

빅데이터 기반 정보 추천 시스템

이종찬¹ · 이문호^{2*}

Big data-based information recommendation system

Jong-Chan Lee¹ · Moon-Ho Lee^{2*}

¹Department of Computer Information Engineering, Kunsan National University, Gunsan 54150, Korea

^{2*}Department of Multimedia Science, Chungwoon University, Incheon 22100, Korea

요 약

삶의 질의 향상으로 인하여 건강관리는 현대인의 주요 관심 사항이며 자연스럽게 헬스케어 시스템에 대한 요구가 증가하고 있다. 그러나 인터넷 상에는 다양한 의료 관련 정보가 존재할 뿐만 아니라 이 정보들의 신뢰성 또한 가늠하기 힘든 것이 현실이므로, 특정 사용자에게 적합한 맞춤형 웰니스 정보 제공은 어려운 것이 현실이다. 본 연구에서는 빅데이터를 텍스트 마이닝으로 분류하여 사용자 맞춤형 의료정보를 제공함으로써 단순 검색기능이 아닌 사용자에게 적합한 맞춤 서비스를 제공할 수 있는 사용자 중심의 서비스 제공 방법을 제안한다. 효율적인 빅데이터 분석을 위해 하둡 슬레이브 노드를 증가하면서 데이터 처리시간을 실험하였다. 기존 시스템보다 빅데이터 시스템을 구축하는 것이 효율적임을 확인하였다.

ABSTRACT

Due to the improvement of quality of life, health care is a main concern of modern people, and the demand for healthcare system is increasing naturally. However, it is difficult to provide customized wellness information suitable for a specific user because there are various medical information on the Internet and it is difficult to estimate the reliability of the information. In this study, we propose a user - centered service that can provide customized service suitable for users rather than simple search function by classifying big data as text mining and providing personalized medical information. We built a big data system and measured the data processing time while increasing the Hadoop slave node for efficient big data analysis. It is confirmed that it is efficient to build big data system than existing system.

키워드 : 헬스케어, 하둡노드, 빅데이터, 텍스트 마이닝, 웰니스

Key word : Healthcare, Hadoop Node, Big data, Text Mining, Wellness

Received 6 November 2017, Revised 8 November 2017, Accepted 15 January 2018

* Corresponding Author Moon-Ho Lee(E-mail:mhlee@chungwoon.ac.kr, Tel:+82-032-770-8201)

Department of Multimedia Science, Chungwoon University, Incheon 22100, Korea

Open Access <http://doi.org/10.6109/jkiice.2018.22.3.443>

pISSN:2234-4772

©This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Copyright © The Korea Institute of Information and Communication Engineering.

I. 서론

u헬스케어3.0 시대는 행복한 삶의 질을 목적으로 하는 의료소비자가 증가하면서 ICT 융복합 기술을 적용하여 언제 어디서나 의료서비스가 가능한 건강관리서비스를 선보이는 등 웰니스(Wellness) 산업에 대한 관심이 지속적으로 확대되고 있다. 신체적·정신적·사회적·환경적 건강을 통한 행복한 삶을 추구하는 웰니스의 가치가 건강식품, 레저, 스포츠, 문화 등으로 빠르게 확산되고, 이에 따라 웰니스 서비스가 점차 개인화함에 따라 개인 맞춤형 웰니스 서비스에 대한 수요도 지속적으로 증가하고 있다[1-5]. 하지만 웰니스 서비스의 확대에도 불구하고 비즈니스 측면에서 성공적으로 평가받는 분야는 ‘개인건강관리’ 및 ‘다이어트 코칭’ 서비스가 대부분을 차지하고 있다. 국내 웰니스 산업은 병원 솔루션을 제공 중인 중소·중견기업을 중심으로 센서·단말 등 디바이스 분야에 집중되고 있으며, ‘웰니스 지식 분야’는 빅데이터 기술 등 요구기술은 확보되어 있으나 표준화 및 UI/UX 부재 등으로 인해 현장에 안정적인 상용화로는 이어지지 못하고 있는 실정이다.

인터넷, 모바일, SNS 등에서 폭증하는 비정형 텍스트 데이터는 사실에 입각한 지식 이외에 작성자의 의견이나 느낌, 관심, 선호, 예상 등을 직간접적으로 반영하고 있기 때문에, 이를 찾아내어 활용하고자 하는 니즈(Needs)는 점차 강해지고 있다. 그러나 이러한 데이터의 출처와 유형 및 내용에 따른 분석 기법도 매우 다양하고[6-10], 언어가 갖고 있는 다양한 특성들로 인하여 그 의미를 분석, 도출하는데 따른 어려움이 발생하므로 구체적인 성과는 아직 미흡한 실정이다. 구체적으로 개발 시 어려움을 기술하면 아래와 같다.

- 수집된 웰니스 콘텐츠 DB의 저장관리 및 분류체계 미흡으로 인해 대용량성, 비정형, 실시간성이라는 데이터의 특징을 수용할 수 없었다.
- 질환별 관심키워드를 수동으로 입력하여 웰니스 콘텐츠를 제공함으로써 인해 사용자의 다양한 질환별 요구에 대응하기 미흡했다.
- 웰니스 의료, 음식, 운동, 문화, 상품 정보 등 다양하고 수많은 데이터가 생성되고 있지만 이를 일괄 및 실시간으로 처리를 하지 못해 사용에 어려움이 많았다.
- 엑셀 및 CSV 등 지정 형식의 자료를 편집하여 업

로드하는 방식과 개별 콘텐츠를 RSS(Really Simple Syndication) 도구를 이용하여 수집하는 방법을 병행하였으나 데이터의 원본 소스가 많아지고, 포맷이 다양해지면서 시간 및 비용이 증가하였다.

따라서 이러한 문제점을 해결하고 웰니스 콘텐츠 정보를 개인 맞춤형으로 쉽게 제공 받을 수 있으며 찾고자 하는 정보를 타인의 웰니스 경험을 통해 손쉽게 공유할 수 있는 서비스의 개발이 필요하다. 이를 위하여 본 연구에서는 빅데이터 기술[11-16]을 활용함으로써 웹에 분산되어있는 의료정보 정보를 맞춤형 정보로 제공함으로써 의료정보 블로그(개인 운영, 병원운영 포함)를 객관적으로 판단할 수 있도록 하는 웰니스 시스템을 제안한다.

II. 제안 시스템의 구조

2.1. 시스템 구성도

그림 1과 같이 제안된 시스템은 크롤링(Crawling)된 데이터의 분석을 위하여 텍스트 마이닝을 적용하고 이를 통해 데이터 속에 숨겨진 패턴을 발견하고, 올바른 콘텐츠의 판단과 사용자의 상황 정보에 맞게 웰니스 콘텐츠를 추천한다.



Fig. 1 System structure for wellness service

제안 시스템의 기본 기능은 아래와 같다.

- 데이터 품질 및 표준에 대한 기준 마련을 통해 웰니스 콘텐츠의 분류 체계를 수립한다.
- 다양한 빅데이터 웰니스 콘텐츠 수집을 위해 비구조적 데이터를 확보한다.
- 빅데이터를 수집·분석하기 위한 원활한 수집체계 및 분석기반을 마련함으로써 양질의 웰니스 콘텐츠 서비스 제공한다.

자료수집/처리 단계에서는 웹 크롤러를 이용하여 자료를 수집하고 이를 MongoDB에 저장한다. 저장된 데이터를 텍스트 마이닝(자연어 처리와 형태소 분석)을 적용하여 데이터 전처리를 수행한다. 분석단계에서는 비구조화 텍스트 데이터를 구조화자료 형태로 단어-문서로 행렬로 변환하여 분석에 적합한 형태로 데이터를 가공한다. 이후, 빈도 및 비율분석을 실시하고 텍스트에 TF-IDF(Term Frequency-Inverse Document Frequency) [17-18]를 적용하여 가중치를 산출한다. 다음으로 텍스트 마이닝한 결과와 가중치가 부여된 결과의 비교를 위하여 빈도 및 비율분석을 실시하고, 이를 바탕으로 추출된 의료시설 블로그를 제공한다.

2.2. 빅데이터 수집 및 통합

다양한 빅데이터 웰니스 콘텐츠 수집을 위해 비구조적 데이터의 데이터 확보가 필요하다. 본 시스템에서는 데이터의 최신 상태 유지를 위해 웹 크롤링을 수행한다. Url, 규칙, 주기, 처리방식을 적용한 크롤링 자동화 기능을 구축하고 링크 체크나 HTML 코드 검증과 같은 웹 사이트의 자동 유지 관리 작업 기능뿐만 아니라 자동 이메일 수집과 같은 웹 페이지의 특정 형태의 정보를 수집 등도 수행한다.

웰니스 데이터의 수집을 위한 기법으로 별도의 스토리지 서버(Storage Server)를 구축하고 인위적인 수집이 아닌 인터넷 포털 사이트의 특정 텍스트를 스크래핑(Scraping)하는 크롤러(Crawler)를 이용하여 데이터를 산출하고 가공 단계를 거쳐서 보관하며, 스토리지 서버에 저장된 데이터를 빅데이터 처리를 위한 마스터 서버(Master Server)와의 네트워크 연동을 통하여 데이터 분석을 수행한다.

2.3. 빅데이터 저장 및 관리 문제

기존의 데이터 저장 구조는 처리 속도가 늦고 관리가

어려웠다. 또한 저장할 데이터양의 증가로 읽기/쓰기에서 RDBMS의 제약이 생기고, 수평 확장에도 한계가 있었다. 본 시스템에서의 데이터인 웰니스 데이터는 비정형 데이터이므로 새로운 저장 기술을 적용한다. 이에 몽고 DB로 웰니스 빅데이터 저장 구조를 구축하였다.

그림 2에 데이터 처리과정을 통하여 빅데이터 시스템의 MongoDB에 저장된 데이터 구조를 보인다. 본 서비스는 의료관련사이의 정보를 빅데이터 화하고 이를 분석하여 사용자에게 다양한 건강정보를 개인 맞춤형으로 추천 제공한다. 분산되어 있는 의료정보 블로그를 활용한 고품질의 다양한 서비스 제공이 가능해짐에 따라 개인 맞춤형 의료정보 블로그를 제공하며 신뢰성 확보를 위한 협업 필터링, 콘텐츠 기반 필터링의 적용 및 각 블로그의 판단 척도를 자사의 내부 정보가 아닌 객관적인 정보를 제공 하여 사용자는 신뢰성 있는 고품질의 의료정보 서비스 제공받을 수 있다.

그림 2에서 의료 정보 데이터는 질환명_data라는 항목을 통해서 날짜별로 관리되고 activities라는 하위 항목에 질환 특성으로 분류하였다. 데이터 분석과정에서는 이와 같이 MongoDB에 저장된 데이터를 Map/Reduce 과정을 통하여 분석한다. 이런 과정을 통하여 분석된 결과는 웹 페이지를 통해서 확인할 수 있도록 한다.

```
{
  "_id" : 12a98b
  "disease name" : "diabetes"
  "Related site" : "www.samsunghospital.com"
  "diabetes_data" : [
    {
      "date" : 2017/09/21
      "activities" : [
        {
          "symptom": weight loss
          "cure" : blood sugar management
          "taking": incretin preparation
          "food": salt
          "exercise": walking
        },
        {
          "symptom": urorrhagia
          "cure": diet
          "taking": meglitinide
          "food": multi-grain rice
          "exercise": swimming
        }
      ]
    }
  ]
},
{
  "date" : 2017/09/22
  "activities" : [
```

Fig. 2 Storage structure of Mongo DB

Table. 1 Analysis of wellness data

Steps	Functions	Development Methodology
Level1: Collection/processing	Selection of data collection object	- Individual and group websites - Website of middle and large hospital - Website of the pharmaceutical company
	Collection of data	- Data collection through web crawler
	Storage of data	- MongoDB-based information storage
	Processing of data	- Natural language processing - Morphological analysis - Keyword extraction using R
Level 2: Analysis	Data conversion	- Structured data conversion of unstructured data - Conversion of unstructured data into word-document matrix
	Weighting	- Weight calculation using TF-IDF
	Conduct of analysis	- Ratio analysis / frequency analysis using R
Level 3: Service provision	Provision of service	- Providing extracted medical information - Providing extracted medical facility information

2.4. 빅데이터 플랫폼 구축

대량의 데이터를 효율적으로 처리하기 위해서는 다수의 서버를 통한 분산처리가 필수적이다. 그림 3과 같이 빅데이터 서버는 하둡 엔진과 빅데이터 미들웨어로 구성된다. 제안한 시스템은 확장성과 비용절감 측면을 고려하여 하둡 프레임워크를 적용하였다. 빅데이터 미들웨어를 통해 가공된 데이터를 MongoDB에 저장하고, 분류 및 분석 작업을 거쳐 맞춤 의료 정보를 생성한다. 이 시스템은 20대의 서버를 사용하여 그 중 1대 서버를 마스터 서버, 나머지 19대의 서버를 슬레이브 서버로 설정하여 분산처리 클러스터를 구축한다.

2.5. 빅데이터 분석

웹 크롤링된 정보로부터 사용자의 상황 정보에 맞는 신뢰성 있는 웰니스 콘텐츠 제공을 위해서는 효과적인 데이터 분석이 필요하다. 따라서 텍스트 마이닝을 통해 데이터 속에 숨겨진 패턴을 확인하고, 키워드 색인 생성 및 우선순위 부여 기능을 구축한다.

표 1에서 나타난 바와 같이 자료 분석 체계는 단어-문서 행렬로 변환된 웰니스 자료를 바탕으로 텍스트 데이터를 의미 있는 정보로 저장하기 위하여 단어별 빈도가 아닌 정보 검색과 텍스트 마이닝에서 사용되고 있는 TF-IDF를 이용하여 각 단어별 가중치를 산출한다. 가중치가 부여된 단어들을 바탕으로 명사로 분리하여 빈도분석 및 비율분석을 통해 웰니스 구성요인을 파악하기 위한 변수를 도출한다. 다음으로 R을 적용하여 탐색적 요인분석을 통하여 웰니스의 구성요인을 도출하고 이를 바탕으로 질환별 판별요인을 파악하여 질환 유형에 따른 웰니스 요인을 결정한다.

2차, 3차 카테고리 세분화 후 검색영역 확장 및 디테일 검색 기능을 구축하고 TF-IDF를 사용하여 구와 절 수준에서의 분석을 수행한다. 즉, 어떤 단어의 중요도는 그 단어가 문서에 나온 총 횟수에 비례하고, 그 단어가 있는 모든 문서의 총 수에 반비례한다. 이러한 TF-IDF 방식을 이용하면 하나의 문서 중에서 가장 가중치 값이 큰 단어가 그 문서에 키워드로 채택한다.

그림 4와 같이 전용 크롤러를 통해 수집된 웹문서의 온라인 리뷰를 분석하기 위해 텍스트 마이닝과 R언어를 이용하여 질환의 유사도를 측정하고 이를 자연어 처

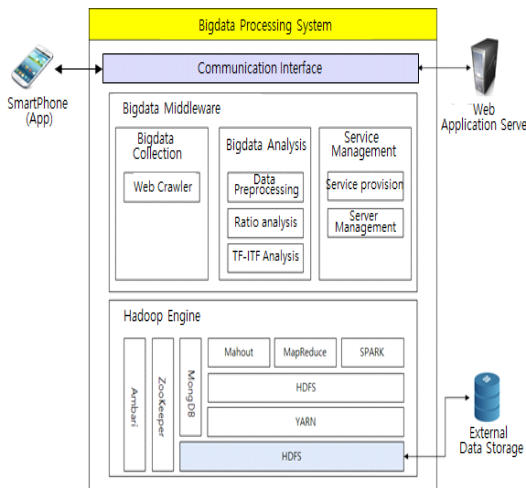


Fig. 3 Big data processing system

리를 통하여 모든 형태소를 소문자로 변환한 후 구두점과 공백, 숫자를 제거한다. 또한 문장 내용을 나타내는데 큰 역할을 하지 않는 관사, 전치사, 조사, 접속사 등의 기능어 또한 제거함으로써 각 단어로 자연어 처리되었다. 그리고 어근을 추출하거나 접사를 제거하는 작업인 어간추출을 통하여 동일한 의미의 단어들을 하나의 단어로 처리하여 텍스트 처리의 효율성을 높인다.

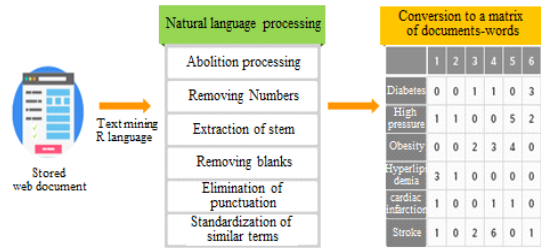


Fig. 4 Classification by disease

그림 5는 텍스트 마이닝이 적용된 화면 설계를 보인다. 텍스트 마이닝을 적용하여 질환별로 크롤링된 데이터의 패턴을 분석하고, 입력된 사용자 정보를 바탕으로 맞춤형 웰니스 콘텐츠를 제공한다.



Fig. 5 Service design with text mining

III. 구현 및 분석

3.1. 테스트 환경

테스트를 위하여 표 2와 같이 질환별 질환의료키워드 사전을 구축하였다. 의료관련 사이트로부터 크롤링으로 수집된 정보를 문장 단위로 분리하고 각 문장을 형태소 분석 후 토큰 배열로 변환한다. 변환된 토큰을 질환 웰니스 사전에 저장한다.

Table. 2 Dictionary of diseases

Wellness Dictionary	disease related medical keyword (disease)
Diabetes	blood sugar, insulin, hyperglycemia
High blood pressure	hypertension, diastole, atherosclerosis
Obesity	Overweight, body/abdominal obesity
Hyperlipidemia	dyslipidemia, cholesterol, High bleeding
Atherosclerosis	heart disease, cerebrovascular disease
Angina pectoris	heart disease, ischemia, cardiomyopathy
Myocardial infarction	cardiovascular, blood clot, electrocardiogram
Stroke	paralysis, ischemia, ischemic attack
Lung disease	pneumonia, pulmonary tuberculosis, asthma
Liver disease	sarcoidosis, fatty liver, liver cancer
Gastric disease	gastritis, gastric ulcer, gastric cancer
Joint disease	joints, arthritis, degenerative, rheumatic
Osteoporosis	skeletal system, fracture, bone density
thyroid	thyroid cancer, thyroiditis
headache	migraine, tension headache, temporal arteritis
Rhinitis	sinusitis, sinusitis, sore throat
Urinary disease	urinary incontinence, prostate, prostatitis
Heart disease	cardiovascular, cardiovascular, arrhythmia
Eye diseases	conjunctivitis, cataract, glaucoma
Skin disease	dermatology, atopy, atopic dermatitis
Depression	depressive disorder, sleep depression
Stress	adaptive disorder, anxiety disorder
Gastric cancer	gastric cancer, lymph node, gastroscopy
Colon cancer	colon, colon polyps, colitis
Liver cancer	cirrhosis, liver cirrhosis, jaundice,
Lung cancer	malignant tumor, tumor, chemotherapy

표 3은 데이터 크롤링 시에 음식 키워드와 운동 키워드 분류 기준을 나타낸다.

Table. 3 Classification of food and exercise keywords

Criteria	Related Keywords	Example
Food	Name of a disease + Name of food(noun)	Good food for diabetes
	Disease related medical keyword + Name of food(noun)	Good food for high blood sugar
	Name of a disease + Name of food	Onions are good for diabetes
	Disease related medical keyword + Name of food	Garlic is good for insulin suppression

Criteria	Related Keywords	Example
Exercise	Name of a disease + Name of food(noun)	Good exercise for diabetes
	Disease related medical keyword + Name of food(noun)	Good exercise for hyperglycemia
	Name of a disease + Name of food	Swimming is suitable for the treatment of diabetes
	Disease related medical keyword + Name of food	Walking is effective for insulin suppression

본 논문에서 제안하는 시스템의 평가를 위하여 표 4와 같이 하둡 에코시스템을 설치한다. 분산데이터를 저장하는 HDFS와 분석 데이터를 처리하는 맵리듀스가 하둡 코어 시스템에 해당하고, 웰니스 정보를 제공하기 위하여 부가적으로 필요한 하둡 에코 시스템으로서 MongoDB, Zookeeper, Sqoop, Pig 등을 적용한다.

Table. 4 Composition of Big Data System

Classification	Components	Version
Master server/ Name node	OS	Cent OS 7.0
	Platform	JDK 7.0
	Hadoop	Apache Hadoop 2.2.0
	MongoDB	2.5.0
	Zookeeper	3.3.5
	Sqoop	1.3.0
	Mahout	0.13.0
Slave server/ Data node	OS	Cent OS 7.0
	Platform	JDK 7.0
	Hadoop	Apache Hadoop 2.2.0
	MongoDB	2.5.0
	Zookeeper	3.3.5

3.2. 성능 평가

그림 6은 각 슬레이브 서버에 데이터를 분산 저장하고 분석하며 그 결과를 종합하는 과정에서 소요되는 시간을 기존의 일반 서버 시스템과 비교한 결과이다. 빅데이터의 3V (데이터의 양(volume), 데이터 입출력의 속도(velocity), 데이터 종류의 다양성(variety)) 중에서 서비스 만족도 향상을 위한 속도 측면에서의 성능을 평가한다. 분석 시간은 30%이상 향상 되는 결과를 보였다. 따라서 수집된 데이터의 양이 증가할수록 슬레이브 서버를 증설하여 분산 처리를 수행하는 것이 보다 효율적임을 알 수 있다.

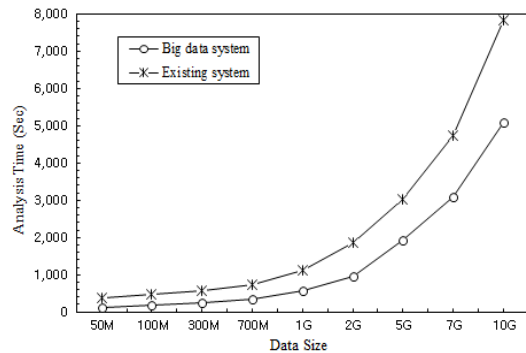


Fig. 6 Comparison of processing time

3.3. 개발 결과

데이터 분석을 위해 맵리듀스 프로그램을 실행하여 처리한 후 아래와 같은 10 종의 메인 앱 페이지를 그림 7과 같이 개발하였다.

기존 방법에 비하여 빅데이터 분류 체계를 개선했으며 텍스트 마이닝으로 키워드 색인 생성 및 우선순위 기능을 부여함으로써 2차, 3차 카테고리 세분화 후 검색 영역을 확장하고 디테일 검색 기능을 강화하였다. 또한 유사, 유관 콘텐츠 검색을 위한 연관 검색어 맵 등 록기능을 추가하였다. 또한 크롤링 자동화 기능을 이용하여 데이터를 수집, 분류하는 시스템 도입하였고 빅데이터 처리(레이어 분산처리) 위한 NoSQL 도입과 데이터 프락시 서버 기능의 적용을 통하여 분석의 효율성을 극대화하였다.

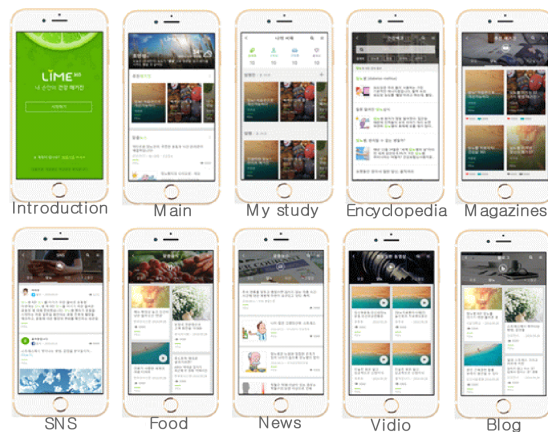


Fig. 7 Developed wellness site

IV. 결 론

웰니스에 대한 관심이 증가함에 따라서 사용자들은 무자각 상태에서 자신의 건강 문제를 지속적으로 점검하고 가장 적절한 시점에 가장 적절한 조치를 취함으로써 자신의 건강상태를 최상으로 유지하는 것이 가능할 것으로 기대된다. 본 논문에서는 빅 데이터의 결과를 기반으로 웰니스 서비스제공 블로그를 사용하였다. 웹에 분산되어있는 의료정보 블로그 정보를 개인 맞춤형 정보로 제공함으로써 개인 의료정보 블로그 및 병원의 의료정보 블로그를 객관적으로 판단 할 수 있도록 함으로써 근거중심의 의료서비스를 통한 진정한 질병치료 및 질병예방이 가능하게 하였다. 향후 분석의 정확성을 확보하기 위하여 키워드 분석에 기계학습을 적용하는 연구가 필요하다.

ACKNOWLEDGEMENT

This research was supported by a grant (16163 MFDS001) from Ministry of Food and Drug Safety in 2016, Republic of Korea.

REFERENCES

- [1] A. S. Oh, "A Study on Design of Health Device for U-Health System," *Bio-Science and Bio-Technology*, vol. 7, no. 2, pp. 79-86, April 2015.
- [2] R. J. Conejar, H. K. Kim, "Designing U-Healthcare Web Services System," *Software Engineering and Its Applications*, vol. 9, no. 3, pp. 209-216, Jan. 2015.
- [3] J. Domingue, N. Lasierra, A. Fensel, T. Kasteren, M. Strohbach, and A. Thalhammer, "Big Data Analysis," *New Horizons for a Data-Driven Economy*, GB: Springer International Pub., ch. 5, pp. 63-86, Sep. 2016.
- [4] Y. Y. Sok and S. H. Kim, "Integrated Medical Information System Implementation for the u-Healthcare Service Environment," *Journal of the Korea Contents Association*, vol. 14, no. 5, pp. 1-7, May 2014.
- [5] Y. E. Gelogo and H. K. Kim, "Integration of Mobile Computing to Ubiquitous Healthcare," *Software Engineering and Its Applications*, vol. 9, no. 9, pp. 295-302, Sep. 2015.
- [6] M. Rostami, S. Ayat, I. Attarzadeh, and F. Saghari, "Proposing a Method to Classify Texts Using Data Mining," *Advances In Computer Research*, vol. 6, no. 4, pp. 125-137, Oct. 2015.
- [7] Y. Li, A. Algarni, M. Albathan, Y. Shen, and M. Arif Bijaksana, "Relevance Feature Discovery for Text Mining," *IEEE Transactions on Knowledge and Data Engineering*, vol. 6, no. 1, June 2015.
- [8] P. Mohata and S. Dhande, "Web Data Mining Techniques and Implementation for Handling Big Data," *Computer Science and Mobile Computing*, vol. 4, no. 4, pp. 330-334, April 2015.
- [9] S. Tang, B. S. Lee, and B. He, "Dynamic job ordering and slot configurations for MapReduce case loads," *IEEE Transactions On Services Computing*, vol. 9, no. 1, pp. 4-17, Jan. 2016.
- [10] M. K. Kim, N. K. Kim, and I. H. Jung, "Including Automatic Keyword Extraction Techniques of Shopping Through Internet Search History Analysis," *Journal of Intelligent Information Systems*, vol. 20, no. 2, pp. 123-136, June 2014.
- [11] B. Dayley, *Node.js, MongoDB, and AngularJS Web Development*, 1th ed. Boston, BO: Addison-Wesley, May 2014.
- [12] H. I. Jung, J. G. Yang, J. I. Woo, B. M. Lee, and Y. H. Lee, "Evolutionary rule decision using similarity based associative chronic disease patients," *Cluster Computing*, Vol. 18, No. 1, pp. 279-291, March 2015.
- [13] F. N. Afrati and J. D. Ullman, "Optimizing Multiway Joins in a Map-Reduce Environment," *IEEE Transactions on Knowledge and Data Engineering*, vol.23, no.9, pp.1282-1298, Feb 2011.
- [14] B. Singh and S. Kumar, "An Effective Information Retrieval with Keyword Optimization," *Computer Technology and Applications*, vol. 5, no. 1, pp. 174-177, Jan. 2014.
- [15] D. H. Shin, "Demystifying Big Data: Anatomy of Big Data Developmental Process," *Telecommunication Policy (ELSEVIER)*, vol. 40, no. 9, pp. 837-854, Sep. 2015.
- [16] S. W. Ha, S. Y. Lee, K. C. Lee, "Development of Reference Architecture Based on Big Data Ecosystem", *Journal of Security Engineering*, vol.11, no.6 pp. 511-522, Sep. 2014.
- [17] M. H. Seo, S. W. Lee, and J. Y. Seo, "A Web-document Recommending System using the Korean Thesaurus," *Journal of the Korea Institute of Information and Communication Engineering*, vol. 13, no. 1, pp. 103-109, Jan. 2009.
- [18] K. A Yang, D. W. Lee, K. H. Kim, and H. J. Yoon, "Analysis of Security Threat and Security Requirements of the Bigdata System", *Journal of Security Engineering*, Vol.13, No.6, pp. 501-514, Dec. 2016.



이종찬(Jong-Chan Lee)

1996년 송실대학교 컴퓨터공학과 (공학석사)
2000년 송실대학교 컴퓨터공학과 (공학박사)
2000년~2005년 한국전자통신연구원 선임연구원
2005년~현재 국립군산대학교 컴퓨터정보통신공학부
※관심분야 : 이동통신, 이동멀티미디어, 정보보안



이문호(Moon-Ho Lee)

1977년 서울대학교 전자공학과 (공학사)
1993년 송실대학교 컴퓨터공학과 (공학석사)
1996년 송실대학교 컴퓨터공학과 (공학박사)
1979년~1984년 한국전자통신연구원 연구원
1997년~ 현재 청운대학교 멀티미디어학과
※관심분야 : 이동통신, 이동멀티미디어, 멀티미디어 통신