

<https://doi.org/10.7236/JIIBC.2019.19.6.219>  
JIIBC 2019-6-31

## 다중 회귀 분석을 이용한 한자 난이도 예측 기법 연구

# Prediction Techniques for Difficulty Level of Hanja Using Multiple Linear Regression

최정환, 노지우, 김순태\*

Jeongwhan Choi, Jiwoo Noh, Suntae Kim\*

**요약** 한자 급수와 같이 기존 한자 난이도 선정 방식에 문제점이 있다. 실생활에서 쓰이는 한글 단어와 차이가 나며 해당 급수가 실제로 얼마나 많이 쓰이는지 알 수가 없다. 이러한 문제를 해결하기 위해 빈도수를 이용하여 다중 회귀 분석을 이용하여 한자 난이도를 측정한다. 초등 교과서를 기반으로 한자활용빈도수와 한글의미빈도수를 집계한다. 두 빈도수와 획수를 함께 사용하여 설문지를 작성하여 해당 한자의 학습 적정 시기를 답변 받아 이를 회귀에서 사용할 타겟 변수로 이용한다. 단계별 회귀분석을 이용하여 적절한 피처를 선택하고 다중 선형 회귀 분석을 한다. 모델의  $R^2$ 는 0.1105가 나왔으며 RMSE는 0.1105의 결과가 나왔다.

**Abstract** There is a problem with the existing method of selecting the difficulty levels of Hanja characters. Some Hanja characters selected by the existing methods are different from Sino-Korean words used in real life and it is impossible to know how many times the Hanja characters are used. To solve this problem, we measure the difficulty of Hanja characters using the multiple regression analysis with the frequency as the features. Based on the elementary textbooks, *FWS* and *FHU* are counted. A questionnaire is written using the two frequencies and stroke together to answer the appropriate timing of learning the Hanja characters and use them as target variables for regression. Use stepwise regression to select the appropriate features and perform multiple linear regression. The  $R^2$  score of the model was 0.1105 and the RMSE was 0.1105.

**Key Words** : Big Data Analysis, Feature Selection, Hanja, Multiple Linear Regression, Text Processing

### 1. 서론

기존 한자 난이도 선정 방식의 문제점이 있다. 첫 번째 문제점은 실제 해당 학년에서 얼마나 많이 쓰이는 단어 인지 파악이 되지 않는 것이다. 전문가가 임의로 선정하

여 난이도의 평가 기준이 모호하다. 이는 한국어문화 한 자능력시험에서 배정된 급수에서 확인할 수 있다. 예를 들어 8급 급수에서는 대한민국(大韓民國) 구성 한자가 모두 배정되어 있다. 이는 미취학아동이 대한민국을 쓰게 하기 위한 의도가 있지만, 미취학아동에게는 학습하기 어

\*정회원, 전북대학교 소프트웨어공학과  
접수일자: 2019년 10월 15일, 수정완료: 2019년 11월 15일  
게재확정일자: 2019년 12월 6일

Received: 15 October, 2019 / Revised: 15 November, 2019 /  
Accepted: 6 December, 2019

\*Corresponding Author: stkim@jbnu.ac.kr  
Dept. of Software Engineering, Jeonbuk National University,  
Korea

려운 한자들이 포함되어 있다.

두 번째 문제점은 한자의 의미와 실생활에서 쓰이는 한글 단어 차이가 난다는 점이다. 한자의 의미와 실생활의 한글과 의미상 연상이 되지 않는다. 이는 한자를 학습하는 데 있어서 어려움이 있을 수 있다. 예를 들어 8급에서 '쇠 금'의 뜻에 해당하는 '쇠'는 '철'을 일상적으로 이르는 말이다. 하지만 초등학교 사용 단어에서 '쇠'라는 말은 출현하지 않는다. 또한, 7급에서 '매양 매'의 '매양'은 '항상'과 같은 말이지만 '매양'이라는 단어는 초등학생이 일상에서 사용하지 않으며 교과서에서도 출현하지 않는다. 마찬가지로 7급에서 '저자 시'의 '저자'는 시장의 예스럽게 이르는 말이다. 이와 같이 옛말이 의미로 되어있는 한자는 초등학생이 학습하기에 적절하지 않을 수 있다.

본 논문에서는 초등학교 교과서에 출현하는 어휘들과 한자들을 기반으로 다중 선형 회귀 분석(Multiple Linear Regression Analysis)를 적용하여 한자 난이도를 예측하고자 한다. 예측한 한자 난이도를 기반으로 학년별 학습대상 한자와 단어를 추출한다.

다음 2장에서는 배경지식 및 관련 연구에 대한 내용을 기술하였고, 3장에서는 본 논문에서 주로 다루고자 하는 다중 회귀 분석을 이용한 한자 난이도 측정을 절차 별로 기술하였다. 마지막 4장에서는 결론 및 향후 연구에 대한 내용을 다룬다.

## II. 관련연구

교육용 한자의 문제점은 여전히 제기되고 있다. H.Chul<sup>[1]</sup>은 중고등 한문 교과서를 통해 교육용 한자에 대해 다시 논의되어야 할 필요가 있음을 제기하고 있다. 해당 논문에서는 한문 교과서의 본문에 사용된 한자를 통계하여 기초한자의 사용량을 밝히고 있다. 현행 교육용 기초한자가 실제적 언어 재료인 한문 고전이나 언어생활, 혹은 한자어 사용에 얼마나 깊은 관여를 하고 있는지를 알고 있다. 이는 여전히 한문교육용 기초한자 즉 상용한자로 선정된 1800자가 실제 교과서에서 어떻게 활용되고 있으며 그 때 발생하는 문제가 무엇인지에 대해서 검토된 바가 없다는 점을 문제 삼고 있다.

C.In-sook<sup>[2]</sup>의 연구에서는 초중고 교과서를 기반으로 초중고 텍스트 통합공식, 중고등학교 텍스트 전용공식, 초등학교 텍스트 전용공식을 구성하여 집단의 특성을 잘 반영하는 공식을 회귀분석을 통해 도출하였다. 그리고 텍스트의 언어적 난이도에 영향을 미치는 요인들을 사용

하여 텍스트수준 결정요인이 무엇인지 규명하고자 하였다. 하지만 C.In-sook의 연구에서는 한자어와 한자에 대한 난이도를 이용하여 측정하지는 않아 본 논문과 차이가 있다.

K.Eun<sup>[3]</sup>의 연구에서 또한 한자어 난이도 위계 설정에 대한 연구가 있었지만 이는 일본인 학습자를 중심으로 이루어졌고 일본어와의 유사성을 바탕으로 난이도를 선정하였다. 이는 학습대상과 난이도 결정 요인에 본 연구와 차이가 있다.

본 논문은 초등학교 교과서를 통해서 상용한자 1800자에 대한 난이도를 다시 재 검토하고자 한다. 이에 대한 내용은 다음 목차에서 구체적으로 다룬다.

## III. 회귀분석을 이용한 한자 난이도 측정

회귀 분석을 이용한 한자 난이도를 측정하기 위해서 초등 교과서와 상용한자 1800자를 사용한다. 초등교과서는 국어 교과서와 바른생활·슬기로운생활·즐거운생활 교과서와 사회 교과서를 사용한다. 초등학생이 일상적으로 사용하는 어휘들이 국어 교과서에 적절하다고 판단하였다. 수학과 과학 과목의 경우 특정 분야의 전문적인 어휘가 사용되어 실생활 어휘와 거리감이 존재하여 연구 초기단계부터 배제하였다. 그리고 사회 교과서를 추가하였는데, 1학년과 2학년은 사회 교과서가 존재하지 않는다. 따라서 1·2학년은 바른생활·슬기로운생활·즐거운생활 교과서로 대체하였다. 총 교과서 데이터셋은 19권, 11,524 문장, 97,345 단어이다.

### 1. 빅데이터 기반 빈도 추출

한자 난이도를 측정하기 위해서 초등 교과서를 기반으로 한자활용빈도와 한글의미빈도를 추출한다. 두 빈도를 추출하는 방법을 다음 장에서 설명한다.

#### 가. 한자활용빈도 추출

한자활용빈도(FHU; Frequency of the Hanja Usage)는 교과서 내에서 한자가 활용된 빈도를 말한다. 즉, 문장의 단어 중 한자가 얼마나 출현하는가를 말하는 데 이를 집계하기 위해서 한자 단어를 병행으로 표기하는 절차가 필요하고 이에 대한 예시는 표 1과 같다.

병행 표기를 위해 UTagger를 기반으로 하는 한글 한자 자동 변환 시스템을 사용한다<sup>[4]</sup>. 이는 한글에서 한자

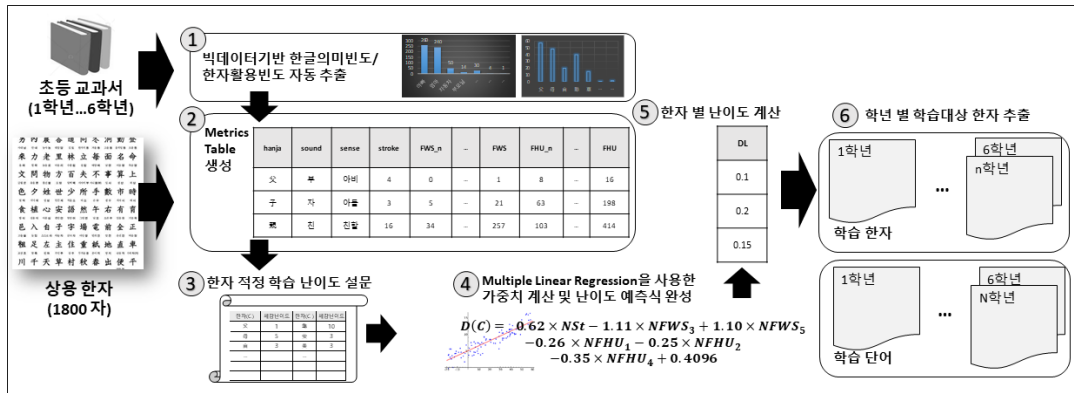


그림 1. 연구 방법 개요  
 Fig. 1. Overview of Research Methods

로 자동으로 변환하는 도구이다. 하지만 병기 표기로 변환 후에 검토하는 작업이 필요하다. 변환 시스템의 정확도를 확인할 수 없기 때문이다. 잘못된 한자로 변환된 문장도 검토를 해주어야 한다. 그리고 사람의 이름을 한자로 변환하는 경우가 있는데, 이는 다루고자 하는 한자가 아니기 때문에 병기 표기를 제거해주어야 한다. 그리고 병행 표기된 한자들과 해당되는 문장을 매핑하는 과정을 가진다.

표 1. 한자 병행 표기된 문장 샘플  
 Table 1. A sample of the Marked Sentence with Hanja Characters

문장	학교에서 만난 사람들과 인사해 봅시다.
병행표기	학교(學校)에서 만난 사람들과 인사(人事)해 봅시다.

나. 한글의미빈도 추출

한글의미빈도(FWS: Frequency of the Word Sense)는 교과서 내에서 한자 의미의 유의어가 쓰인 빈도를 말한다. 예를 들어, '母'의 의미인 '어미', '어머니', '엄마'가 몇 번 출현하는가를 말하는데 이를 위해서는 한자 의미의 유의어를 선정하는 과정이 필요하다.

한자 의미의 유의어를 선정하기 위해 네이버 국어사전에서 제공하는 '낱말 어휘정보처리연구소'의 유의어/반의어 사전을 사용한다. 이곳에서 제공하는 유의어들 중 적절한 유의어를 선정한다. 예를 들어, '매양 매'의 유의어로 '항상', '매번', 그리고 '번번이'를 선정한다. 그리고 유의어로 선정된 의미들과 한자 의미들을 매핑하는 과정을 거친다. 문장과 문장에 사용된 의미들을 매핑하는 과정을 통해 한글의미빈도를 집계하는데, 이때 어근화 과정이 필

요하다. 어근화가 필요한 이유는 해당 의미가 문장에서 사용되었는지 확인하기 위해서 필요하다. 어근화를 위해 OKT(Open-Korean-Text) 한글 형태소 분석기를 이용한다<sup>[5]</sup>. 한국어 처리를 통해 문장을 토큰화하고 토큰화된 부분을 어근화한다. 이에 대한 예시는 표 2와 같다.

표 2. 토큰화와 어근화된 문장 샘플  
 Table 2. A Sample of the Tokenized and Stemmed Sentence

문장	학교에서 만난 사람들과 인사해 봅시다.
토큰화	'학교/Noun', '에서/Josa', '만난/Noun', '사람/Noun', '들/Suffix', '과/Josa', '인사/Noun', '해/Verb', '봅시다/Verb', './Punctuation'
어근화	'학교/Noun', '에서/Josa', '만난/Noun', '사람/Noun', '들/Suffix', '과/Josa', '인사/Noun', '하다/Verb', '보다/Verb', './Punctuation'

토큰화와 어근화되어 태깅된 문장과 이와 동일하게 태깅된 한자 의미를 비교하여 매핑한다. 예를 들어, '사람인'의 태깅된 의미는 '사람/Noun'이기에 위의 표에서의 문장과 매핑될 수 있다. OKT에 따르면 간단한 한국어 처리를 통해 색인어를 추출하는 것이 목표이기에 완전한 수준의 형태소 분석을 지향하지 않기 때문에 어근화를 검토하는 과정이 필요하다. 문장과 의미가 매핑된 것 또한 검토가 필요하다. 이때, 고려해야 할 사항이 세 가지가 있다. 첫 번째는 단순히 해당 의미로 매핑이 되었는지 확인하는 것이다. 두 번째는 동음이의어를 확인하는 것인데 예를 들어, 눈과 배 등이 있다. 이를 구분하여 매핑을 검토한다. 세 번째는 본용언과 보조 용언을 확인하는 것이 있다. 본용언은 실질적 의미가 사용되었다고 할 수 있지만 보조 용언은 문법적 의미만이 사용되었다고 할 수 있

표 3. 메트릭 테이블의 일부  
Table 3. Metrics Table

idhanja	hanja	sound	mean	St	FWS <sub>1</sub>	FWS <sub>2</sub>	FWS <sub>3</sub>	FWS <sub>4</sub>	FWS <sub>5</sub>	FWS <sub>6</sub>	FWS	FHU <sub>1</sub>	FHU <sub>2</sub>	FHU <sub>3</sub>	FHU <sub>4</sub>	FHU <sub>5</sub>	FHU <sub>6</sub>	FHU
0	人	인	사람	2	6	15	42	46	51	62	222	14	31	24	29	57	42	197
1	日	일	날	4	1	7	15	7	14	19	63	3	5	9	10	19	15	61
2	十	십	열	2	0	0	1	1	4	3	9	0	0	1	1	3	3	8
3	大	대	큰	3	15	38	87	59	92	78	369	1	3	7	10	20	28	69
4	二	이	두	2	7	6	10	29	51	42	145	0	0	0	0	2	0	2
5	國	국	나라	11	0	3	8	5	33	25	74	0	1	16	12	5	22	56

다. 보조 용언과 매핑된 의미를 제거하는 과정이 필요하다. 예를 들어, '보다/Verb'라는 의미가 '초콜릿을 만들어 보다'에서 실질적 의미로써 사용되지 않기 때문에 이 문장과 매핑 된 것을 제거해야한다.

### 2. 한자 기준 빈도 메트릭 테이블 작성

추출된 한자활용빈도와 한글의미빈도를 취합하여 표 3과 같이 메트릭 테이블을 작성한다. 총 15개의 모든 feature에 대한 정보는 다음과 같다.

- 획수(Stroke): *St*
- 학년별 한글의미빈도수:  $FWS_1, FWS_2, FWS_3, FWS_4, FWS_5, FWS_6$
- 총 한글의미빈도수: *FWS*
- 학년별 한자활용빈도수:  $FHU_1, FHU_2, FHU_3, FHU_4, FHU_5, FHU_6$
- 총 한자활용빈도수: *FHU*

### 3. 한자 체감 난이도 설문

메트릭 테이블을 기반으로 설문지를 작성한다. 이 설문지는 해당 한자가 몇 학년, 몇 학기에 학습하는 것이 적절한지에 대한 응답을 받는 것이다. 임의적으로 선정된 한자들로, 총 100문항으로 진행되며 총 7명의 설문자에게 응답을 받는다. 각 학년별 *FWS*와 *FHU*에 대한 정보를 그래프로써 제공하고 설문자는 한자 학습 적정 학년/학기를 응답한다. 설문자의 답변에 대한 분포를 정규 분포로 가정하고 이상치(outlier)를 제거하기 위한 절차를 가진다. 정규 분포 중 평균에서 양쪽으로  $2\sigma$ 의 범위에 있는 모든 값인 약 95%를 사용한다. 즉, 나머지 5%는 이상치로 간주한다. 그리고 이 범위 내에서의 조정 평균을 사용한다.

### 4. 다중 선형 회귀 모델을 사용한 가중치 계산 및 한자 난이도 예측식 완성

#### 가. Feature Selection

Stepwise Regression Analysis의 Bidirectional elimination을 이용하여 피처를 선택한다. 이는 모델에서 가장 중요한 피처부터 시작하고 단계별로 피처를 추가하는 forward selection과 모델의 모든 피처를 적용 후 각 단계의 가장 중요한 피처를 제거하는 backward elimination을 조합한 것이다<sup>6)</sup>.

*FHU*(총 한자활용빈도수)의 기준을 달리하며 피처 선택을 한다. 이때, 해당 한자가 *n*번 이상 사용된 한자만을 사용하고 *n*번 미만인 한자는 제외한다. 표 4는 설문 결과를  $2\sigma$ 로 평균의 조정한 것이고 표 5는  $\sigma$ 로 조정한 데이터의 결과이다.

가장 높은  $R^2$ 는 표 4의 0.852이며 이는 *FHU*가 2 이상일 때이다. 그리고 bidirectional elimination으로 선택된 피처들은 *St, FWS<sub>3, FWS<sub>5, FHU<sub>1, FHU<sub>2, FHU<sub>4</sub></sub></sub></sub></sub>*와 같다.  $R^2$ 는 추정된 선형 모형이 주어진 자료에 적합한 정도를 계산하는 척도인 결정계수(coefficient of determination)이다. 즉 모델의 독립변수에 의해 설명된 종속변수의 분산의 비율을 나타낸다.

표 4.  $2\sigma$ 로 평균 조정한 데이터의 피처 선택 결과와 회귀 결과  
Table 4. Feature Selection and Regrssion Results Using Data Tuned Average by  $2\sigma$

<i>FHU</i> ≥ <i>n</i>	$R^2$	RMSE
n=0	0.798	0.1247
n=1	0.817	0.1127
n=2	0.852	0.1105
n=3	0.828	0.1283
n=4	0.812	0.1048

#### 나. Multiple Linear Regression

$FHU$ 가 2 이상일 때를 토대로 다중 회귀를 진행한다. 이때 Min-Max 정규화를 사용하여 데이터를 최대값과 최소값이 각각 1과 0이 되도록 조정한다.

훈련 데이터셋과 테스트 데이터셋의 비율은 8:2로 설정한다. 다중 선형 회귀 분석 결과의 메트릭으로 RMSE(Root Mean Squared Error)를 사용한다. 이는 모델 예측한 값과 실제 값의 차이를 다룰 때 사용한다.

회귀 결과 RMSE 값은 0.11046으로 나왔다. 그리고 해당 모델의 계수들은 표 5과 같고 각각의 계수들은 정규화된 feature들의 계수이다.

표 5. 다중 선형 모델의 계수들의 값  
 Table 5. The Coefficient Values of the Multiple Linear Model

Coefficient	Value
$NSt$	0.6164
$NFWS_3$	-1.1085
$NFWS_5$	1.1038
$NFHU_1$	.02659
$NFHU_2$	0.2542
$NFHU_4$	-0.3491

#### 5. 한자별 난이도 계산

다중 선형 회귀분석에 의한 예측식  $D(C)$  는 식(1)과 같다. 이 식을 통해 한자별 난이도를 계산한다.  $C$ 는 하나의 한자를 의미한다. 한자별 난이도를 계산하고 이를 오름차순으로 정렬한 결과의 일부는 표 6과 같다.

$$D(C) = 0.62 \times NSt - 1.11 \times NFWS_3 + 1.10 \times NFWS_5 - 0.16 \times NFHU_1 - 0.25 \times NFHU_2 - 0.35 \times NFHU_4 + 0.4096 \quad \dots (1)$$

표 6. 계산된 한자 난이도 결과의 일부  
 Table 6. Samples of the Predicted Difficulty Level of Each Hanja Character

hanja	sound	sense	strokes	difficulty
生	생	날	5	0
活	활	살	9	0.1070
一	일	한	1	0.1273
夕	석	저녁	3	0.1349
方	방	모	4	0.1751

#### 6. 학년별 대상 학습 한자 및 단어

학년별 대상 학습 한자와 단어를 선정하기 위해  $FHU$

가 5 이상인 한자 데이터들로 필터링한다. 즉, 총 한자활용빈도수 5회 이상인 한자만을 사용한다. 이때 총 한자수는 704자이다.

#### 가. 학년별 한자 추출

학년별로 한자를 추출하기 위해서 각 학년별로 난이도 범위를 휴리스틱 방법으로 선정한다. 학년별 난이도 범위와 이에 해당하는 한자 개수는 표 7과 같으며 6학년의 한자 난이도 범위를 조정할 때, 1.0보다는 더 낮은 난이도로 선정하였다. 그 이유는 초등 6학년이 1800자의 가장 높은 난이도까지 학습할 필요가 없기 때문이다. 이때 고려할 점은 각 학년에서 한자를 추출할 때, 해당 학년의  $FHU_n$ 은 1번 이상 사용되한다.

표 7. 학년별 난이도 범위와 해당 한자 개수  
 Table 7. The Difficulty Level Range by Grade and the Number of Hanja Characters

grade	range	No. hanja
1	$0 \leq d < 0.35$	35
2	$0.35 \leq d < 0.45$	32
3	$0.45 \leq d < 0.455$	108
4	$0.455 \leq d < 0.53$	122
5	$0.53 \leq d < 0.57$	148
6	$0.57 \leq d < 0.65$	127

#### 나. 학년별 단어 추출

##### (1) 단어 난이도로 학년별 단어 추출

단어에 구성된 한자들의 난이도를 식(2)와 같이 계산하여 단어의 난이도를 추출하는 방법이다. 이때  $h_i$ 는 단어의  $i$ 번째 한자이고, 각 한자들의 난이도의 평균을 단어의 난이도로 표현한다. 학년별 한자와 마찬가지로 단어 학년별  $FHU_n$ 이 1 이상인 단어들만을 추출한다. 표 8은 1학년 대상 한자어를 추출한 것들 중에서 네 단어만을 보여주며  $FW_1$ 는 1학년 교과서에서의 해당 한자어의 빈도수를 나타낸다.

$$Difficulty = average(h_1, \dots, h_i) \quad \dots (2)$$

표 8 단어 난이도로 추출한 1학년 대상 학습 한자어  
 Table 8. Samples of words for Grade 1 Extracted by the difficulty of the word

word	hanja word	difficulty	$FW_1$
삼촌	三寸	0.2421	1
추석	秋夕	0.2010	14
생일	生日	0.0535	5
생활	生活	0.1473	1

(2) 단어 빈도수로 학년별 단어 추출

각 단어의 출현 빈도수를 계산하여 각 학년별로 가장 많이 나온 단어들로 추출하는 방법이다. 이 방법을 선택한 이유로는 단어의 난이도를 계산하는 것이 타당하지 않기 때문이다. 위의 표 9에서 단어의 출현 빈도수가 고려되지 않는 경우가 있기 때문이다.

표 9. 단어 출현 빈도수로 추출한 1학년 대상 학습 한자어  
Table 9. Samples of words for Grade 1 Extracted by Frequency of the Word

word	hanja word	$FW_1$
문장	文章	47
모음자	母音字	46
자신	自身	29
선생님	先生님	25

IV. 결론 및 향후 연구

초등 교과서를 기반으로 한자활용빈도수와 한글의미 빈도수를 집계한다. 이 두 빈도수와 획수를 사용하여 설문지를 작성하여 해당 한자의 학습 적정 시기를 답변 받아 이를 회귀에서 사용할 타겟 변수로 이용한다. 단계별 회귀분석을 이용하여 적절한 피처를 선택하고 다중 선형 회귀 분석을 한다. 이를 통해 한자 난이도 예측식을 완성하고 한자별 난이도를 계산한다. 계산된 한자 난이도를 바탕으로 초등학교 학년별 학습 대상 한자와 단어를 추출한다. 한자 난이도 예측 모델의 RMSE는 0.1105,  $R^2$ 는 0.852의 결과가 나왔다. 학년별 대상 한자와 단어 모두 타당하게 추출된 것 같지만, 좀 더 많은 교과서 데이터와 일상생활에서 주로 사용하는 데이터가 필요하다고 판단된다.

해당 연구를 기반으로 한자 교육 플랫폼에 기여할 수 있다. 한자 난이도에 대한 객관적인 근거를 초등 교과서를 통해 제공할 수 있으며 학교 교과서와 연계한 교육 효과를 기대할 수 있다. 해당 연구를 기반으로 한글 문서의 난이도를 측정하는 연구 또한 진행할 수 있다. 이를 통해 문서의 한자의 음과 의미에 대한 난이도를 계산하여 가독성에 대한 정도를 측정할 수 있을 것이다.

References

[1] Heo Chul, "A study of using various corpus on set

grade Basic Han-characters and Sino-Korean words for Korean learning(1) -Review basic Han-character set for Hanmun education", Journal of Korean classical Chinese education52, no.0:95-114, 2019

[2] Choe In-Sook, "A Comparative Study on Modelling Readability Formulas: Focus on Primary and Secondary Textbooks", Journal of the Korean Society for information Management 22, No. 4: 173-95, Dec 2005.  
DOI: <https://dx.doi.org/10.3743/KOSIM.2005.22.4.173>

[3] Kim Eun, "A Study on Setting Difficulty Hierarchy for Sino-Korean Words for Korean Language Education - Focusing on Japanese Learners", Contemporary Society and Multiculture Volume: 8 Issue 1 : 166-189, 2018.  
DOI: <https://dx.doi.org/10.35281/csm.2018.06.8.1.166>

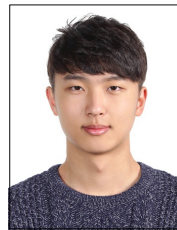
[4] "UTagger,[Korean Language Processing Laboratory]", Accessed August 7, 2019,  
<http://nlplab.ulsan.ac.kr/doku.php?id=utagger>.

[5] open-korean-text, "Open-Korean-Text: Open Korean Text Processor - An Open-Source Korean Text Processor", Accessed August 7, 2019,  
<https://github.com/open-korean-text/open-korean-text>.

[6] M. Wang, J. Wright, R. Buswell, and A. Brownlee, "A comparison of approaches to stepwise regression for global sensitivity analysis used with evolutionary optimization.", Proc. BS 2013 13th Conf. Int. Build. Perform. Simul. Assoc., pp. 2551-2558, 2013.  
DOI: <https://doi.org/10.1016/j.enbuild.2016.05.065>

저 자 소개

최 정 환(준회원)



- 전북대학교 학사 ('16-現)
- E-Mail: jeongwhan.choi@jbnu.ac.kr

노 지 우(준회원)



- 전북대학교 학사 ('18)
- 전북대학교 석사 ('18-現)
- E-Mail: wldn306@jbnu.ac.kr

김 순 태(정회원)



- 중앙대학교 컴퓨터공학과 학사('03),  
서강대학교 컴퓨터공학과 석사('07),  
박사('10)
- 전북대학교 소프트웨어공학과 교수  
( '14~現)
- E-Mail: stkim@jbnu.ac.kr

※ 본 연구는 중소기업청의 첫걸음기술개발사업의 일환으로 수행하였음. [1801001298, 한자-중국어 병행 맞춤형 자기 학습을 위한 학습 역량 진단 평가 알고리즘 개발]