

Korean to Korean Translation Based Learning Contents Management System for Parents of Multi-Cultural Family

Yunhee Kang[†] · Myungju Kang^{††}

ABSTRACT

One of the main reasons of information divide of multi-cultural family is caused by language barrier that is associated with low education level. In addition the social problem can be triggered by the information divide that may increase the gap of economic inequality. With respect to the overall capability of accessibility of digital devices and the level of data utilization, the parent of multi-cultural family's level is inferior to that of the parents of an ordinary family. However the traditional learning contents management system for those parents is not appropriate to decrease the gap of the information divide. To handle this problem, it is necessary to construct a customized learning contents management system that is used to support the education of the parents of multi-cultural family depending on the level of understanding the learning contents written in korean. In this paper we design the korean to korean translation based learning contents management system and show the result of its prototype.

Keywords : Multi-Cultural Family, Information Divide, Korean to Korean Translation, Customized Learning Contents

다문화 가정 학부모를 위한 한한변환 기반 학습콘텐츠 관리 시스템

강 윤 희[†] · 강 명 주^{††}

요 약

다문화 가정의 정보 격차의 주요 원인인 언어장벽은 다문화 가정 자녀의 낮은 교육수준과 높은 연관성을 가지며, 경제불균형을 높임으로써 추가적인 사회문제가 될 것으로 예상된다. 다문화가정 학부모는 디지털 기기에 대한 접근성과 자료 활용 능력이 일반국민에 비해 현저하게 떨어지는 정보격차로 인해 기존의 교육 콘텐츠의 효과적 활용은 제약점을 갖고 있다. 이러한 제약점을 해결하기 위해서는 학습자의 이해수준에 적합한 콘텐츠를 제공하는 맞춤형 학습콘텐츠 지원체제의 구축은 필수적이다. 이 논문에서는 다문화가정 학부모의 정보격차 해소를 위한 한한변환 기반 학습콘텐츠 시스템을 설계하고 이를 위한 사용자 맞춤형 학습콘텐츠 지원을 위한 한한변환 시스템의 프로토타입을 결과로 제시한다.

키워드 : 다문화 가정, 정보격차, 한한변환, 맞춤형 학습콘텐츠

1. 서 론

정보화 사회에서 지식과 정보는 중요한 자원으로 구성원의 새로운 지식과 정보에 대한 접근 능력과 활용 능력차이가 점차 커지고 있다. 정보 격차(Information Divide)는 디지털 정보와 정보기술에 접근이 가능한 사람들과 가능하지 않은 사람들 사이의 격차를 의미했으나 정보이용 역량, 활용

성 및 효율성 등으로 기준이 확대되고 있다[1]. 이러한 정보 격차는 경제적, 사회적 격차로 이어질 것으로 예상되며, 정보화 사회의 문제로 등장하고 있다[1-3]. 미래창조과학부와 한국정보화진흥원의 2014년도 정보격차 실태조사에서는 일반국민과 주요 소외계층인 장애인, 저소득, 장노년, 농어민 및 북한이탈주민, 결혼이민여성을 대상으로 정보격차를 평가하였다. 본 실태조사에서 주요 소외계층의 PC 활용 정보화 수준은 지속적으로 향상된 것으로 있으며, 이에 따른 정보격차는 매년 감소하고 있는 추세인 것으로 조사되었다. 그러나 일반인과 소외계층의 컴퓨터 및 인터넷 사용능력 수준을 의미하는 정보이용 역량과 컴퓨터·인터넷 사용의 질적 수준 격차는 개선되지 못하고 있다.

광의적 개념으로 다문화 가정은 자국 내에 거주하는 한명

※ 이 논문은 2016년도 한국정보처리학회 춘계학술발표대회에서 '다문화 가정 학부모를 위한 학습콘텐츠 추천 시스템 설계'의 제목으로 발표된 논문을 확장한 것임.

† 정 회 원 : 백석대학교 정보통신학부 부교수

†† 비 회 원 : 트리니티(주) 이사

Manuscript Received : June 28, 2016

First Revision : August 30, 2016

Accepted : August 30, 2016

* Corresponding Author : Yunhee Kang(yhkang@bu.ac.kr)

이 국적을 취득해 구성된 가정을 의미한다. 우리나라에서는 결혼 이주민의 증가로 전체 가정에서 다문화 가정이 증가하는 추세이다. 다문화 가정 정보격차의 주요한 원인인 언어장벽은 다문화 가정 자녀의 낮은 교육수준과 높은 연관성을 가지며, 경제 불균형을 높임으로서 추가적인 사회문제화가 될 것으로 예상된다.

국내외적으로 스마트러닝(Smart Learning)은 자기주도형 학습과 같은 새로운 학습모델을 통해 기존 학습과 비교하여 높은 성과를 보이고 있다. MOOC (Massive Open Online Courses) 서비스는 양질의 교육콘텐츠를 제공함으로써 교육격차문제의 해결방안이 제시되고 있다[4-7]. 다문화가정 학부모는 현저하게 떨어지는 한국어 이해 수준으로 인한 디지털 기기에 대한 접근성과 자료 활용 능력이 일반국민에 비해 기존의 교육 콘텐츠의 효과적 활용에서 제약점을 갖는다. 이러한 제약점을 해결하기 위해서는 학습자의 이해수준에 적합한 콘텐츠를 제공하는 맞춤형 학습콘텐츠 지원체계의 구축은 필수적이다.

이 논문에서는 다문화가정 학부모의 정보격차 해소를 위한 한한변환 기반 학습콘텐츠 관리 시스템의 소프트웨어 아키텍처를 설계하고 사용자 맞춤형 학습콘텐츠 지원을 위한 한한변환 시스템의 프로토타입을 제시한다. 이 논문의 2장에서는 학습데이터 저장 및 수집을 위한 MapReduce 기반 오픈소스인 Hadoop 및 Apache Flume의 특징과 기술을 기술한다. 3장에서는 설계된 시스템 아키텍처 및 구현 결과를 기술한다. 4장에서는 결론을 기술한다.

2. 관련 연구

2.1 MapReduce 및 Hadoop

MapReduce는 대용량 데이터를 병렬 및 분산처리하기 위한 프로그래밍 모델이다[8, 9]. MapReduce 미들웨어인 Hadoop [10]은 Apache의 오픈소스 프로젝트로서 분산 파일 시스템(Distributed File System)인 HDFS (Hadoop Distributed File System)에 저장된 데이터를 처리한다. 이를 위해 MapReduce 응용은 HDFS 자료를 입력받아 처리하며 처리 결과를 HDFS에 저장한다. HDFS는 높은 확장성(scalability)과 데이터 신뢰성(reliability)을 제공하기 위해 중복(replication)을 갖는 다수의 저가 서버를 클러스터로 구성한다.

MapReduce 프로그래밍 모델은 key와 value의 쌍을 기본 자료구조로 사용한다. key와 value의 자료형으로는 정수, 실수, 문자열, 바이트열 또는 임의의 클래스로서 정의될 수 있다. MapReduce 프로그램은 Map과 Reduce의 함수로 이루어진다. 다음은 두 함수의 구조를 보인 것이다.

Map :: (key1, value1) → list((key2, value2))

Reduce :: (key2, list(value2)) → list((key3, value3))

Map 함수는 주어진 블록의 데이터를 읽고 응용 처리를 수

행한 후 기본 자료구조인 key-value의 쌍인(key2, value2)의 리스트를 생성한다. Reduce 함수는 동일한 중간 key와 관련된 모든 값을 key-value의 쌍인(key2, list(value2))으로 입력받아 새로운 key-value 쌍인(key3, value3)의 리스트를 출력한다. 이 과정에서는 중간 값에 대한 shuffle과 sort를 통한 그룹핑을 수행한다[7].

MapReduce는 프로그래머로부터 시스템 수준의 세부사항을 숨길 수 있는 추상화를 제공하는 장점이 있다. 이는 독립적인 작업으로 대량의 데이터 집합을 병렬 및 분산처리를 할 수 있도록 한다. 병렬 및 분산처리를 위해 입력 데이터는 MapReduce 작업을 위해 특정한 크기의 작은 블록으로 파티션 한다. 이후 Map함수와 Reduce함수의 수행을 위한 Mapper 태스크와 Reducer 태스크가 독립적으로 시작한다.

HDFS 클라이언트는 필요한 데이터 블록을 네임 노드(Name node)에 요청한다. 네임 노드는 파일을 구성하는 데이터 블록의 메타데이터를 유지하며, 클라이언트에 블록의 위치정보를 제공한다. 이후 HDFS 클라이언트는 파일에 대한 실제 연산을 수행하기 위해 데이터 노드(Data node)에 접근하여 처리한다. Fig. 1은 HDFS에서 네임노드와 데이터 노드 간의 상호작용을 보인 것이다.

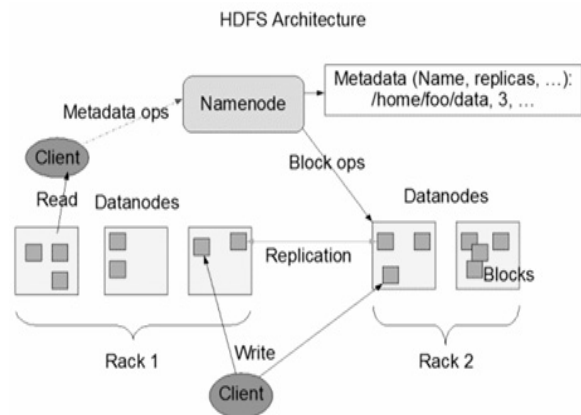


Fig. 1. HDFS Architecture

2.2 Apache Flume

Apache Flume은 다수의 정보소스로부터 대량의 데이터를 안정적이고 효율적으로 HDFS와 같은 목적지에 전달할 수 있는 오픈소스 기반의 시스템이다[11]. Flume은 분산 구조로 데이터양이 급증해도 기존 서버의 설정이나 구조 변경 없이 손쉽게 확장할 수 있는 이점이 있다. Hadoop 응용에 필요한 자료를 처리하기 위해 응용개발자가 HDFS에 파일을 저장하거나 프로그램을 통해 파일 단위로 저장할 수 있지만, 서버 수의 증가에 따라 생성된 로그 파일을 지속적으로 HDFS에 유지하기는 쉽지 않다. 이러한 이유로 Flume은 로깅 시스템(logging)에 사용하며, 다양한 환경에서 생성된 데이터를 Hadoop의 HDFS에 저장하기 위해 사용한다.

Flume 기반의 데이터 수집은 3-tier 구조로써 에이전트로부터의 로그데이터를 수집기를 통해 HDFS에 단일방향으로

저장된다. HDFS에 저장된 로그 데이터를 분석 및 처리하기 위해서는 MapReduce 프로그램을 사용한다. Fig. 2는 Flume을 사용한 데이터 수집 프로세스를 보인 것이다.

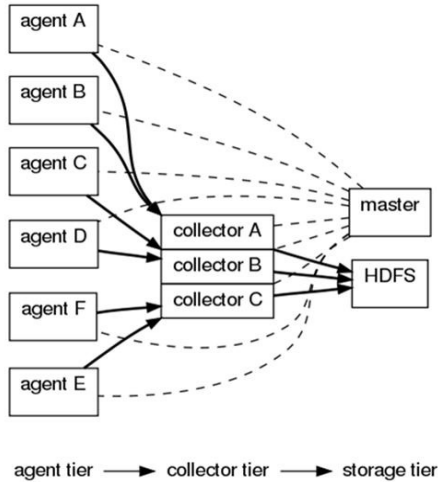


Fig. 2. Data Aggregation Process

3. 다문화 맞춤형 학습콘텐츠 시스템 설계

3.1 사용자 및 시스템 요구사항 정의

다음은 다문화 가정 구성원에게 추천콘텐츠를 제공하기 위한 한한변환 학습콘텐츠 관리시스템의 요구사항이다.

- 다문화 가정 구성원의 학부모는 학교 및 가정 구성원과 원활한 의사소통이 가능하도록 언어적 장벽을 낮추어주기 위한 서비스가 요구됨
- 다문화 가정 구성원의 문화공유 및 학습지원을 위해 다양한 상황에 따른 대처가 가능하고 다문화 구성원이 갖는 다양성을 일반인과 공유할 수 있는 양방향 오픈 커뮤니티가 요구됨
- 개별적인 다문화 가정 구성원의 학습현실과 필요성에

따라 맞춤형 스마트러닝 콘텐츠를 제시하는 시스템이 필요하며, 다문화 가정의 학부모는 학습 성향 및 학습 결과를 분석 한 후 분석결과를 기반으로 학습콘텐츠를 제공하는 것이 요구됨

정의된 요구사항을 기반으로 다문화 맞춤형 학습콘텐츠 관리 시스템 설계 과정에서는 개별적인 다문화 구성원의 학습수준에 적합한 맞춤형 스마트러닝 콘텐츠의 시스템의 주요기능을 다음과 같이 제시한다.

- 다문화 가정 구성원의 학습 효율성을 위해서는 학습자의 수준에 대한 평가가 이루어져야 함
- 다문화 가정 학부모의 수준별 콘텐츠 활용을 기반으로 수준에 따른 학습콘텐츠 추천이 요구됨
- 다문화 가정 학부모를 위한 학습콘텐츠를 위한 메타데이터 표준 활용 및 필요에 따른 확장이 필요함

Fig. 3은 다문화 가정 학부모의 한국어 학습 지원을 위한 한한변환 기반 교육콘텐츠 시스템을 위한 전체 시스템 구성을 보인 것이다. 한한변환 기반 교육콘텐츠 시스템(Korean-Korean translation based learning contents management system)은 사용자환경(user environment), 한한변환 시스템(K2K translation system), 빅데이터 처리 시스템(BigData processing system) 및 학습콘텐츠 추천시스템(Recommendation System)으로 구성된다. 사용자환경에서는 학습에 필요한 학습서비스를 요청하여 학습콘텐츠를 전달 받으며, 한글문장에 대한 한글 변환이 필요한 경우 한한변환 시스템에 요청한다. 한한변환 시스템의 문장전처리 기능을 갖는 형태소 분석기(Morphological analysis system)를 사용하여 한한변환 요청을 처리한다. 한한 변환수행 결과는 로그로 저장하게 되며, 콘텐츠 수집기(Contents aggregator)를 통해 수집된 로그데이터는 빅데이터 처리 시스템의 HDFS에 저장한다. 수집된 한한변환로그는 학습 콘텐츠 추천시스템으로 입력되어 처리되며 이를 통해 추천 학습콘텐츠를 사용자에게 제공한다. 이 논문에서는 한한변환 시스템의 구성을 위한 서브시스템에 한정하여 내용을 기술한다.

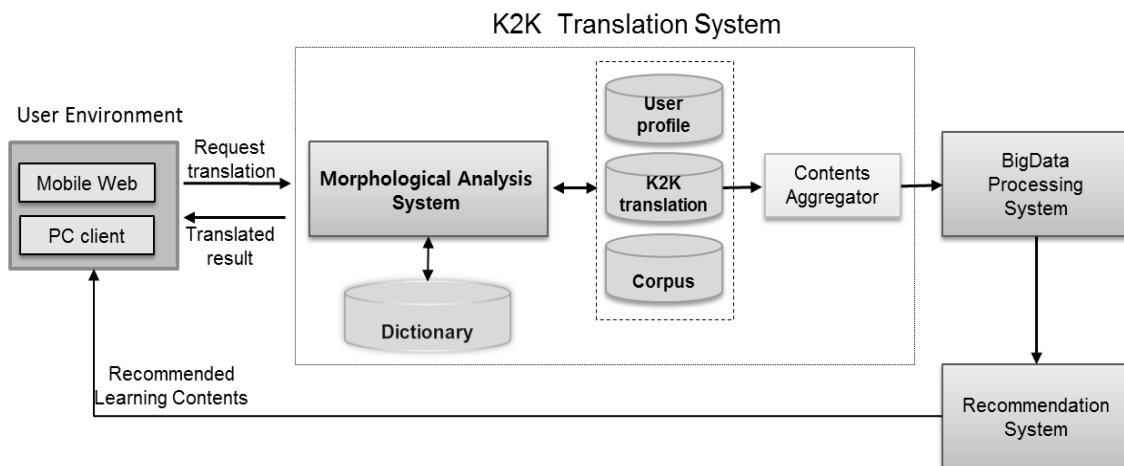


Fig. 3. System Structure of Korean-Korean Translation Based Learning Contents Management System

3.2 Apache Flume 기반 학습콘텐츠 수집

다문화 맞춤형 학습콘텐츠 추천 시스템은 학습콘텐츠의 저장을 위해 대용량 데이터 처리를 위한 Hadoop의 분산 파일 시스템인 HDFS를 사용한다. HDFS는 수집된 자료를 블록단위로 저장함으로써 대용량 데이터의 순차처리의 성능을 개선하도록 한다. 그러나 관리 측면에서 학습콘텐츠를 생성하는 사용자 수의 증가에 따라 생성된 콘텐츠를 지속적으로 HDFS에 저장하기는 쉽지 않다. 이를 해결하기 위해 여기서는 Apache Flume을 사용한 수집기를 구성한다. Apache Flume은 대량의 데이터를 안정적이고 효율적으로 스트림(Stream) 기반으로 전달할 수 있는 오픈소스 기반의 시스템이다[9]. Flume은 분산 구조로 데이터양이 급증해도 기존 서버의 설정이나 구조 변경 없이 손쉽게 확장할 수 있는 이점이 있다. 개발된 수집기는 데이터를 HDFS에 저장하기 위해 입력소스(Input Source)로 한한변환 결과 로그파일을 설정하고 출력싱크(Output Sink)로 HDFS 위치를 설정한다.

Fig. 4는 수집기의 동작을 위해 구성된 Apache Flume의 형상 파일(Configuration File)을 보인 것으로 해당 정보는 이름과 값의 쌍으로 이루어진다. 작성된 형상파일은 구성요소 이름과 이들 구성요소의 설정값의 쌍으로 표현한다. 수집기는 채널을 생성한 후 수집된 자료를 색인 구축과 같은 데이터 처리를 위해 HDFS에 저장한다. HDFS에 저장된 자료는 Hadoop 응용에서 처리하도록 제공한다. Fig. 5에서 a1.sources와 a1.sinks는 설계된 시스템의 소스와 싱크를 선언한 것으로 채널은 소스에서 싱크로 자료의 전달이 이루어지며, a1.channels으로 채널을 선언한다. 소스로부터 자료는 명령어 tail을 사용하여 읽은 후 싱크 유형인 a1.sinks.k1.type에 따라 HDFS에 값을 저장한다. 데이터 전달이 이루어지는 채널의 형식과 전송데이터 크기는 a1.channels.c1.type과 a1.channels.c1.capacity에 값을 지정한다. HDFS에 저장된 학습콘텐츠는 빅데이터 시스템을 구성하는 MapReduce 응용인 색인기의 입력으로써 사용한다.

```
# Name the components on this agent
a1.sources = r1
a1.sinks = k1
a1.channels = c1
# Describe/configure the source
a1.sources.r1.type = exec
a1.sources.r1.command =tail -F K2K_Tran/log..txt
a1.sources.r1.channels = c1
# Describe the sink
a1.sinks.k1.type = hdfs
a1.sinks.k1.hdfs.path= hdfs://localhost/flume/TranData
# Use a channel which buffers events in memory
a1.channels.c1.type = memory
a1.channels.c1.capacity = 1000
```

Fig. 4. Configuration for Apache Flume

3.3 한한변환 기반 용어 추천시스템

다문화 가정 구성원의 정보격차를 해결하기 위한 시스템의 주요 구성요소인 한한 변환 시스템의 구성은 Fig. 5와 같다. Fig. 5의 입력 처리기(Input Handler)는 사용자의 입력 또는 텍스트 파일로부터 선택되어 추출된 텍스트를 입력받는 부분으로 입력텍스트는 문장단위로 요청텍스트로 구성하여 작업 스케줄러(Job scheduler)에 대기큐(Ready queue)로 전달한다. 작업 스케줄러의 대기큐에 전달된 텍스트 문장은 형태소 분석 데몬(Morphological Analysis Daemon)에 해당 문서의 형태소 분석을 위해 전달한다. 형태소 분석 데몬은 용어추천을 위해 사전관리기(Dictionary manager)를 참고한다.

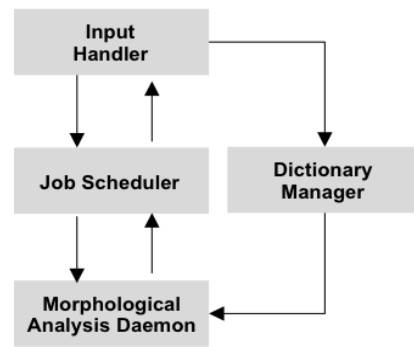


Fig. 5. Morphological Analysis System

Fig. 6은 형태소 분석 데몬(daemon)의 처리과정을 보인 것으로 REST 기반의 형태소 분석요청을 통해 전달된 문장은 단어 단위로 분리된다. 분리된 단어는 사전을 탐색한 후 해당 단어에 매칭 되는 추천단어를 찾아 전달한다. 일례로 입력단어 한국은 검색 후 등록된 단어인 대한민국을 추천단어로 매칭 후 전달한다.

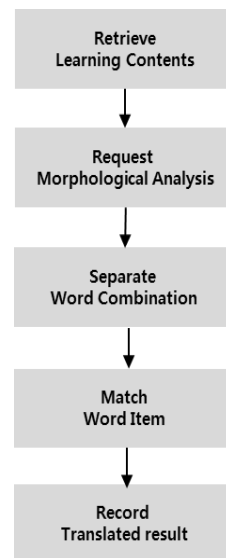


Fig. 6. The Morphological Analysis Process of Korean Text

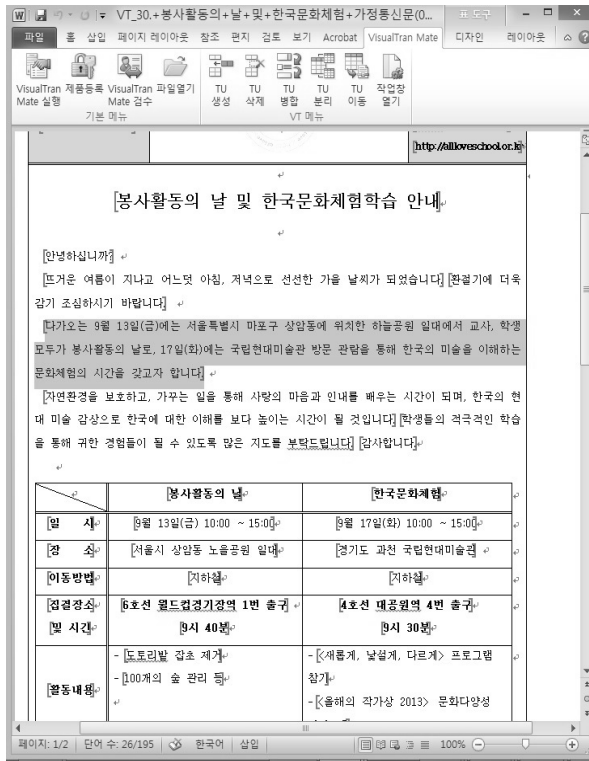


Fig. 7. Text Area Selection for Korean to Korean Translation

한한변환 시스템은 Fig. 7과 같이 클라이언트 모듈에서 변환이 필요한 문서의 일부를 선택하고 선택된 영역은 자동변환 시스템으로 전달한다. 이를 전달받은 자동변환 시스템은 불용어 및 특수문자 제거와 같은 문장 전처리를 수행하며, 형태소 분석과 구문분석을 수행한다.

클라이언트는 편집기에서 선택된 영역에 대한 한국어 텍스트 형태소 분석의 수행결과를 Fig. 8과 같이 제공한다. Fig. 8은 Fig. 7에서 선택된 영역의 텍스트인 국립현대 미술관을 포함한 5개의 한국어 명사에 대해 해당 용어의 쉬운 한국어 추천이 이루어지는 것을 확인할 수 있다. 이 과정에서 단어의 추천은 형태소 분석 결과로 얻은 파란색으로 강조된 한국어 단어는 관리/사전/문장 DB의 검색을 통해 얻는다.

4. 결론

다문화 가정과 같은 소외계층은 컴퓨터 및 인터넷 사용능력 수준인 정보이용 역량과 컴퓨터·인터넷 사용의 질적 수준인 정보격차를 줄이기 위해서는 언어장벽 해결이 필요하다. 이 논문에서는 다문화가정 학부모의 정보격차 해소를 위한 한한변환 기반 학습콘텐츠 시스템의 소프트웨어 아키텍처를 설계하고 이를 위한 사용자 맞춤형 학습콘텐츠 지원을 위해 한한변환 시스템 프로토타입을 제시하였다. 설계된 한한변환 기반 학습콘텐츠 시스템은 다문화가정 학부모의 정보격차 문제점을 해결할 수 있을 것으로 예상된다. 향후 학습자의 이해수준에 적합한 콘텐츠를 제공하는 사용자맞춤형 학습콘텐츠 지원체제의 구축할 예정이다.

References

- [1] Alexander JAM, van Deursen and Jan AGM van Dijk, "The digital divide shifts to differences in usage," *New Media & Society*, Vol.16, No.3, pp.507-526, 2014.
- [2] C. Bure, "Digital Inclusion without Social Inclusion: The Consumption of Information and Communication Technologies (ICTs) within Homeless Subculture in Scotland," *The Journal of Community Informatics*, Vol.2, No.2, pp.116-133, 2006.
- [3] P. Norris, "Digital Divide: Civic Engagement, Information Poverty, and the Internet Worldwide," Cambridge University Press, 2001.
- [4] Y. Albulut and C. S. Cardak, "Adaptive educational hypermedia accommodating learning styles: A content analysis of publications from 2000 to 2011," *Computers & Education*, Vol.58, No.2, pp.835-842, 2012.
- [5] K. Verbert et al., "Context-Aware Recommender Systems for Learning: A Survey and Future Challenges," *IEEE Transactions on Learning Technologies*, Vol.5, No.4, pp.318-335, 2012.

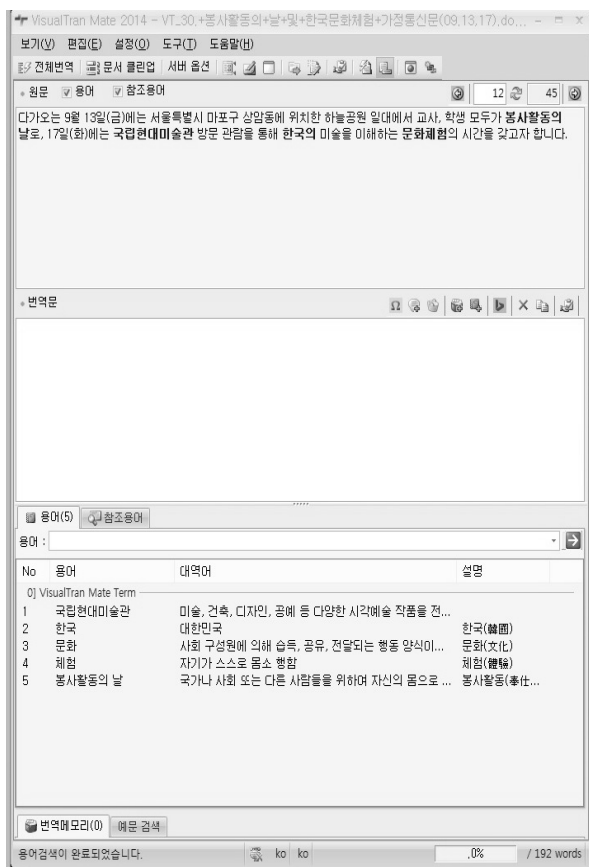


Fig. 8. The Result of Korean to Korean Translation

- [6] P. Adamopoulos, "What Makes a Great MOOC? An Interdisciplinary Analysis of Student Retention in Online Courses," in *Proceedings of the 34th International Conference on Information Systems, ICIS*, 2013.
- [7] SangHyun Jang, "A Comparative Study on the Development of K-MOOC Platform," *Journal of Platform Technology*, Vol.4, No.1, pp.33-28, 2016.
- [8] Tharindu R. Liyanagunawardena, "MOOC experience: a participant's reflection," *SIGCAS Comput. Soc.*, Vol.44, No.1, pp.9-14, 2014.
- [9] J. Dean and S. Ghemawat, "MapReduce: A Flexible Data Processing Tool," *Communications of the ACM*, Vol.53, No.1, pp.72-77, 2010.
- [10] Yun-Hee Kang, "Construction of a MapReduce Application Running on Twister in Cloud Computing Environments, FutureGrid," *Journal of KIIT*, Vol.9, No.4, pp.147-154, 2011.
- [11] Apache Hadoop, What Is Apache Hadoop? [Internet], <http://hadoop.apache.org/index.html>.
- [12] Apache Flume, Welcome to Apache Flume [Internet], <http://flume.apache.org/index.html>.



강 윤 희

e-mail : yhkang@bu.ac.kr
1989년 동국대학교 컴퓨터공학과(공학사)
1991년 동국대학교 컴퓨터공학과
(공학석사)
2002년 고려대학교 컴퓨터과학과
(이학박사)

2000년~현 재 백석대학교 정보통신학부 부교수
관심분야: 클라우드컴퓨팅, 그리드 컴퓨팅, 결합포용



강 명 주

e-mail : mjkgang@trinity.kr
1988년 동국대학교 컴퓨터공학과(공학사)
1991년 동국대학교 컴퓨터공학과
(공학석사)
1993년~1997년 제주산업대학교
전자계산과 조교수

1998년~2006년 (주)클릭큐 연구소장
2006년~2012년 자바정보기술(주) 이사
2013년~현 재 트리니티(주) 이사
관심분야: 자연어처리, 기계번역, 인공지능, 빅데이터