日本語學硏究 第65輯(2020. 9. 20. pp.131~142) http : //dx.doi.org/10.14817/jlak.2020.65.131

빅데이터 기반 다중언어 문화이미지프레임망 구축 구상

이준서(李埈瑞) 한경수(韓京洙) 노웅기(盧雄基)*

〈 Abstract 〉

A trial to construct the Cultural-Image-Frame-Network based on Big-data framework

This research is to construct the multi-lungual Cultural Image Network Element(CIFN) based on big-data framework. Our research team has already made a desktop application version Korean-Japanese CEMS at LEE & Han(2016) with a purpose of enhancing language education efficiency. Since then, we could produce several achievements which extract cultural elements from the corpus of each language. But we could find that the CEMS has several limitations i.e. 1. basically, CEMS is a desktop version application with lack of openness, 2. CEMS handles limited languages, *Korean* and *Japanese*, 3. The corpuses which CEMS depends on have only the fixed data. In this paper, we try to find out the way out to overcome the limitations which CEMS has by constructing the Cultural-Image-Frame-Network.

Field : Semantics
Keywords : CEMS(Cultural Element Mining System), CIFN(Cultural Image Frame Network), Big-data, Corpus, Cultural Element, Cultural Image

1. 들어가며

언어와 문화의 불가분성은 주지의 사실이다. 언어가 의사소통 수단, 즉 의사전달의 매개체로서의 역할뿐만 아니라 인류 문명 및 문화의 창출과 발전에 필요불가결한 존재라는 점에서, 문화가 투영되는 거울이라고도 할 수 있는 언어 속에서 각 언어권의 다양한 문화적 요소를 찾아낼 수 있다.

기존 연구에서는 거시적인 차원에서 언어와 문화의 관련성 및 언어의 보편적 기능적 측면에서 유발되는 언어 속 문화적 표출에 대한 담론이 주를 이루는 경우가 많았던 것이 사실이다. 그러나 우리의 문화는 인류 보편성에 기인한 일반성을 지니는 동시에 각 언어권의 공동체들에 노출되는 특수적 개별 환경에 따라서 다른 언어권과 차별된 상대성 및 특수성도 담지할 것으로 예측할 수 있다. 일례로 어느 문화에서나 존재하는 식사법, 인사예절 등 인류 보편적인 문화행위도 각 언어권의 문화적 상대성에 기인한 문화요소를 발견할 수 있을 것이다.

본고는 지금까지의 문화적 보편성에 기반한 담론 수준의 논의를 뛰어넘어 문화적 상대성과 특수성을 담보한 문화요소를 언어 속에서 발견해내기 위한 구체적이고 기술적인 방법의 하나로 빅데이터 기반 문

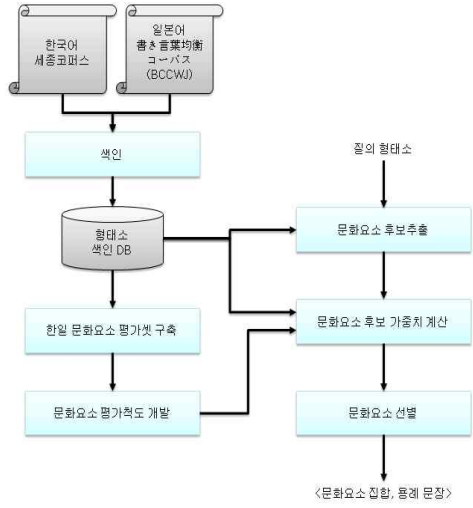
* 성결대 동아시아물류학부 교수(주저자), 성결대 컴퓨터공학부 부교수(교신저자), 가천대 소프트웨어학과 부교수(교신저자)
이 논문은 2019년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임.
(NRF-과제번호)(NRF-2019S1A5A2A03046676)

화이미지프레임망 구축 구상을 제안한다.

2. 웹 기반 다중언어 문화요소추출시스템

이준서·한경수(2016)에서는 데스크톱 애플리케이션 형태의 한·일 문화요소 추출 시스템을 개발하였다.

그림 1 2016 데스크톱 애플리케이션 한·일 문화요소 추출과정 구동화면



〈그림 2〉 2016 데스크톱 애플리케이션 한·일 문화요소추출시스템 구성

문화요소 추출을 위해서는 문화적 현상이 언어로 표현되어 있는 각국의 정제된 말뭉치로부터 각 형태소 단위의 통계를 분석해야 한다. 또한, 대용량 코퍼스를 효율적으로 처리하기 위해 코퍼스를 형태소 단위로 분리하여 색인하는 과정을 거치고, 이러한 색인 결과를 바탕으로 형태소 사이의 연관도를 계산하여 연관도 높은 형태소들을 문화요소로 추출하는 것이다.

문화요소추출시스템은 이렇게 각 언어를 대표하는 신뢰성 높은 대용량 코퍼스 데이터를 기반으로 각 언어 특유의 문화요소를 발견하고 이들의 대조 연구를 수행하기 위한 도구라고 할 수 있는데, 이미 이를 활용한 다양한 연구 성과가 만들어지고 있다(이준서, 2016 & 2017, 김혜연, 2020 등).

본 연구는 한국어와 일본어의 문화요소 추출 및 대조 연구에 활용해온 기존 데스크톱 애플리케이션 버전

의 한·일 문화요소 추출 시스템에 중국어를 추가하고, 동시에 사용자 편의성을 한층 더 증대시킨 웹 기반 한·중·일 다중언어 문화요소추출시스템을 구축하려는 것이다.

2.1 형태소 색인

추출 대상 코퍼스는 각 언어의 형태소 분석 결과가 포함된 코퍼스를 선정하였다. 한국어는 문화관광부의 '21세기 세종 계획' 사업 결과물인 세종 코퍼스(김흥규 등, 2007)를, 일본어는 일본 국립국어연구소에서 배포하는 '書き言葉均衡 코퍼스(BCCWJ)'를 사용하였다. 중국어 코퍼스로는 LDC의 Chinese Treebank 9.0을 사용하였다.

색인 단계에서는 코퍼스에 존재하는 형태소 중에서 문화요소 추출에 활용도가 낮은 형태소들은 불용어(조사, 접속사 등의 기능어-function words-)로서 제거하고, 문화요소 추출에 영향을 미칠 가능성이 높은 형태소들만 추출하여 데이터베이스에 저장한 후 인덱싱한다. 한국어 코퍼스에는 총 45가지의 품사가, 중국어에는 33가지의 품사가, 일본어에는 54가지의 품사가 존재하였으나 본 연구에서는 다음과 같은 품사의 형태소들만을 색인하였다.

한국어: 일반명사(NNG), 고유명사(NNP), 동사(VV), 형용사(VA), 일반부사(MAG), 어근(XR)
중국어: 명사(NN), 고유명사(NR), 동사(VV), 술어형용사(VA), 부사(AD), 기타 명사 수식어(JJ)
일본어: 보통명사(普通名詞), 형용동사-일반(形容動詞-一般), 형용동사-タリ(形容動詞-タリ), 부사(副詞), 동사(動詞), 형용사(形容詞), 접미사-명사적-조수사(接尾辞-名詞的-助数詞)

2.2 문화요소 추출

문화요소 추출을 희망하는 기준이 되는 형태소(질의 형태소)와 코퍼스에서 빈번히 같이 사용되는 형태소(문맥 형태소)를 문화요소 후보로 간주한다. 동일 문장에서의 공기(co-occurrence) 빈도를 기반으로 질의 형태소와 문맥 형태소 사이의 연관도를 계산하여 연관도 순으로 랭킹하여 상위 랭킹된 문맥 형태소를 문화요소로 추출한다. 연관도 계산은 연어 추출 연구에서 널리 사용되는 t-점수를 사용하였다. 질의 형태소 q와 문맥 형태소 w 사이의 t-점수 t(q,w)는 다음과 같이 계산된다(Church et al., 1991; 강범모, 2010; 이준서·한경수, 2016).

$$t(q,w) = \frac{cofreq(q,w) - \frac{freq(q)}{N} \times freq(w)}{\sqrt{cofreq(q,w)}}$$

수식 1 공기(co-occurrence) 빈도 계산법

cofreq(q,w)는 질의 형태소 q와 문맥 형태소 w가 동일 문장에서 동시에 등장한 빈도이고, freq(q)와 freq(w)는 각각 q와 w가 코퍼스에서 단독으로 등장한 빈도이다. N은 어절의 개수로 계산되는 코퍼스의 크기이다.

2.3 웹 기반 문화요소추출시스템

본 연구진은 기존에 개발된 데스크톱 애플리케이션 형태의 한·일 문화요소 추출 시스템(이준서·한경수, 2016)을 웹 애플리케이션 형태로 변환하여 편의성 높은 웹 기반의 문화요소 추출 시스템을 구현하였다. 이 시스템은 Tomcat 8.0에서 동작하는 JSP로 개발되었다.

〈그림 3〉 웹 기반 문화요소 추출 시스템(중국어 동작 예시)

한·중·일 다국어 문화요소를 추출하게 됨에 따라 사용자가 입력하는 질의 형태소에 따라 언어를 자동으로 판별한다. 입력된 형태소의 유니코드 값에 따라 언어를 판별하는데, 한자처럼 언어 구분이 모호한 경우에는 사용자가 직접 언어 선택 값을 변경할 수 있다. 자동 판별이든 수동 선택이든 이렇게 선택된 언어의 종류에 따라 질의 형태소나 문맥 형태소의 품사 리스트가 해당 언어의 품사 리스트로 자동 설정된다. 품사 리스트에서 희망하는 문맥 형태소의 품사들을 원하는 만큼 복수 개 선택하여 문화요소를 추출할 수 있다. 추출된 결과는 텍스트 파일 형태로 저장할 수 있다. 현재 버전은 한·중·일 다국어의 문화요소를 추출하는 기능 중심으로 개발되었으나, 향후 문화요소의 발견 및 대조 연구를 충분히 지원할 수 있는 편의 기능 추가와 문화요소 추출 속도 향상 등의 기능 개선이 단계적으로 진행될 예정이다.

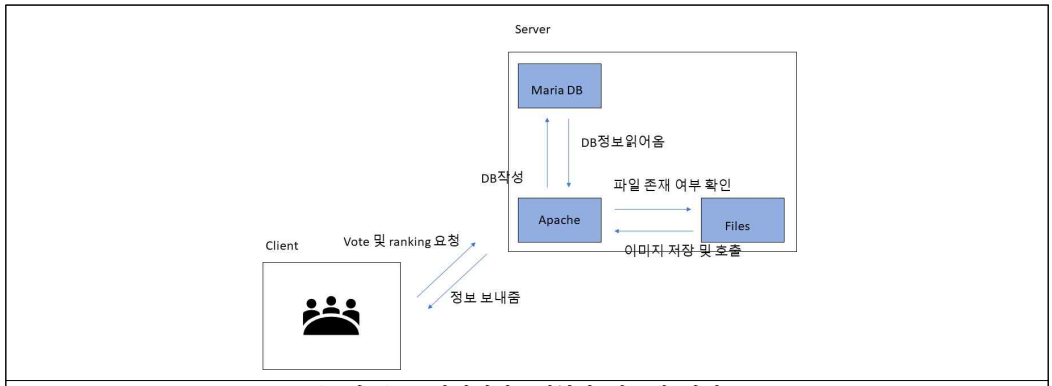
3. 빅데이터 기반 다중언어 문화이미지프레임망

한·중·일 문화요소추출시스템의 공기정보는 기존의 특정 대용량 코퍼스를 이용하여 얻어진 값이므로 언어의 동적인 변화를 발견하기 어렵다는 한계가 있다. 문화이미지프레임망은 이러한 한계를 보완하기 위하여, 네이버, Google, Yahoo 등의 검색 엔진에서 제공하는 연관 이미지 검색 서비스를 활용하여, 다수결 투표(majority voting) 방식으로 한국어, 중국어, 일본어의 국적별, 세대별, 성별 문화이미지를 선정하기 위한 플랫폼이라고 할 수 있다.

3.1 문화이미지프레임망 시스템의 전체 구조

문화이미지프레임망 시스템은 크게 클라이언트와 서버로 구분된다. 클라이언트는 인터넷을 통하여 실제 사용자가 접속하여 시스템을 이용하고 관리하기 위한 인터페이스를 제공하는 소프트웨어이고, 서버는 이미지

데이터와 클라이언트 사용자의 입력 및 관리 내용을 저장하고, 검색하는 소프트웨어 및 하드웨어이다. 서버에서는 용량이 큰 이미지 외의 모든 데이터는 데이터베이스 시스템에 저장하며, 클라이언트의 요청에 따라 원하는 데이터를 데이터베이스에서 검색하여 전송한다. 클라이언트는 인터넷 익스플로러, 크롬, 사파리와 같은 다양한 인터넷 브라우저는 물론 모바일 환경에서도 운용이 가능하도록 설계되었다.



〈그림 4〉 문화이미지프레임망 시스템 설계도

〈그림 4〉는 문화이미지프레임망 시스템의 전체 구조를 간략하게 보인 것이다. 클라이언트 측에서 인포먼트 그룹이 특정 단어에 대하여 가장 적절하다고 판단한 문화이미지를 선택하면 해당 정보가 서버 측으로 전송된다. 서버 측에서는 다국적 인포먼트 그룹의 투표로 얻어진 선택 정보를 모아서 데이터베이스에 저장한다. 이때 인포먼트 그룹으로부터 가장 많이 선택된 이미지 순으로 정렬하여 통계적 랭킹(ranking) 정보를 생성하고 데이터베이스에 저장한다. 문화이미지 파일은 데이터베이스 외부에 저장되나, 이미지에 대한 다양한 정보는 데이터베이스 내에 저장된다. 예를 들어, 이미지 파일의 이름, 이미지의 크기, 색상 공간, 연관된 단어, 이미지 입수 경로 등의 정보이다. 클라이언트 측에서 특정 단어에 대한 이미지 랭킹에 대한 정보를 요청하면, 서버 측에서는 해당 정보를 데이터베이스에서 검색하여 추출한다. 이때 랭킹에 포함된 이미지 정보도 함께 얻어지며, 그 정보를 이용하여 실제 이미지 파일을 찾는다. 이렇게 구해진 랭킹 정보와 이미지 파일을 클라이언트 측으로 전송한다.

3.2 데이터베이스 구조

term	korean	chinese	japanese	_count	super
forget.v	빼먹다,잊다	忘记	忘れる	0	Abandonment
ready.a	준비	null	null	0	Activity_ready_state
leave.v	떠나다	離開	去る	0	Abandonment
abandonment.n	포기	放棄	放棄	0	Abandonment
accuracy.n	정확성	準確性	準確性	0	Accuracy
addict.n	중독자	瘾君子	中毒者	0	Addiction
abandon.v	null	扔掉	捨てる	0	Abandonment
breakfast.v	아침,조식	早餐	朝ご飯,朝食	0	Ingestion
ingestion.n	젓가락	筷子	お箸	0	Ingestion

〈그림 5〉 데이터베이스 구조 예시 1

term	_count	nationality	origin	_date	image	super
忘れる	1	japanese	forget.v	2020-04-10 11:05:07	0	Abandonment
忘记	2	chinese	forget.v	2020-05-14 13:42:31	0	Abandonment
준비	0	korean	ready.a	0000-00-00 00:00:00	0	Activity_ready_state
떠나다	0	korean	leave.v	0000-00-00 00:00:00	0	Abandonment
포기	0	korean	abandonment.n	0000-00-00 00:00:00	0	Abandonment
離開	0	chinese	leave.v	0000-00-00 00:00:00	0	Abandonment
去る	0	japanese	leave.v	0000-00-00 00:00:00	0	Abandonment
정확성	0	korean	accuracy.n	0000-00-00 00:00:00	0	Accuracy
準確性	0	chinese	accuracy.n	0000-00-00 00:00:00	0	Accuracy
準確性	0	japanese	accuracy.n	0000-00-00 00:00:00	0	Accuracy
중독자	0	korean	addict.n	0000-00-00 00:00:00	0	Addiction
瘾君子	1	chinese	addict.n	2020-04-21 21:45:42	0	Addiction
中毒者	0	japanese	addict.n	0000-00-00 00:00:00	0	Addiction
扔掉	1	chinese	abandon.v	2020-04-22 00:06:45	0	Abandonment
捨てる	1	japanese	abandon.v	2020-04-22 00:07:03	0	Abandonment
잊다	0	korean	forget.v	0000-00-00 00:00:00	0	Abandonment
아침	15	korean	breakfast.v	2020-05-13 21:11:52	0	Ingestion
조식	0	korean	breakfast.v	0000-00-00 00:00:00	0	Ingestion
早餐	0	chinese	breakfast.v	0000-00-00 00:00:00	0	Ingestion
朝ご飯	1	japanese	breakfast.v	2020-05-05 10:22:33	0	Ingestion
朝食	0	japanese	breakfast.v	0000-00-00 00:00:00	0	Ingestion
젓가락	4	korean	ingestion.n	2020-05-14 12:01:25	0	Ingestion
筷子	3	chinese	ingestion.n	2020-05-08 07:44:28	0	Ingestion
お箸	3	japanese	ingestion.n	2020-05-08 07:43:38	0	Ingestion

〈그림 6〉 데이터베이스 구조 예시 2

문화이미지프레임망 시스템에서 사용하는 데이터베이스 내의 대표적인 테이블은 매칭 테이블, 단어 테이블, 투표 테이블이 있다. 〈그림 5〉는 매칭 테이블의 일부를 보인 것이다. 이 테이블에서 term(Lexical Unit 단어 및 품사), korean(한국어 단어), chinese(중국어 단어), japanese(일본어 단어) 컬럼은 매칭되는 단어들을 저장한 것이며, super 컬럼은 Frame Index에서의 해당 단어의 상위 단어를 나타낸다. _count 컬럼은 해당 단어에 대한 검색 횟수를 저장하기 위한 것이다.

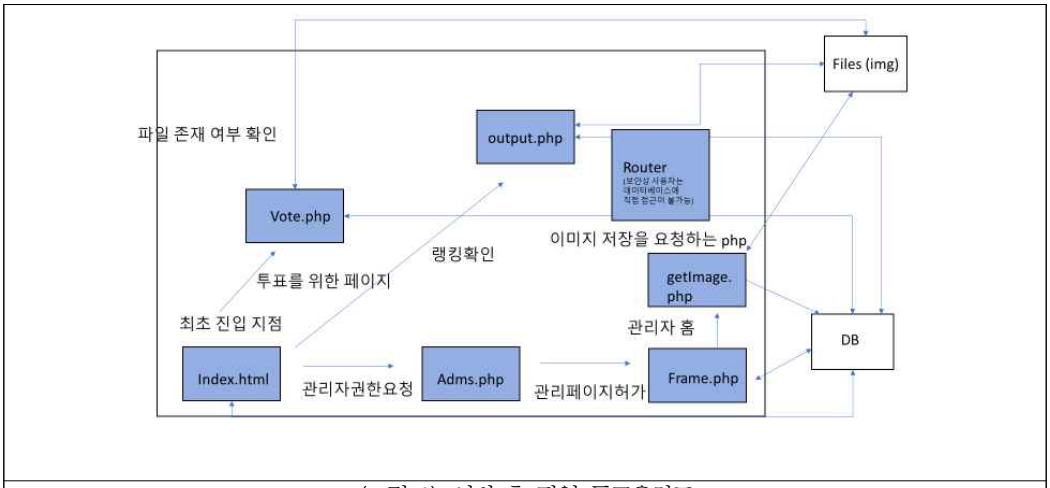
〈그림 6〉은 시스템 구축 초기 단계의 단어 테이블의 일부를 보인 것이다. 이 테이블에 저장된 단어는 모두 매칭 테이블의 korean, chinese, japanese 컬럼에 포함되어 있는 단어만 나타날 수 있다. 각 단어에 대하여 _count 컬럼은 해당 단어가 검색된 횟수이며, nationality 컬럼은 한국, 중국, 일본 중의 하나의 값을 가진다. origin 컬럼은 해당 단어에 대한 Lexical Unit, 즉 위의 매칭 테이블에서의 term 컬럼에 저장된 값이다. _data 컬럼은 해당 단어에 대한 정보가 최후로 수정된 시간을 나타내고, super 컬럼은 매칭 테이블에서의 super 컬럼과 동일하다.

투표 테이블은 단어 테이블에 저장된 각 단어에 대하여 하나씩 존재한다. 이때 테이블의 이름은 해당 단어가 된다. 예를 들어, '아침'이라는 단어에 대한 투표 테이블은 '아침'이라는 이름의 테이블이다. 투표 테이블은 index 컬럼과 male0, male1, ..., male6, female0, female1, ..., female6 컬럼으로 구성된다. index 컬럼은 해당 단어에 대한 이미지에 대한 인덱스이며, 하나의 인덱스는 하나의 이미지에 대응된다. 'male0, male1, ..., male6' 컬럼은 각각 특정 인덱스의 이미지를 선택한 10대~60대 남성 사용자의 수를 저장한다. 마찬가지로 'female0, female1, ..., female6' 컬럼은 해당 이미지를 선택한 10대~60대 여성 인포먼트 그룹의 수를 저장한다.

〈그림 7〉 데이터베이스 구조

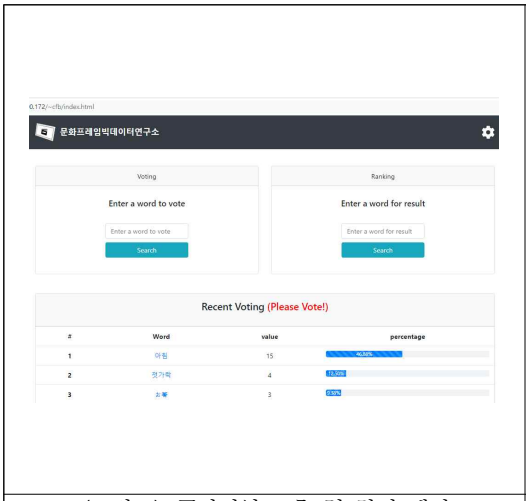
〈그림 7〉는 '아침' 단어에 대한 투표 테이블이다. 이 단어와 연관된 이미지는 각각 index 컬럼 내의 '0, 1, ... , 24' 인덱스 값에 대응되는 25개의 이미지가 존재한다. 각각의 이미지에 대하여 몇 명의 인포먼트 그룹이 선택하였는지가 저장되어 있다. 예를 들어, 시스템 구축 초기 단계에서는 5번 이미지에 대하여 10대 남녀 중에는 선택한 사람이 없고(즉, male0 = female0 = 0), 20대 남자 중에는 4명, 20대 여자 중에는 0명, 30대 남자 중에는 1명, 30대 여자 중에는 0명이 각각 선택한 것을 알 수 있다.

3.3 문화이미지프레임망 시스템 서버측 작업

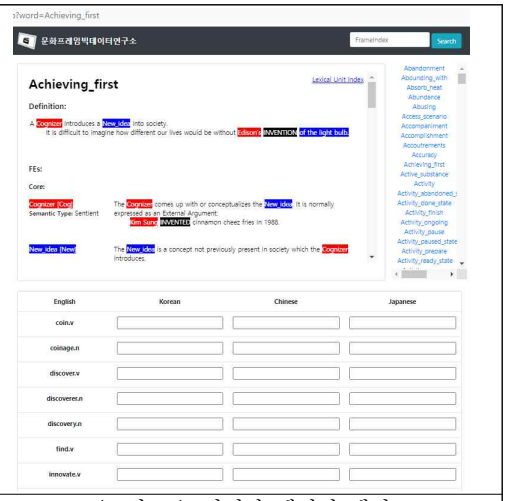


〈그림 8〉 서버 측 작업 플로우차트

〈그림 8〉은 서버 측에서 수행되는 작업 프로세스를 나타낸 것이다. 여기에서 index.html 문서는 클라이언트에서 서버를 이용하기 위하여 최초로 접속하는 웹 페이지이다. 모발일 환경에도 대응하는 이 페이지 내에서 사용자는 문화이미지 투표, 랭킹 확인, 관리자 페이지의 세 가지 중에서 한 가지를 선택할 수 있다.

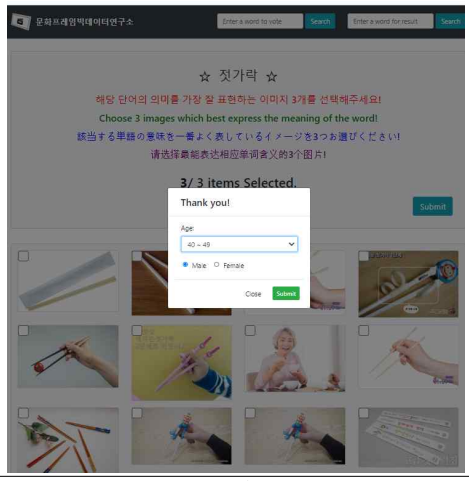


〈그림 9〉 클라이언트 측 첫 화면 예시

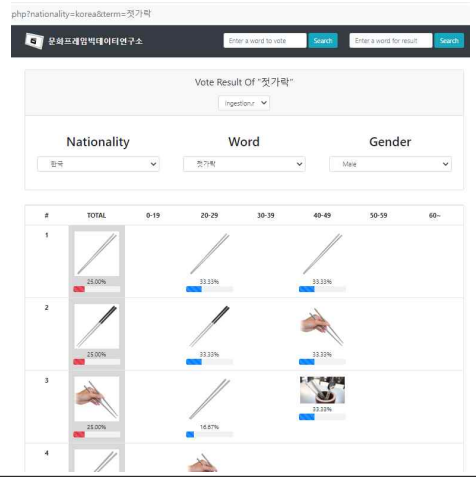


〈그림 10〉 관리자 페이지 예시

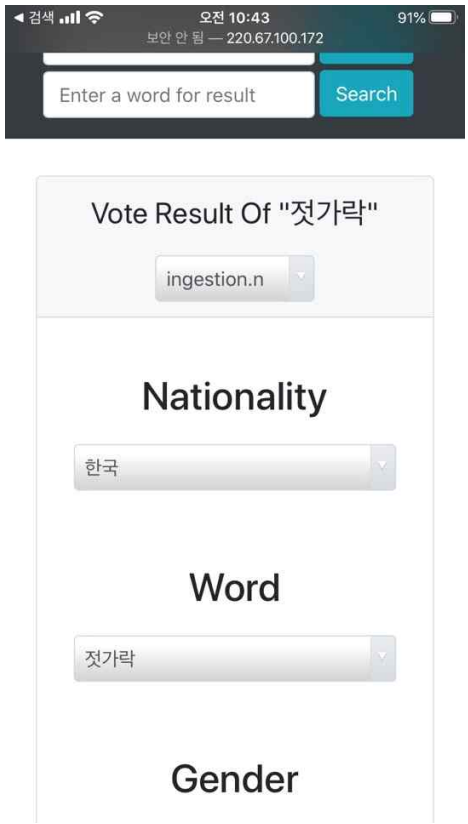
특정 단어에 대한 문화이미지 투표를 선택하면 해당 단어가 한국어, 중국어, 일본어의 3개 국어로 연동된 투표 인터페이스가 연관 이미지들과 함께 출력된다(〈그림 11〉). 이 이미지들은 구글 이미지 검색서비스에서 제공하는 연관 이미지 검색 결과물로 해당 언어의 단어 입력으로 얻어진 값들로 문화이미지 후보군이 된다. 인포먼트 그룹은 이 문화이미지 후보군 중에서 해당 언어에 대응하는 단어를 해당 언어 문화권에서 문화적으로 가장 잘 표현한다고 판단하는 이미지 3개를 선택한다. 이러한 과정은 vote.php 프로그램을 통하여 진행된다.



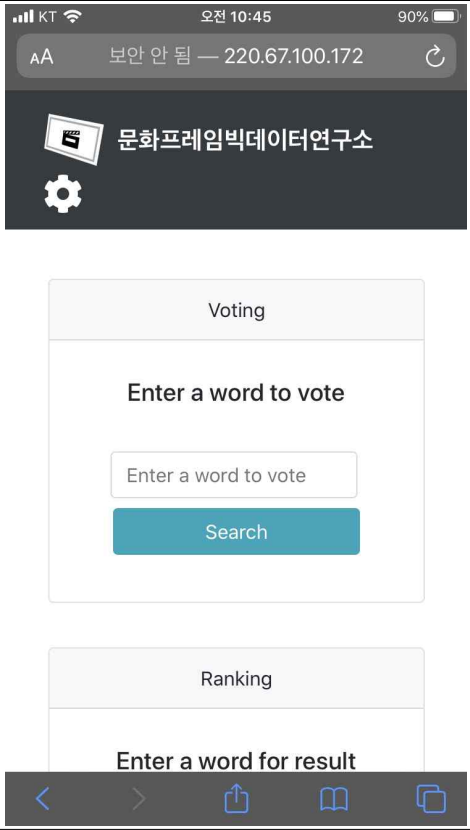
〈그림 11〉 투표용 페이지 예시



〈그림 12〉 랭킹 결과 페이지 예시



〈그림 13〉 모바일 화면 예시 1



〈그림 14〉 모바일 화면 예시 2

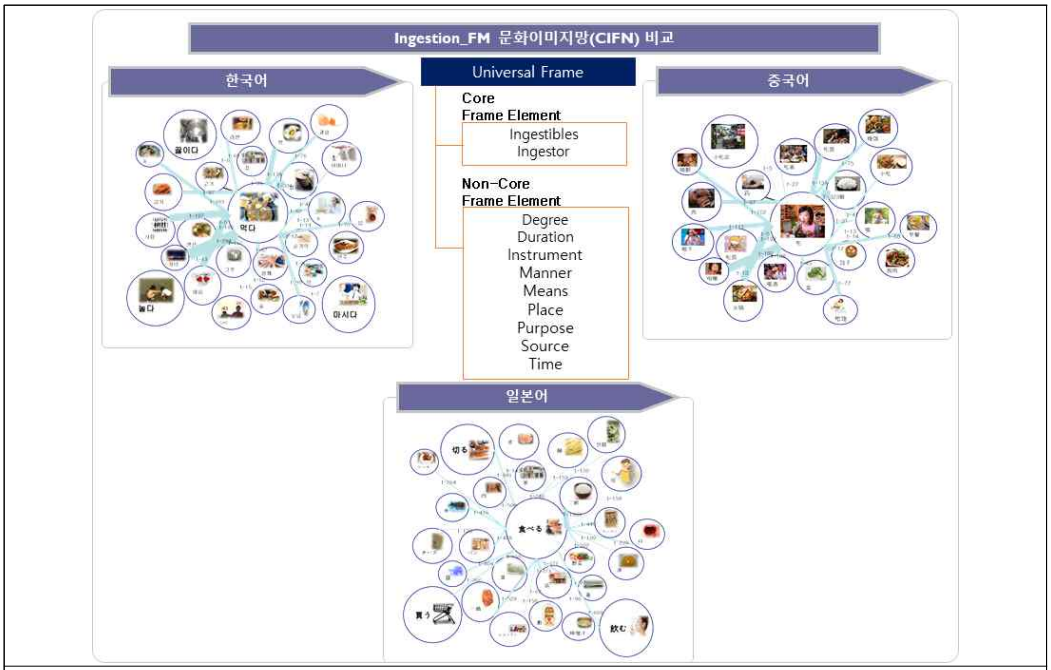
vote.php 프로그램은 필요에 따라서 언제든지 생성 가능하며, 클라이언트 측의 인터넷 브라우저에서 출력 가능한 웹 페이지로 얻어진다. 이러한 웹 페이지는 파일 또는 데이터베이스에 저장되는 것이 아니라, 필요에

따라 수시로 만들어져서 클라이언트 측으로 전송된다. 이때 만들어진 웹 페이지는 클라이언트 측으로의 전송이 완료되면 바로 삭제된다.

투표에 참여하는 다양한 인포먼트 그룹(성별, 세대별, 국가별)에 의해 생성되는 동일 단어에 대한 문화이미지 랭킹의 출력은 output.php 프로그램이 실행한다. 이 프로그램은 위에서의 vote.php 프로그램과 마찬가지로 클라이언트 인포먼트 그룹들의 입력값(한·중·일 단어)에 따라 해당 단어의 성별, 세대별, 국가별 문화 이미지 정보를 데이터베이스로부터 추출하고 해당 단어의 문화이미지 파일을 찾아서 하나의 웹 페이지로 통합하여 클라이언트 측에 전송한다〈그림 12〉.

vote.php와 output.php 프로그램은 모두 데이터베이스를 액세스한다. 즉, vote.php는 성별, 국가별, 세대별로 구분된 각 언어권의 인포먼트 그룹이 특정 단어에 대하여 어떤 문화 이미지를 선택하였는가를 데이터베이스에 저장하고, output.php에서는 이렇게 선택한 이미지들의 랭킹 정보를 데이터베이스로부터 읽어 온다. 이때, 두 php 프로그램은 관리자가 아닌 투표에 참여하는 인포먼트 그룹을 포함한 일반 사용자를 위한 것이므로, 해킹 등 보안 상의 문제로 인하여 데이터베이스를 직접 액세스 불가능하고 간접 액세스를 지원하는 router 모듈을 이용한다. 또한, 데이터베이스 외의 이미지 파일을 직접 액세스 가능하다. 이 파일들은 주기적으로 백업되어, 해킹 또는 하드웨어 오작동 등 예기치 못한 돌발 상황에서도 복구가 용이하다.

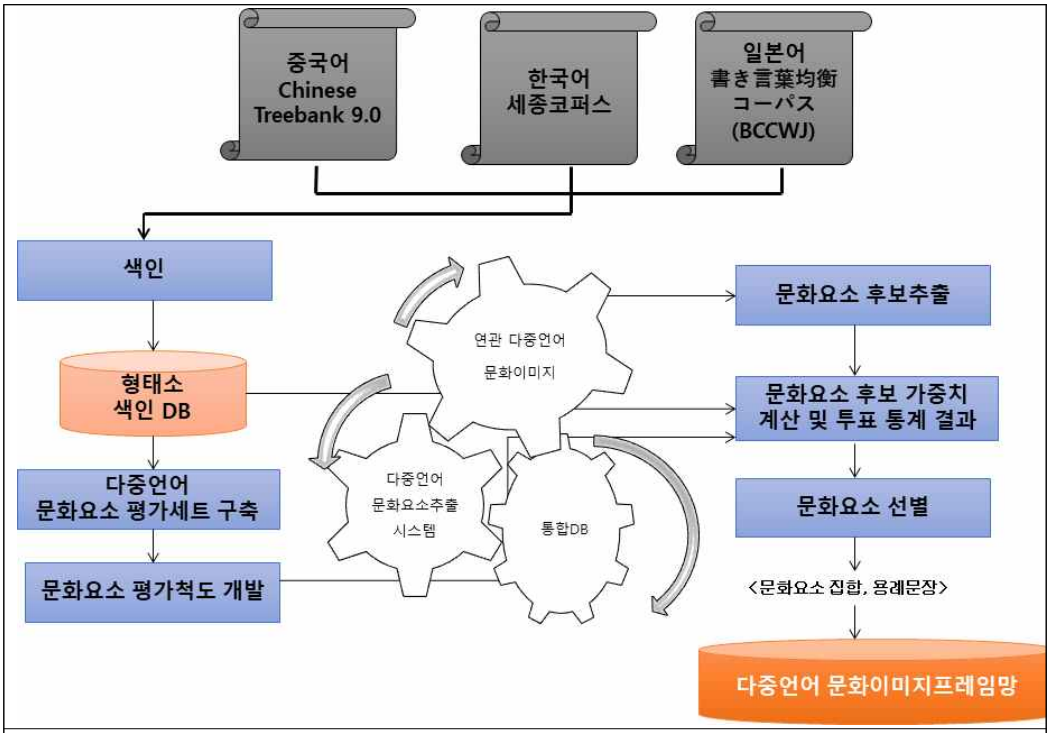
최초의 index.html 페이지에서 관리자 권한을 요청하면 adms.php 프로그램이 실행된다. 이 프로그램에서는 패스워드 등 관리자 인증을 위한 정보를 요청하며, 관리자 인증에 성공하면 frame.php 프로그램이 실행된다. 이 프로그램에서는 각 언어권의 인포먼트 그룹에 의하여 얻어진 한·중·일 단어-이미지 정보에 대한 통계 데이터 등을 볼 수 있으며, 투표 시스템 관리를 위한 다양한 설정이 가능하다〈그림 10〉. 또한, 데이터베이스를 직접 액세스하고 데이터베이스 관리(백업, 투표관리, 문화이미지 관리 등)에 대한 사용자 편이 기능을 갖춘 다양한 작업의 수행이 가능하다.



〈그림 15〉 문화이미지프레임망 시각화(visualization) 예시

마지막으로, 각 문화권에 대응하는 연관 이미지를 기반으로 하는 문화이미지 후보군과 연동된 국적별, 세대별, 성별 다수결 투표(majority voting)로 축적되는 빅데이터는 향후 〈그림 15〉의 예시와 같이 다중언어 문화이미지프레임망 구현을 위한 다양한 시각화(visualization) 작업이 점진적으로 이루어질 예정이다.

4. 나가며



〈그림 16〉 문화요소추출 및 문화이미지프레임망 시스템 연동 구조

웹기반 문화요소추출시스템은 정제된 말뭉치(코퍼스) 빅데이터를 기반으로 각 언어의 형태소 단위의 통계적 분석을 통해 각 언어문화권 특유의 문화요소를 발견하고 이들의 대조 연구를 수행하기 위한 시스템으로 본 연구를 통하여 필요에 따라서 얼마든지 다국어로 확장될 수 있는 가능성을 보여준 것이라고 할 수 있다. 또한, 문화요소추출시스템의 공기정보가 기존의 특정 대용량 코퍼스를 이용하여 얻어진 값이므로 언어의 동적인 변화를 발견하기 어렵다는 한계성을 지니는데, 본 연구를 통하여 이를 극복할 수 있는 방안을 제시한 것이다. 즉, 문화요소추출시스템과 연동된 연관 이미지 검색 서비스를 활용하여, 다수결 투표(majority voting) 방식으로 한국어, 중국어, 일본어의 국적별, 세대별, 성별 문화이미지프레임을 선정할 수 있는 플랫폼인 본 연구의 문화이미지프레임망은 문화요소추출시스템의 단점을 보완하며, 동시에 무한한 다중언어 인포먼트 그룹의 통계 데이터의 축적을 통하여 기존의 문화적 보편성에 기반한 담론 수준의 논의를 뛰어넘어 문화적 상대성과 특수성을 담보한 문화요소를 언어 속에서 발견해낼 수 있는 구체적이고 기술적인 방법을 제시

한 것이다.

향후 이용자의 편의성을 증대시키기 위한 컴스터마이징, 시스템 안정화를 위한 하드웨어 증축 및 다국어 인포먼트그룹 확보 등의 문제가 남아있지만, 본 연구의 빅데이터 기반 다중언어 문화이미지프레임망 구축 구상을 통하여 유관 연구자들의 점진적인 공감대 형성의 계기가 될 것으로 기대한다.

【참고문헌】

강범모(2010), 「공기 명사에 기초한 의미/개념 연관성의 네트워크 구성」『한국어의미학』 32, pp.1-28.

김흥규, 강범모, 홍정하(2007) 「21세기 세종계획 현대국어 기초말뭉치: 성과와 전망」『제19회 한글 및 한국어 정보처리 학술대회 논문집』한국정보과학회 언어공학연구회 pp.311-316.

이준서·한경수(2011) 「일본어교육용 이미지 검색엔진 구축」『일본어교육연구』20 한국일어교육학회 pp. 159-169.

이준서·한경수(2016) 「다국어 '문화요소추출시스템(CEMS)' 개발 구상」『일본어교육연구』20 한국일어교육학회 pp. 289-304.

이준서(2016) 「'sleep' 프레임 동사의 한·일 문화요소 비교」『일본어학연구』49 한국일본어학회 pp. 79-90

이준서(2017) 「'Visit_host' 프레임의 일본어 문화이미지프레임(CIF) 구성에 관한 일고찰」『일어일문연구』101 한국일어일문학회 pp. 91-106.

김혜연(2020) 「공기하는 의미 요소의 특징에서 본「목욕」의 한일 이미지 대조연구* - 텍스트마이닝 결과를 바탕으로 -」『일본어학연구』63 한국일본어학회 pp. 153-167

Church, K., W. Gale, P. Hanks, and D. Hindle (1991), "Using Statistics in Lexical Analysis", in U. Zernik (ed.), Lexcial Acquisition: Exploiting on-line resources to build a lexicon. Hilldale: Lawrence Erlbaum, pp.115-164

Fillmore C(1975) An Alternative to Checklist Theories of Meaning", Proceedings of the First Annual Meeting of the Berkeley Linguistics Society, pp. 123-131

Fillmore C(1977) Scenes-and-frames semantics, Linguistic Structures Processing, North-Holland, Amsterdam, pp. 55-81

Fillmore C(1982) Frame Semantics. In Linguistic Society of Korea (ed.), Linguistics in the Morning Calm. Seoul, Hanshin, pp.111-138

Fillmore C, Atkins B(1992) Toward a frame-based lexicon: The semantics of RISK and its neighbors. In Lehrer A, Kittay E (eds), Frame, fields, and contrasts: New essays in semantic and lexical organization. Hillsdale, Erlbaum, pp. 75-102

Fillmore C, Atkins B(2000) Describing Polysemy: The Case of 'Crawl.' In Ravin Y, Laecock C (eds), Polysemy. Oxford, Oxford University Press, pp. 91-110

Chinese Treebank 9.0, https://catalog.ldc.upenn.edu/LDC2016T13

〈 要 旨 〉

ビッグデータ基盤多重言語文化イメージフレーム網構築構想

本硏究はビッグデータ基盤多重言語文化イメージフレーム網を構築するためのものである。我々硏究チームは李&韓 (2016) で言語取得の効率を高めるためのツールとしてデスクトップバージョンアプリケーションである文化要素抽出システム(CEMS)を作り上げたことがある。それによって、日本語と韓国語から様々な文化要素を見いだすことができ、いくつかの成果があげられたのである。しかし、文化要素抽出システム(CEMS)にはいくつかものかの限界が見つかる。1．文化要素抽出システム(CEMS)は基本的に日本語と韓国語に限られている。2．文化要素抽出システム(CEMS)はデスクトップバージョンアプリケーションであるため、限られた環境でしか使えない。3．文化要素抽出システム(CEMS)は特定のコーパスにしか対応されていない。一般公開を前提に開発された本硏究のビッグデータ基盤多重言語文化イメージフレーム網はこのような限界を克服できるものと期待できる。また、中国語にも対応でき、より多くの多重言語に広がる可能性も大いに期待できるのである。

論文分野：意味論
キーワード：文化要素抽出システム（CEMS）、文化イメージフレームネットワーク(CIFN)、ビッグデータ(Big-data)、コーパス、文化要素、文化イメージ

▣ 이준서(李埈瑞), Lee JUNSEO) 성결대 동아시아물류학부 교수
한경수(韓京洙, Han, Kyoungsoo) 성결대 부교수
노웅기(盧雄基, Roh, woonggi) 가천대 부교수
jslee@sungkyul.ac.kr

■投稿日 : 2020년 6월 24일
■審査開始 : 2020년 7월 20일
■審査完了 : 2020년 8월 2일
■掲載確定 : 2020년 8월 21일