

# 소프트웨어 정의 네트워크를 이용한 한중일 한자-한국어 변환 키워드 도메인 이름 시스템

이승훈<sup>○\*</sup>, 조승철<sup>\*\*</sup>, 薛園園<sup>\*\*\*</sup>, 路凯<sup>\*\*\*\*</sup>, 向天歌<sup>\*\*\*\*</sup>, 한선영<sup>†\*\*\*\*</sup>

건국대학교 스마트ICT융합학과\*, 건국대학교 상허교양대학\*\*, 건국대학교 중한비교어문학과\*\*\*, 건국대학교 컴퓨터 공학과\*\*\*\*  
asia\_a@cclab.konkuk.ac.kr, cschol@cclab.konkuk.ac.kr, erya0905@gmail.com, otkabsol@naver.com,

singskyx@gmail.com, syhan@cclab.konkuk.ac.kr

## CJK Chinese Character-Korean Character Conversion Keyword Domain Name System in Software Defined Network

SeungHun Lee<sup>○\*</sup>, SungChol Cho<sup>\*\*</sup>, Yuanyuan Xue<sup>\*\*\*</sup>, Kai Lu<sup>\*\*\*\*</sup>, Tiange Xiang<sup>\*\*\*\*</sup>, Sunyoung Han<sup>†\*\*\*\*</sup>

Department of Smart ICT\*, SangHuh College\*\*, Dept. of Chinese-Korean Comparative Literature and  
Linguistic\*\*\*, Department of Computer Science&Engineering\*\*\*\*, Konkuk University

### 요 약

본 논문에서는 소프트웨어 정의 네트워크를 이용한 한중일 한자-한국어 변환 키워드 도메인 이름 시스템을 제안하였다. 한자 체계를 주로 사용하는 한국, 중국, 일본에서 세 나라의 한자 수량이 너무 많기 때문에 우선 한국, 중국, 일본이 공용으로 사용하는 한자 체계인 CJK808을 가지고 연구하였다. 연구를 통해 CJK808 한자 체계에서 각 나라의 한자 특징도 많이 발견하였고, 그 중에서 표준자와 이체자의 다양성이 더욱 두드러졌다. SDN을 이용함으로써 관리 측면에서 다양한 이점을 얻을 수 있다. 제안하는 시스템을 통하여 사용자들은 한국, 중국, 일본 한자를 입력하면 SDN에서 관리하는 도메인 네임 서버를 통해 IP 주소를 얻을 수 있다.

주제어: 소프트웨어 정의 네트워크, 한중일 한자, 도메인 네임 시스템

### 1. 서론

현재 스마트폰으로부터 대규모 웹 사이트의 콘텐츠를 서비스하는 서버에 이르기까지 인터넷상의 모든 컴퓨터는 IP주소를 이용하여 서로 통신한다. 컴퓨터 통신을 위하여 IP주소를 모두 기억하기에는 어려움이 있으며 이를 해결하기 위해서 도메인 네임 시스템(Domain Name System, DNS)이 개발되었다. DNS는 호스트의 도메인 이름을 호스트의 IP 주소로 바꾸거나 그 반대의 변환을 수행할 수 있다. 기본적으로 영문 도메인 네임 시스템은 존재하나 한자 문화권에서는 같은 뜻에 서로 다른 모양을 갖는 이체자까지 포함하기 때문에 한자 도메인 서버에 이를 같은 글자로 처리해주어야 한다. 기존의 DNS에 한자 키워드 기반의 한자 도메인 시스템을 도입하기 위해서는 모든 컴퓨터에 서비스하는 보조 DNS 주소를 입력해야 하는 어려움이 있다. 하지만 소프트웨어 정의 네트워킹(Software Defined Networking, SDN)과 같은 중앙 집중형 네트워크는 SDN 컨트롤러에 DNS 주소를 한 번만 입력하면 SDN의 모든 호스트에게 적용이 되기 때문에 한자 DNS 서버를 용이하게 찾을 수 있다. 한자 DNS를 SDN에 적용할 경우 다음과 같은 장점이 있다. 네트워크 프로그래밍 측면에서는 CLI, API를 통해 관리 플랫폼을 사용하여 네트워크 생성 및 제거, 관리 작업을 하기 때문에 한자 DNS 서비스를 자동화할 수 있다.

본 연구에서는 소프트웨어 정의 네트워크를 이용한 한중일 한자-한국어 변환 키워드 도메인 이름 시스템을 제안하였다. 한자 체계를 주로 사용하는 한국, 중국, 일본

에서 세 나라의 한자 수량이 너무 많기 때문에 우선 한국, 중국, 일본이 공용으로 사용하는 한자 체계인 CJK808을 가지고 연구했다.

### 2. 관련 연구

#### 2.1 표준자와 이체자

표준자 즉 상용한자는 한자문화권에 속하는 각 국가에서 규정하여 일상생활에서 쓰도록 권장하는 한자의 모음을 말한다. 상용한자는 일반인들이 다 외우기 어려운 한자의 방대한 자수를 고려하여 한자 교육의 편의 및 일상에서 쓰이는 한자의 제한과 정자로의 통일을 위해 고안되었다. 중국은 간화자, 일본은 신자체, 한국, 대만, 홍콩 및 마카오에서의 한자는 정체자를 참고하였다.

이체자는 표준자의 상대적인 개념이다. 중국 본토, 대만지역, 홍콩, 한국 및 일본에서 각 나라에서 정해진 표준 한자 이외에 있는 한자를 모두 이체자라고 한다. 구체적으로 말하면 한 표준자를 정하면 나머지 같은 뜻과 발음을 가진 여러 가지 글자를 모두 이체자라고 한다.



[그림 1] 표준자와 이체자

유니코드 한자 체계에서 중국 본토, 대만지역, 홍콩,

한국, 일본에서 유니코드 값이 같은 한자를 포함하고 유니코드 한자 체계에서 말하는 이체자는 표준 한자와 상대적으로 유니코드 값은 다르지만, 의미와 발음이 같은 글자를 가리킨다.

### 2.2 CJK808

CJK808[6]은 중국, 일본, 한국에서 흔히 쓰이는 한자의 모음이다. 808자 중에서 801자는 중국의 현대 중국 공통 문자에 속해있고, 710자는 일본의 문자교육에 속한다. 그리고 808자 모두는 한국의 중고등학교 한자 시간에 속한다. 이 중 약 540글자가 같은 한자이다. 200여 글자는 구분에 있어서 차이점을 가지고 있다.

공통으로 사용되는 한자의 도표를 만들자는 생각은 2005년 포럼에서 중국 대표가 제안해 일본, 한국과 공동으로 진행하였다. 중국인민대학은 2010년 한중일 3국의 문화교류와 공통발전을 촉진하기 위한 콜롬 문자표를 제안하였고, 3국에서 6년간의 심의 끝에 목록 및 초안을 작성하였다. 그리고 2014년 CJK808자가 정식으로 승인되었다.

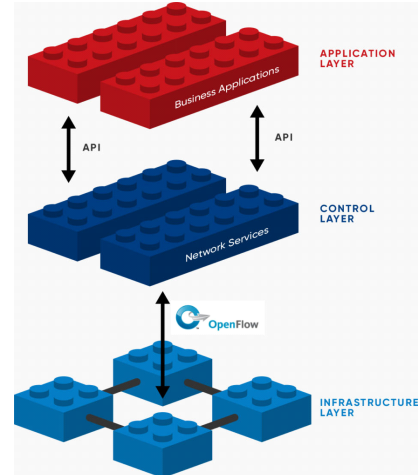
이를 사용함으로써 얻는 이점은 다른 국가의 언어들을 알지 못하더라도 표에 있는 한자를 통해서 다른 국가 사람들간에 간단한 의사소통을 할 수 있어 많은 편의를 제공한다.

### 2.3 소프트웨어 정의 네트워킹(SDN)

소프트웨어 정의 네트워킹(Software Defined Networking, SDN)[7]은 개방형 API를 통해 네트워크의 트래픽 전달 동작을 소프트웨어 기반 컨트롤러에서 제어/관리하는 접근 방식이다. 트래픽 경로를 지정하는 컨트롤 영역과 트래픽 전송을 수행하는 데이터 영역이 분리되어 있다. SDN의 동작 원리는 SDN 컨트롤러(제어부)와 SDN 전송 장비(전송부), 컨트롤러와 전송 장비와의 통신 규약(RPC)으로 이루어져 있다.

SDN 컨트롤러는 여러 네트워크 장비와 통신할 수 있도록 API(South-bound API)를 제공 및 추가할 수 있습니다. 또한 여러 가지 기능의 애플리케이션을 개발하고 다른 운영 도구와 통신할 수 있게 해주는 API(North-bound API)도 제공합니다. SDN 전송 장비의 경우 화이트박스 스위치의 등장으로 새로운 변화를 가져왔는데 기존 레거시 장비의 제어 방식이 공개하지 않는 것과 반대로 네트워크 장비의 동작 방식을 사용자가 결정하고 투명하게 공개되어 있다는 의미로 지정되었다.

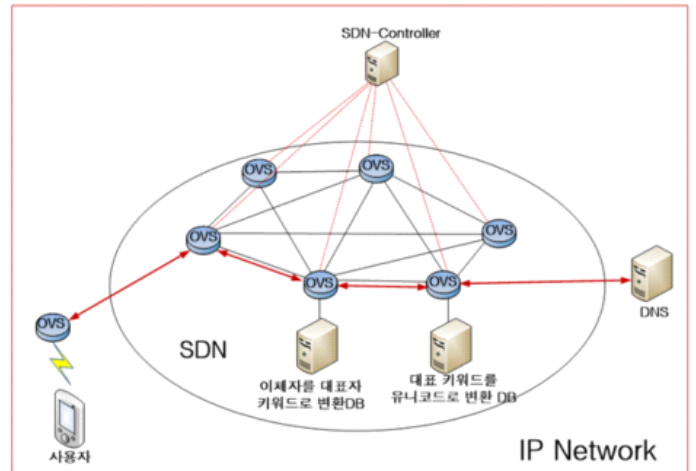
SDN에서 제어부와 전송부와의 표준 통신 구격이 필요한데 이 중 하나가 오픈플로우(OpenFlow)이다. 오픈플로우는 SDN 환경에서 제어 계층과 데이터 계층 간의 통신 프로토콜로 컨트롤러와 하드웨어 장비 간의 통신 내용을 정의하는 SDN의 기반이 되는 기술로써, 데이터 경로를 설정함에 따라 네트워크 장비의 프로그램 가능성을 제공하여 트래픽을 어떤 네트워크 장비 속으로 흐르게 할 것인지 정의할 수 있다. 구성으로는 전송부를 제어하기 위한 명령어 및 네트워크 장비의 정보를 수집하기 위한 명령어로 이루어져 있다.



[그림 2] 소프트웨어 정의 네트워킹

### 3. 한자 키워드 도메인 이름 시스템

본 연구에서 제안하는 시스템 전체 구성도는 [그림 3]과 같다.



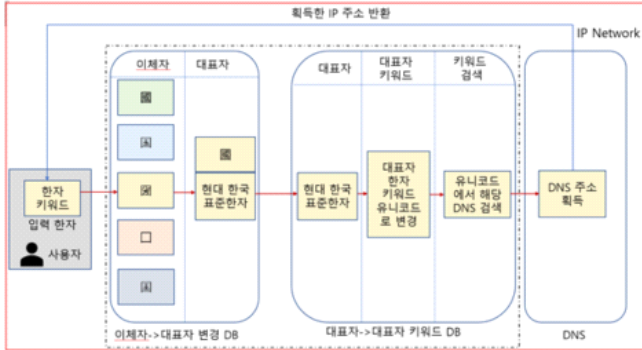
[그림 3] 시스템 전체 구성도

시스템을 구성하는 요소는 SDN Network를 관리하는 SDN 컨트롤러와 가상의 스위치 역할을 하는 SDN Switch, 사용자들에게 서비스를 제공하기 위해 구축하는 서버, 사용자가 서비스를 이용할 수 있는 역할인 호스트가 있다.

SDN 컨트롤러로는 대표적으로 사용되는 ONOS를 사용하였고, SDN Switch로는 실제 하드웨어 스위치 대신 가상의 하드웨어 스위치 역할을 할 수 있는 Open vSwitch (OvS)를 사용하였다. 서버는 서비스를 제공하기 위해서 이체자를 대표자 키워드로 변환하는 DB서버와 대표 키워드를 유니코드로 변환하는 DB서버가 있고, 사용자들은 PC, Mobile 기기 등을 이용하여 SDN에 접속하여 서비스를 이용한다.

SDN을 이용함으로써 관리자는 도메인 네임 서버를 사용자 별로 할당하지 않고, 중앙에서 전체 호스트에게 통과하는 도메인 네임 서버를 관리할 수 있고, 운영하고자 하는 서비스 또한 손쉽게 관리할 수 있게 됨에 따라 전체 네트워크 토폴로지 관리도 용이하다. 사용자의 입장

에서는 별다른 소프트웨어나 하드웨어 없이 SDN에 컨트롤러에서 제공하는 서비스 및 애플리케이션을 SDN에 속한 호스트에서 이용할 수 있다.



[그림 4] 한자 키워드 도메인 이름 시스템 구성도

[그림 4]는 사용자의 입력에서부터 DNS 주소를 통해 IP 주소를 얻는 과정이다. 사용자는 한자 키워드를 입력하고, 이체자의 경우 현대 한국 표준 한자로 변환한다. 이후 현대 표준 한자에 해당하는 유니코드 값을 도메인 네임 키워드 데이터에서 해당하는 도메인 네임을 가져오고, SDN 컨트롤러에서 정의한 DNS서버에 전달하여 해당하는 IP 주소를 받아온다.

[그림 4]에서 설명한 과정에서 이체자를 대표자로 변환하기 위해 [그림 5]와 같이 CJK808 유니코드 값을 표로 정리하였다.

한국	유니코드	중국본토	유니코드	일본	유니코드
1 價	U+50F9	价	U+4EF7	価	U+4FAI
2 街	U+8857	街	U+8857	街	U+8857
3 假	U+5047	假	U+5047	仮	U+4EEE
4 歌	U+6B4C	歌	U+6B4C	歌	U+6B4C
5 加	U+52A0	加	U+52A0	加	U+52A0
6 可	U+53EF	可	U+53EF	可	U+53EF
7 家	U+5BB6	家	U+5BB6	家	U+5BB6
8 各	U+5404	各	U+5404	各	U+5404
9 角	U+89D2	角	U+89D2	角	U+89D2
10 看	U+770B	看	U+770B	看	U+770B
11 間	U+9593	间	U+95FA	間	U+9593
12 敵	U+6562	敵	U+6562	敵	U+6562
13 感	U+611F	感	U+611F	感	U+611F

[그림 5] CJK808 한중일 한자 유니코드 표

[그림 5]표를 보면 각 나라의 한자가 같은 경우 유니코드 값도 동일하지만, 모양이 다른 이체자의 경우 3번 줄과 같이 유니코드 값이 다르게 나왔다. 그러므로 이체자를 표준 한자로 변환하기 위해 이체자가 입력되는 경우 같은 행에 해당하는 표준 한자 유니코드 값으로 변환하였다.

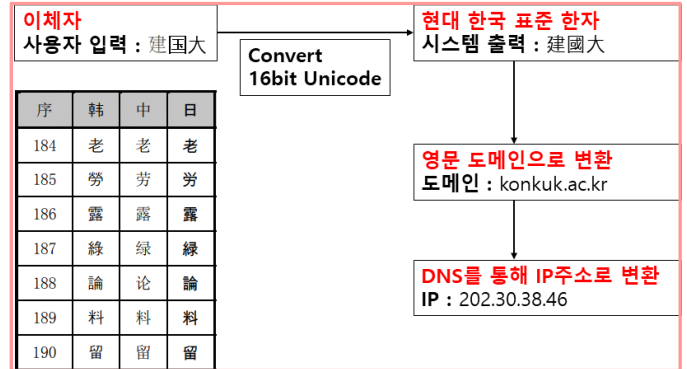
한국 유니코드	중국 유니코드	일본 유니코드	영문 도메인
1 U+5EFAU+570BU+5927	U+5EFAU+56FDU+5927	U+5EFAU+56FDU+5927	konkuk.ac.kr
2 U+6C11U+9858	U+6C11U+613F	U+6C11U+9858	minwon.go.kr
3 U+570BU+9632U+90E8	U+56FDU+9632U+90E8	U+56FDU+9632U+90E8	mnd.go.kr

[그림 6] 도메인 네임 키워드 데이터

유니코드에서 해당 DNS를 검색하는 과정은 [그림 6]과 같이 유니코드 값에 맞는 도메인 네임 키워드 데이터를 생성하였다. 아직 한글 도메인이 널리 알려지지 않고,

일반 한글 형태는 도메인 네임과 적합하지 않기 때문에 유니코드 값을 키워드로 하여 변환하였다.

해당 도메인 네임을 이용하여 IP 주소를 얻어오는 부분은 SDN 네트워크 내에서 한국 도메인 네임 서버 kns.kornet.net을 통하여 IP 주소를 얻어오는 방식을 이용하였다.



[그림 7] 애플리케이션 동작 과정

```

input hanja : 建国大
konkuk.ac.kr
202.30.38.46
input hanja : 建國大
konkuk.ac.kr
202.30.38.46
    
```

[그림 8] 애플리케이션 동작

[그림 7], [그림 8]은 한자 키워드 도메인 이름 시스템을 통해 만든 애플리케이션 실행 화면이다. [그림 7]은 애플리케이션 구성도이고, [그림 8]을 보면 두 가지 키워드에서 '建国大' '建國大'의 경우 '세울 건'은 중국식 한자와 한국식 한자의 경우 미세한 자형의 차이가 있고, '나라 국'의 경우에는 사용하는 한자의 모양이 명확히 다르다. 특히 '세울 건'과 같이 각 국가별로 미세한 자형의 차이가 있는 경우가 많다.

#### 4. 결론

본 논문에서는 소프트웨어 정의 네트워크를 이용한 한중일 한자-한국어 변환 키워드 도메인 이름 시스템을 제안하였다. 전체 시스템은 SDN을 이용함으로써 여러 이점을 얻을 수 있었고, 실제 사용자들이 이용할 수 있는 한자 키워드 도메인 이름 시스템은 한중일이 공통으로 자주 사용하는 808자의 한자를 조합한 CJK808를 이용함으로써 공통점은 최대화하고 차이점은 최소화하였다. 사용자들은 이 시스템을 통하여 이체자를 포함한 한중일 표준 한자를 이용하여 한국 사이트에서의 정보 검색, 분류 등 다양한 분야에서 도움이 되리라 생각된다.

그러나 현재 논문에서 제안한 CJK808 한자만으로는 한국 사이트에서의 정보 검색, 분류하기에는 부족하다. 또한 도메인 네임 키워드 데이터를 추가하는 방법에서 현재는 모든 도메인 주소를 직접 추가하는 방식이기에 이

를 개선할 필요가 있다.

향후 연구로는 앞에서 제안한 문제점을 해결하기 위해 CJK808 이 외 공용으로 사용하는 한자들을 더 추가하고, 도메인 네임 키워드 데이터 추가를 자동화하는 과정에 대해서도 연구가 필요하다. 또한 이 시스템을 애플리케이션에 적용하기 위해서 어떤 방향으로 개선해 나가야 할지에 대한 고민도 필요하다.

## 사사의 글

본 과제(결과물)는 교육부와 한국연구재단의 재원으로 지원을 받아 수행된 「대학혁신지원사업」의 연구결과입니다.

## 참고문헌

- [1] 설원원, “한중일 공용한자 808자 자서 중 자형과 육서 분류의 편차 연구”, 건국대학교, 2018.
- [2] 김태호, 이병주, 류경주, 전종대, 차민경, 『韓·中·日 공용한자 808 자』, 서울: 형민사, 2014.
- [3] 808 공용한자편찬위원회, 『한중일 공용한자 808 자』, 서울: 중앙일보, 2015
- [4] 김균태, 「한·중·일 (韓中日) 아래자형 설화의 비교 연구」, 『아시아비교 민간사회』, 2004
- [5] 양원석, 「‘韓·中·日 共用汉字 800 字’의选定与课题」, 『한국문화학회지』, 2013.
- [6] Chinese2020, “CJK 808 Character”, <http://chinese2020.com/CJK808>, 2016,4
- [7] ONF(Open Networking Foundation), “Software-Defined Networking(SDN) Definition”, <https://www.opennetworking.org/sdn-resources/sdn-definition>, 2016.4