중에서 시간 시각화 방식이 적합하며, 그 형태는 그림 1과 같은 타임라인 방식이다. 그림 1의 InteractiveSVG 방식의 경우 시계열에 따라 정보를 직관적으로 확인할 수는 있으나 많은 정보를 담아내지 못하는 문제점이 있으며, Horizontal 방식의 경우 많은 정보를 담아낼 수는 있으나 정보를 한 눈에 확인하기 어려운 문제점이 있다. 그리고 TimelineJS의 경우 많은 정보를 담아낼 수 있으며 시계열에 따른 정보를 직관적으로 한 범위에서 표현 가능하다는 장점이 있다.

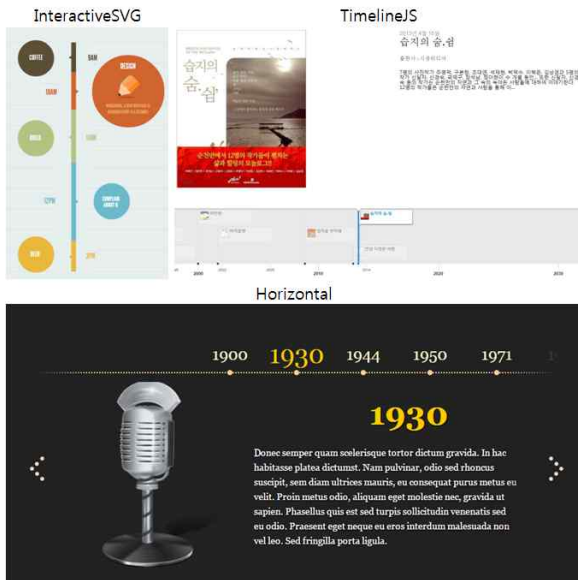


그림 1. 타임라인의 형태
Fig. 1. Type of Timeline

또한, 시각화를 하는 여러 가지 방식 중 내용을 한 눈에 파악하기 쉬운 그래프 방식을 활용한다. 그 중 자주 활용되는 그래프에는 파이 그래프, 막대 그래프, 레이더 차트, 꺾은선 그래프 등이 있다. 파이 그래프는 가장 기본적인 형태의 그래프이며, 막대 그래프는 수량의 증감을 비교, 분석할 수 있다. 레이더 차트는 항목별 강약점을 표시하며 꺾은선 그래프는 항목 간 상대적인 추이를 분석하는데 용이하다[3].

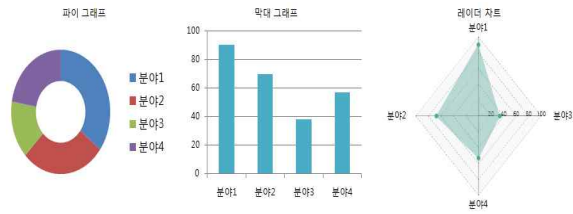


그림 2. 그래프 방식
Fig. 2. Graph Method

2. 도서 정보 시각화

기존의 도서 정보에 대한 시각화 방식에서는 여러 권의 도서를 나열하고 그 중에서 사용자가 선택한 도서의 본문을 어떻게 효과적으로 보여줄 것인가에 초점을 맞추어 시각화 한 것이 대부분이었다.



그림 3. 도서 시각화 예시
Fig. 3. Book Visualization Examples

맞춤형 도서 추천 시스템에서 활용하기 위해서는 도서 정보와 시간 정보를 사용자에게 직관적으로 전달할 수 있는 시각화 기법이 필요하다. 또한 어떤 저자의 도서를 읽었거나 추천 받았을 때, 그 저자가 작성한 다른 도서에도 관심을 가지는 것이 일반적이므로 저자의 도서들을 시계열에 따라 시각화하는 기법이 필요하다. 그러나 한 저자의 도서와 그 정보를 시계열에 따라 시각화한 예시는 미흡하였다.

본 논문에서는 저자의 도서를 시계열 방식으로 분석하고, 이를 시간 시각화와 레이더 차트를 사용하여 도서 정보를 제공하는 방법을 제안하였다.

III. 본 론

본 논문에서는 저자의 도서 정보를 시계열적으로 분석하고, 분석된 결과를 사용자에게 직관적으로 제공하는 시각화 방법을 제안한다. 제안 방식은 시계열에 따른 도서 시각화를 맞춤형 도서 추천 시스템에

서 활용하기 위한 방법으로 시간 시각화 기법과 그래프를 이용하였다. 시간 시각화를 이용하면 사용자는 임의 저자의 도서와 도서 분석 결과, 도서 출판 일자, 도서 구매 일자, 검색 일자 등과 같은 시간 정보를 직관적으로 이해하기 쉽다. 또한 그래프를 이용하면 사용자는 많은 양의 도서 정보를 구체적이고 직관적으로 제공받는다.

그리고 시간 시각화를 위해서 임의 저자의 도서를 타임라인 형태로 제공한다. 타임라인으로 임의 저자의 도서나 주제 등에 따른 도서 결과를 보여줄 경우 저자의 작품 일대기나 문학사조와 같이 온라인 서점에서 제공할 수 없는 정보를 제공할 수 있다. 다양한 타임라인의 형태들을 비교해본 결과 TimelineJS 방식이 시계열에 따른 도서의 흐름과 도서 정보를 표현하기에 적합한 방식이다.

임의 저자의 도서 데이터와 본문 텍스트를 마이닝한 결과를 통해 분야 정보를 추출하여 사용자에게 직관적으로 제공하기 위해 그래프를 사용한다. 분야 정보에는 로맨스, 스릴러, 경제, 여행 등이 있다. 임의 도서1에 대한 마이닝 결과로 분야를 추출하고 그래프를 그려 비교해 본 결과 파이 그래프와 막대그래프보다 레이더 차트가 직관적으로 결과를 확인하는데 더 좋은 것으로 판단되어 본 논문에서는 레이더 차트를 사용한다.

타임라인 시각화와 레이더 차트를 테스트하기 위해 도서 본문에 대한 마이닝으로 획득한 데이터 대신 수작업으로 임의 작가의 도서를 분석하여 데이터로 사용하였다.

그림 4는 비문학 작가인 저자A가 작성한 도서A1, 도서A2, 도서A3, 도서A4에 대하여 타임라인 시각화를 한 결과이다. 그림 4를 통해 저자A의 도서 일대기를 확인할 수 있다.

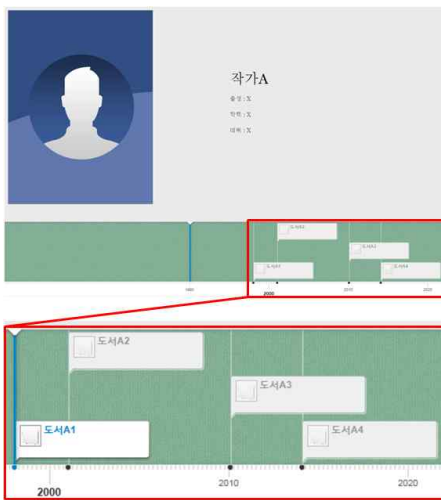


그림 4. 비문학 저자A의 도서 타임라인 시각화 결과
Fig. 4. Result of Non-literature Author A's Book Timeline Visualization

그림 5는 저자A의 도서A1~도서A4를 분석하여 추출한 분야 정보 데이터로 레이더 차트를 그린 결과이다.

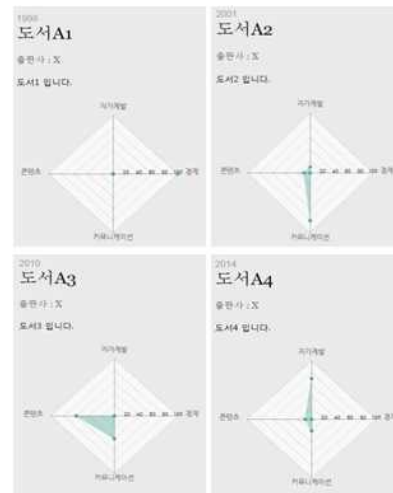


그림 5. 저자A 각 도서의 레이더 차트 결과
Fig. 5. Results of Book Radar Chart of the Author A

저자A의 도서 타임라인 시각화 결과와 레이더 차트 결과를 통해 저자A 도서의 분류가 경제에서 커뮤니케이션, 콘텐츠와 커뮤니케이션, 자기개발로 변화한 것을 확인할 수 있다.

그림 6은 문학 작가인 저자B가 작성한 도서B1, 도서B2, 도서B3, 도서B4에 대하여 타임라인 시각화를 한 결과이며 그림 7은 저자B의 도서B1~도서B4를 분석하여 추출한 분야 정보 데이터로 레이더 차트를 그린 결과이다.



그림 6. 문학 저자B의 도서 타임라인 시각화 결과
Fig. 6. Result of Literature Author A's Book Timeline Visualization

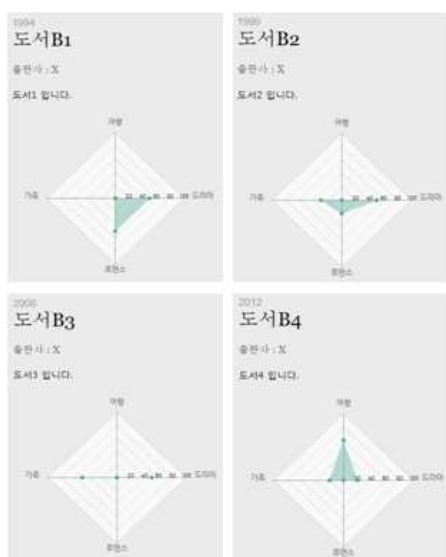


그림 7. 저자B 각 도서의 레이더 차트 결과
Fig. 7. Results of Book Radar Chart of the Author B

저자B의 도서 타임라인 시각화 결과와 레이더 차트 결과를 통해 저자B 도서의 분류가 드라마와 로맨스에서 드라마, 드라마와 가족, 여행으로 변화한 것을 확인할 수 있다.

IV. 결 론

정보의 양이 급증하면서 사용자의 성향과 선호도에 맞는 정보를 추천해주는 서비스가 중요해짐에 따라 도서 추천 큐레이션 플랫폼 기술을 개발하고 있다. 이와 같은 기술 개발에서 임의 저자의 도서에

대해 시계열에 따라 분석한 결과를 사용자에게 직관적으로 제공하기 위한 시각화 모델을 제안하기 위해 분석한 결과 도서를 시계열에 따라 시각화한 예제는 찾을 수 없었다.

여러 시각화 기법에 대해 분석한 결과 도서를 시계열에 따라 시각화하기 위해서는 시간 시각화가 적절했으며 타임라인을 통해 도서의 시간 흐름을 확인할 수 있었다. 그래프의 경우 다른 그래프에 비해 레이더 차트가 도서 분류를 직관적으로 표현하기에 적당했다.

본 논문에서는 타임라인 시각화와 레이더차트를 통해 두 저자의 도서를 시계열에 따라 시각화하였다. 그 결과 저자의 도서 일대기와 분류의 변화를 확인할 수 있었다.

Acknowledge

본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2014년도 문화기술연구개발지원사업의 연구결과로 수행되었음

참고문헌

- [1] Seong-Yeon Ju, Ji-Seong Jeong, Kwan-Hee Yoo, "Big Data Technology Trends Instance of Big Data Visualization and Public Data Visualization", It's Smart Media, pp. 37-43, Sep. 2013.
- [2] Kwang-Sun Choi, Young-Kyoung Ham, Sun-Ho Kim, "Big Data Visualization", Journal of The Korea Society of Computer and Information, Vol. 21, No. 1, pp. 33-43, Jun. 2013.
- [3] Se-Kyoung Pyeon, "Pyeon Se-Kyoung's Practical Teaching Manuals", Communicationbooks, Jan. 2014.