



Interest Data Visualization based on R Language for Personalized Service

Chul-Jin Kim, Su-Chul Hwang^{*}

Department of Computer Systems and Engineering, Inha Technical College

ABSTRACT

Selecting user's interest information in massive contents provided to users of various web services and mobile services gives fatigue to users and lowers efficiency. The personalized service can provide the user with the information of interest, but the existing method of providing the information of interest is not intuitive to the user and the utilization is not high because it provides the uniformized information of the character form. In this paper, we propose a visualization architecture and process of personalization service that can provide users with intuitive information of interest. We propose an Online Data Collector, an Offline Data Collector, a Personalization Service and a R Adapter for each component of the architecture for personalization service visualization. It also defines the sequential processing of the components within the architecture. It collects and analyzes user's interest data on-line and off-line, visualizes it through R language, and provides it to web service and mobile service. In the case study, we verify the suitability of the proposed architecture and the processing process by collect, analysis, and visualize interest information that is transmitted between offline beacon sensors and mobile device users. Through the visualization service proposed in this paper, users can receive intuitive personalization services and service providers can provide effective services.

© 2017 KKITS All rights reserved

KEYWORDS : Visualization, Personalized Service, Interest Data, R Language, Data Collection

ARTICLE INFO: Received 17 October 2017, Revised 2 November 2017, Accepted 8 December 2017.

^{*}Corresponding author is with the Department of Computer Systems and Engineering, Inha Technical

College, 100 Inha-ro Nam-gu Incheon, 22212, KOREA.
E-mail address: schwang@inhac.ac.kr

1. 서론

개인화 서비스는 사용자에게 현명한 소비를 제공할 수 있는 서비스이다. 그러나 기존 모바일 서비스와 웹 서비스에서 제공되는 개인화 서비스는 사용자 검색 키워드나 주문한 상품 정보 분석을 통해 사용자에게 추천 기능으로 제공한다. 이러한 개인화 서비스는 서비스 제공자나 사용자 양측 모두에게 최적의 서비스를 제공하는 데는 한계가 있다. 그래서 본 연구에서는 기존 개인화 서비스의 시너지를 높이고자 온라인과 오프라인 서비스 상에서 수집된 데이터를 기반으로 빅데이터 기반 R 언어를 활용하여 사용자에게 시각적인 개인화 서비스를 제안한다. 제안된 시각화 서비스를 통해 사용자는 직관적인 개인화 서비스를 제공받을 수 있으며 서비스 제공자는 효과적인 서비스를 제공할 수 있을 것이다.

본 논문의 구성은 2장에서 시각화 관련 연구를 분석하며, 3장은 개인화 서비스를 위한 관심정보의 시각화를 위한 아키텍처와 처리 과정을 제안한다. 4장은 본 연구의 아키텍처와 처리 과정의 적합성을 검증하기 위한 사례 연구를 수행한다. 5장 결론에서는 향후 시각화에 대한 추가연구 방안을 제시한다.

2. 관련연구

2.1 개인화 서비스

개인화 서비스를 위한 추천 알고리즘은 협업적 필터링(Collaborative Filtering)과 컨텍스트 기반 필터링(Context based Filtering)으로 구분될 수 있다.[1] 협업적 필터링은 사용자별로 선호하는 정보와 비슷한 사용자들을 찾아 다음 그 사용자가 선호하는 정보를 추천하는 방식이다. 전자상거래 사

이트에서 사용자의 행동과 탐색 패턴을 분석하여 개인화된 정보를 제공하는 방식이다[2,3,4].

컨텐츠 기반 필터링 기법은 사용자의 선호 정보를 활용하는 것이 아니라 사용자에게 의해 접근한 콘텐츠의 속성을 활용하는 방식이다. 콘텐츠의 속성과 사용자가 요구하는 정보 간의 유사도를 계산하고 콘텐츠를 순위화 한다. 이렇게 순위화된 콘텐츠는 사용자에게 개인화된 정보로 제공될 수 있는 방식이다[5,6].

2.2 개인화 공통 프레임워크

연구 [7]은 모바일 환경에서 개인 정보를 분석하여 사용자가 쉽게 필요한 서비스를 발견하고 선택하기 위한 방법론과 접근법을 제시하고 있다. 연구 [7]은 사용자에게 공통된 플랫폼을 제공하기 위해 공통적인 작동구조, 시스템 구성요소, 정보요소에 대한 식별 및 가변적 구성요소를 식별하여 실행시점에 개인화된 서비스를 제공하고자 한다. 연구 [7]은 서비스 측면에서 개인화 방안을 제안하고 있으며 데이터 측면에서는 기존의 연구들과 동일하게 텍스트 측면에서 개인화 정보를 제공하고 있다.

2.3 빅 데이터의 분석 기술

데이터 분석 기술로서 통계 기법을 기반 한 솔루션들이 활용되었으나, 최근 빅데이터 분석에 S 언어 기반으로 개발된 객체지향 패러다임 언어인 R언어가 활용되고 있다[8]. R 언어는 대용량 데이터 처리를 위해 분산 컴퓨팅 기술인 하둡(Hadoop)을 지원한다[9]. 이와 같이 R언어는 하둡 환경에서 맵리듀스(MapReduce)와 연결하여 대용량 데이터를 짧은 시간에 분석할 수 있다[10]. 마훗(Mahout)은 하둡 기반의 맵리듀스 분석 알고리즘을 라이브러리화하여 제공하는 기술로서 분류, 클러스터링, 패

턴 마이닝, 회귀분석, 유사도 측정 등의 분석 기술을 제공한다.

2.4 R 언어

빅데이터 도구인 R 언어는 통계 기법을 지원하는 오픈 소스 소프트웨어로서 시각화를 위한 다양한 기능을 제공한다[11]. R 언어는 CRAN(Comprehensive R Archive Network)[12]을 통해 현재 2만개 이상의 패키지를 이용할 수 있다. R 언어는 패키지 설치를 통해 기능 확장을 구현할 수 있으며, 데이터 조작에 사용되는 디플라이아르(dplyr), 데이터 시각화에 사용되는 지지플롯2(ggplot2), 동적인 표를 생성할 수 있는 아르차트(rCharts), 데이터 수집에 사용되는 아르베스트(rvest), R로 웹앱을 구현할 수 있는 샤이니(shiny)가 있다.

본 연구에서는 워드카운트(wordcount) 패키지를 이용하여 사용자의 수집된 관심정보를 시각화한다.

2.5 데이터 시각화

방대한 정보들 중에서 분석을 통해 의미 있는 정보를 제공하기 위한 기술들이 연구되고 있다[13]. 데이터를 수집하여 분석하는 기법들은 텍스트 마이닝, 소셜 네트워크 분석, 공간 데이터 분석 등이 연구되고 있으며 예측을 위해 회귀분석이나 일반화 선형 모델(GLM) 등이 연구되고 있다[14]. 이러한 분석에 대해 사용자에게 의미 있게 전달하기 위해서는 복잡한 수치형태의 텍스트 정보가 아닌 직관적이고 즉각적인 반응을 할 수 있는 시각적인 정보로 제공될 수 있도록 데이터 시각화가 요구된다. 데이터 시각화는 단순한 도형형태로 제공하는 것이 아니라 사용자의 개인화 성향에 맞게 이해하기 쉽도록 정보 시각화에 대해 연구가 진행되고 있다. <그림 1>에서와 같이 빅데이터 언어인 R 언

어를 통해 SNS를 통해서 분석된 용어들의 네트워크를 시각적으로 나타내므로 얼마나 많은 용어들이 사용되고 관련되고 있는지를 시각적으로 파악할 수 있다.

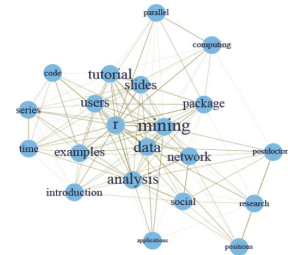


그림 1. R 언어에 의한 소셜 네트워크 분석 사례
Figure 1. Example of Social Network Analysis with R(using igraph package)[15]

연구 [13]에서는 데이터 시각화에 대한 사례연구 분석을 통해 효과적인 정보 전달 방안을 제시하였다. 지도기반 데이터에 대한 시각화 기술로써는 2D, 또는 3D 지형 데이터를 기반으로 지역 특성에 맞는 객체정보를 이미지를 변환하여 시각화할 수 있음을 제시하였다. 텍스트 정보에 대한 시각화 기술은 키워드의 반복횟수, 폰트 사이즈, 회전, 폰트 효과 등을 적용하여 단순 텍스트 정보를 직관적으로 제공할 수 있음을 제시하였다.

3. 개인화 서비스를 위한 관심정보의 시각화

본 연구에서는 사용자의 관심정보를 수집하여 분석한 후 사용자에게 시각적인 관심정보를 제공한다. 사용자의 관심정보는 온라인 서비스와 오프라인 서비스에서 수집하며 오프라인 서비스의 데이터 수집은 비콘 기반으로 관심정보를 수집할 수 있다. 수집된 데이터에 대해서는 R 언어를 분석한 후 시각화하여 사용자에게 직관적인 관심정보를 제공한다.

3.1 관심정보 수집 및 시각화 아키텍처

사용자의 관심정보를 분석하여 시각화하기 위한 아키텍처는 <그림 2>에서와 같이 온라인과 오프라인 서비스 상에서 관심정보를 수집 서버(Collection Server)에 의해 수집된다. 수집한 정보는 원시 데이터로서 수집 서버 내의 분석 모듈에 의해 분석되어 사용자의 관심정보로 제공될 수 있다. 사용자에게 제공되는 관심정보는 직관적으로 사용자에게 제공될 수 있도록 시각화한다.

<그림 3>의 온라인 데이터 수집은 온라인 서비스 상에서 거래되는 데이터를 기반으로 수집할 수 있으며, 오프라인 데이터 수집은 오프라인 상에 설치된 센서(예. 비콘, NFC, 등)와 모바일 디바이스 간에 상호 전송되는 데이터에 의해 수집 할 수 있다.

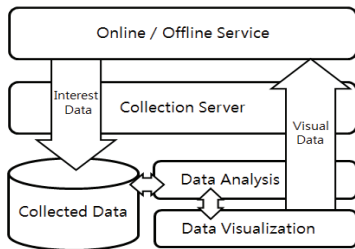


그림 2. 사용자 관심 정보의 시각화 개념도
Figure 2. Concept Diagram for Visualizing User's Interest Information

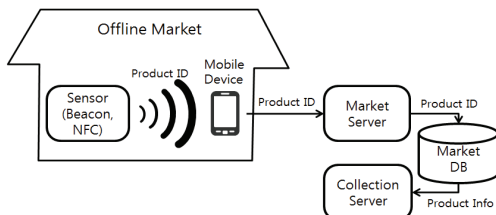


그림 3. 오프라인 상에서의 사용자 관심 데이터 수집 구조
Figure 3. Collection Structure of Offline User's Interest Data

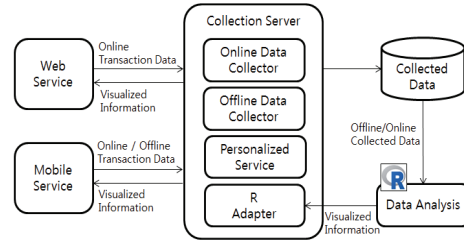


그림 4. 사용자 관심 정보의 시각화 아키텍처
Figure 4. Architecture for Visualizing User's Interest Information

관심정보의 시각화를 위한 시스템 아키텍처는 <그림 4>와 같다. 핵심 요소로서 수집 서버 내에는 온라인 데이터 수집기(Online Data Collector), 오프라인 데이터 수집기(Offline Data Collector), 개인화 서비스(Personalized Service), R 어댑터(R Adapter)로 구성된다.

(1) 온라인 데이터 수집기

온라인 데이터 수집기는 웹이나 모바일 환경에서 사용자가 서비스를 이용하여 거래하는 데이터를 수집하기 위한 모듈이다. 사용자에게 의해 사용하는 서비스는 상품을 주문하거나 검색을 통해 관심도를 갖는 데이터에 대해 수집하게 된다.

(2) 오프라인 데이터 수집기

오프라인 데이터 수집기는 온라인 수집기가 직접적으로 데이터에 대한 온라인 접근을 통해 수집되지만, 오프라인 수집기는 물리적인 환경에서 사용자에게 의해 관심을 갖게 되는 상품에 대해 센서와 사용자 모바일 디바이스 간에 전송되는 데이터에 대해 수집된다. O2O(Online to Offline) 환경에서 온라인과 오프라인 서비스의 경계가 없는 상황에서 오프라인 데이터 또한 사용자의 관심 정보를 도출해 내는데 중요한 자원이 될 수 있다.

(3) 개인화 서비스

개인화 서비스는 온라인과 오프라인 환경에서 수집된 데이터를 분석하기 위한 모듈이다. 관심정보 분석은 수집된 데이터의 개수와 유사도를 체크

하여 동일 그룹으로 구성하고 비중이 높은 데이터를 대표 관심정보로 정의한다. <그림 5>에서와 같이 유사도 체크를 통해 비중이 높은 단어의 그룹으로 포함되며 포함된 단어들은 대표 관심정보의 단어로 변환되어 관리된다.

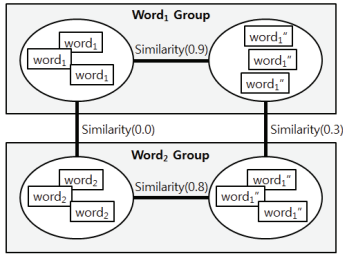


그림 5. 사용자 관심 데이터 단어 그룹화
Figure 5. Word Grouping of User's Interest Data

(4) R 어댑터

R 어댑터는 분석된 관심정보를 추출하여 R 언어에서 처리할 수 있도록 입력 데이터로 전달한다. R 언어는 입력된 데이터를 기반으로 시각화하여 제공한다. <그림 6>에서와 같이 R 어댑터는 유사도에 의한 전처리 과정을 통해 단어의 크기가 결정된다. word1과 word1'를 모두 word1으로 분류하며, word2와 word1''를 word2로 분류하여 전처리한다.

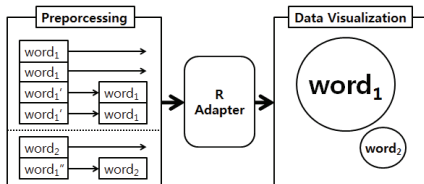


그림 6. R 어댑터의 전처리
Figure 6. Preprocessing by R Adapter

시각화된 관심정보 이미지는 <그림 7>에서와 같이 개인화 서비스로 반환되어 사용자가 상세정보로

연결될 수 있도록 하이퍼링크를 구성한다.

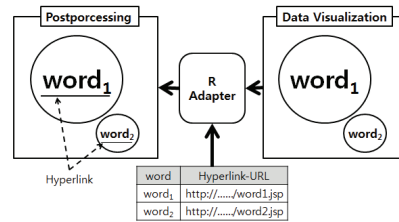


그림 7. R 어댑터의 후처리
Figure 7. Postprocessing by R Adapter

3.2 관심정보 수집 및 시각화 처리 과정

사용자 관심정보의 수집 및 분석 시각화 처리 과정은 <그림 8>의 순차도와 같다. 웹 서비스나 모바일 서비스에서 처리되는 거래 데이터가 수집기에 의해 수집되어 저장된다. 이렇게 수집된 데이터를 기반으로 관심정보를 요청하면 개인화 서비스와 R 언어를 기반으로 시각화된 정보를 제공한다.

개인화를 위한 분석 과정은 <그림 6>에서와 같이 데이터베이스에 저장된 수집 데이터를 읽어 와서 유사도를 측정하여 유사한 단어의 그룹을 구성한다. 해당 단어 그룹의 대표 단어를 추출하고 R 언어에 가중치를 부여하여 개인화 시각화에 대한 정보를 제공한다.

<그림 9>에서와 같이 개인화 서비스는 R 어댑터를 통해 R 언어로 단어와 가중치 정보를 제공하여 시각화 정보를 제공 받는다. 또한 개인화 서비스는 개인화된 시각화 정보에 하이퍼링크를 추가하여 사용자에게 의해 관심정보에 접근할 수 있는 처리를 수행한다.

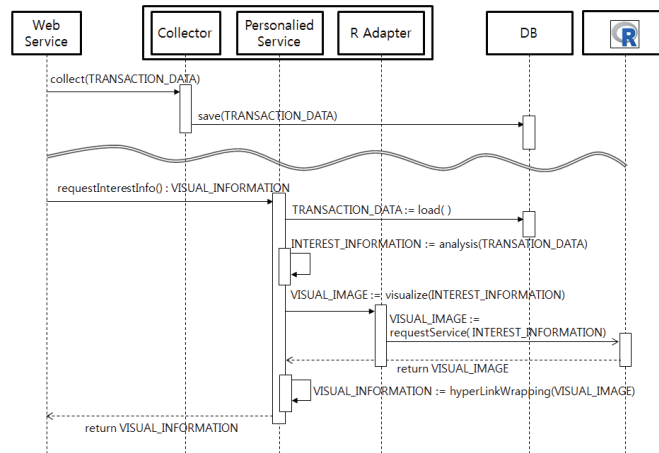


그림 8. 데이터 수집과 시각화 처리 과정
Figure 8. Data Collection and Visualization Process

```

VISUAL_INFORMATION requestInterestInfo( userID )
{
    TRANSACTION_DATA = DB.load( userID );
    INTEREST_INFORMATION = analysis( TRANSACTION_DATA );
    VISUAL_IMAGE = R_Adapter.visualize( INTEREST_INFORMATION );
    VISUAL_INFORMATION = hyperLinkWrapping( VISUAL_IMAGE );
    return VISUAL_INFORMATION;
}

requestInterestInfo( userID ) : Method of User's interest information
VISUAL_INFORMATION : Visualized User's interest information
TRANSACTION_DATA : Transaction data of User in Online and Offline
INTEREST_INFORMATION : Analyzed interest information of User
VISUAL_IMAGE : Generated interest information image by R Adapter
    
```

그림 9. 시각 정보의 생성 알고리즘
Figure 9. Generation Algorithm of Visual Information

지금까지 웹 서비스나 모바일 서비스에서 온라인과 오프라인 상에서 수집된 정보를 기반으로 사용자 관련 처리 정보를 수집하여 개인화 서비스와 R 언어를 통해 시각화된 관심정보를 제공하기 위한 아키텍처 및 처리 프로세스를 제안하였다.

4. 실험 및 평가

본 사례연구에서는 오프라인 상의 비콘 센서와 모바일 디바이스 사용자 간에 전달되는 관심 정보를 수집하여 본 연구에서 제시한 아키텍처와 처리

프로세스에 적용하여 적합성을 검증한다.

<그림 10>은 비콘 센서에 의해 전달되는 데이터를 사용자가 수신하여 클릭이나 거래 등의 처리를 할 경우, 해당 데이터가 서버로 전송되어 데이터베이스에 저장된 관심 데이터를 나타내고 있다.

collect_id	userid	collect_data	collect_path	data_url
1	cjkm	Coffee	online	http://114.71.137.150:8080/Coffee_Service.jsp
2	cjkm	SmartPhone	online	http://114.71.137.150:8080/SmartPhone_Service.jsp
3	cjkm	SmartPhone	online	http://114.71.137.150:8080/SmartPhone_Service.jsp
4	cjkm	SmartPhone	online	http://114.71.137.150:8080/SmartPhone_Service.jsp
5	cjkm	SmartPhone	online	http://114.71.137.150:8080/SmartPhone_Service.jsp
6	cjkm	Book	online	http://114.71.137.150:8080/Book_Service.jsp
7	cjkm	Book	online	http://114.71.137.150:8080/Book_Service.jsp

그림 10. 온라인과 오프라인에서 수집된 관심 데이터 사례
Figure 10. Collected Interest Data from Online and Offline

```

palette<-brewer.pal(9,"Set1")
data1<-readLines("collect_data_2.txt")
wordcount<-table(data1)
wordcloud(names(wordcount),freq=wordcount,scale=c(5,1),rot.per=0
.1,min.freq=1,random.order=F,random.color=T,colors=palette)
legend(0.3, 1, "사용자 선호 상품 ", cex=0.8, fill=NA, border=NA,
bg="white", text.col="red", text.font=3, box.col="red")
    
```

그림 11. R 어댑터의 요청 코드
Figure 11. Request Code of R Adapter

<그림 11>은 서버에 수집된 관심 데이터를 R 언어에 의해 관심 정보로 시각화하기 위한 요청 처리 코드이다. R 언어가 제공하는 패키지 중에 관심 단어에 대한 시각화를 제공하기 위한 wordcloud 패키지를 이용하여 처리한다. <그림 12>는 wordcloud 패키지를 활용하여 데이터 시각화한 결과를 나타낸다.

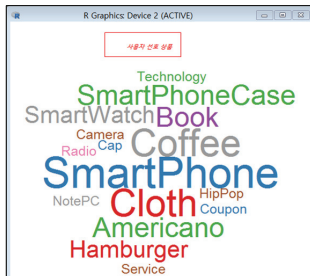


그림 12. R 언어에 의해 생성된 시각화 이미지
Figure 12. Generated Visual Image by R Language

<그림 13>은 R 어택터에 의해 수집 서버로 전달된 데이터에 대한 개인화 서비스에서 하이퍼링크를 구성하여 모바일 서비스로 링크 한 것이다. 하이퍼링크의 정보는 <그림 10>에서와 같이 수집될 때 저장된 정보와 관련된 사이트 이다.



그림 13. 모바일 서비스에 제공된 시각화 정보
Figure 13. Provided Visualization Information to Mobile Service

이와 같이 본 연구에서 제안한 사용자 관심 정보를 시각화하기 위한 아키텍처 및 처리 프로세스의 실현 가능성을 확인할 수 있다. 이러한 시각화를 통해 사용자는 관심 정보에 대해 직관적으로 접근할 수 있으므로 사용성(Usability)을 향상시킬 수 있을 것이다.

5. 결 론

본 연구에서는 온라인과 오프라인 상에서 사용자가 관심을 갖거나 처리하는 데이터를 수집 분석하여 직관적인 개인화 관심 정보를 제공하기 위한 R 언어 기반의 아키텍처 및 프로세스를 제안하였다. 본 연구를 통해 전자상거래에서 제공되는 획일화된 개인화 서비스 범위를 넘어 시각적인 개인화 서비스를 제공함으로써 사용자의 관심도를 높일 수 있는 서비스를 개발할 수 있는 기반을 제공한다. 향후에는 본 연구의 아키텍처를 기반으로 시각적으로 더욱더 다양한 개인화 분석을 제공할 수 있도록 R 언어 외에 다른 빅데이터 분석 도구를 연동할 수 있는 확장성을 연구할 것이다.

References

- [1] Y. S. Kim, *Research trend of recommendation system for personalized service*, ie Magazine, Vol. 19, No. 1, pp. 37-42, 2012.
- [2] H. J. Ahn, *A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem*, Information Sciences, p. 178, 2008.
- [3] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, *Item-based collaborative filtering recommendation algorithms*, Association for Computing Machinery, 2001.

- [4] J. Herlocker, J. Konstan, and J. Riedl, *Explaining collaborative filtering recommendations*, In ACM 2000 conference on computer-supported collaborative work, 2000.
- [5] D. Salber, A. Dey, and G. Abowd, *The context toolkit: Aiding the development of context-aware application*, In the Workshop on Software Engineering for Wearable and Pervasive Computing (Limerick Ireland), Jun. 2000.
- [6] B. Matthias, and D. Scharam, *A survey on context-aware systems*, Distributed Systems Group, Technical University of Vienna, Nov. 2004.
- [7] K. S. Choi, and S. D. Kim, *A common framework for mobile service personalization*, Journal of KISS : Software and Applications, Vol. 39, No. 9, pp. 688-702, Sep. 2012.
- [8] J. S. Kim, *Big data analysis technologies and practical examples*, Korea Contents Association, Vol. 12, No. 1 pp. 14-20, 2014.
- [9] *HDFS(Hadoop distributed file system)*, <http://hadoop.apache.org>, Oct. 2017.
- [10] RHIFE(R and hadoop integrated processing environment), <https://www.rhipe.com>, Oct. 2017.
- [11] *R programming language*, <http://www.r-project.org>, Oct. 2017.
- [12] *The comprehensive R archive network*, <https://cran.r-project.org>, Oct. 2017.
- [13] H. G. Kim, *Case study of bigdata visualization centre around the visual representation form*, Journal of Integrated Design Research, Vol. 13, No. 4 pp. 126-136, 2014.
- [14] H. R. Oh, B. S. Bae, and H. J. Moon,

Analysis of market district hinterland in Seoul by 'Big Data' analysis, Korean Academic Society Of Business Administration, pp. 344-372, 2017.

- [15] Y. Zhao, *An example of social network analysis with R using package igraph*, <https://rdatamining.wordpress.com/2012/05/17/an-example-of-social-network-analysis-with-r-using-package-igraph>, 2012.

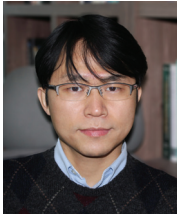
개인화 서비스를 위한 R 언어 기반의 관심 데이터 시각화

김철진, 황수철

인하공업전문대학 컴퓨터시스템과

요 약

다양한 웹 서비스 및 모바일 서비스의 사용자에게 제공되는 방대한 콘텐츠 속에서 사용자의 관심 정보를 선택하는 것은 사용자에게 피로감을 주며 효율성을 저하 시킨다. 개인화 서비스를 통해서 사용자에게 관심 정보를 제공할 수 있으나, 기존의 관심 정보를 제공하는 방법이 문자 형태의 획일화된 정보를 제공하므로 사용자에게 직관적이지 못하며 활용도도 높지 않다. 이에 본 연구에서는 사용자에게 직관적인 관심 정보를 제공할 수 있는 개인화 서비스의 시각화 아키텍처와 프로세스를 제안한다. 개인화 서비스 시각화를 위한 아키텍처의 각 구성 요소인 온라인 데이터 수집기, 오프라인 데이터 수집기, 개인화 서비스 그리고 R 어댑터에 대해 제안한다. 또한 아키텍처 내의 구성 요소들 간에 처리 과정을 순차적으로 정의한다. 사용자의 관심 데이터를 온라인, 오프라인 상에서 수집 및 분석하며 R 언어를 통해 시각화하여 웹 서비스 및 모바일 서비스에 제공한다. 사례연구에서는 오프라인 상의 비콘 센서와 모바일 디바이스 사용자 간에 전달되는 관심 정보를 수집, 분석, 시각화하여 본 연구에서 제시한 아키텍처와 처리 프로세스의 적합성을 검증한다. 본 논문에서 제안한 시각화 서비스를 통해 사용자는 직관적인 개인화 서비스를 제공받을 수 있으며 서비스 제공자는 효과적인 서비스를 제공할 수 있을 것이다.



Chul Jin Kim received his B.E. degree in Computer Science from the Kyonggi University in 1996. He received both M.S. and Ph.D. degree in Computer Science from the Soongsil University in 1998 and 2004. He worked as a senior researcher in SAMSUNG Electronics from 2004 to 2009. He has been a professor in the Department of Computer Systems and Engineering at Inha Technical College since 2009. He research focus is software engineering, object-oriented & component technique, software architecture, customization, and embedded software.

E-mail address: cjkim@inhac.ac.kr



Su Chul Hwang received the bachelor's, the M.S. and Ph.D. degree in the Department of Computer Science from the Chung-ang University in 1986, 1988 and 1993, respectively. He has been a professor in the Department of Computer Systems and Engineering at Inha Technical College since 1991. His main research fields include artificial intelligence, intelligent systems, internet application, IT fusion system etc. Recently he is interesting in an interface application programming based on Visual C++ MFC development environment. He is a life member of the KKITS.

E-mail address: schwang@inhac.ac.kr