

한국어 어휘자동획득 시스템

임희석^{1*}

An Automatic Korean Lexical Acquisition System

HeuiSeok Lim^{1*}

요약 본 논문은 인간의 언어 획득 원리를 반영한 계산주의적 한국어 어휘 자동 획득 시스템을 제안한다. 제안하는 시스템은 인간의 언어 생활을 모델링한 한국어 코퍼스를 입력 받아 언어 인식을 위하여 사용할 수 있는 어절 사전과 형태소 사전의 어절과 형태소를 자동으로 획득할 수 있다. 1천만 어절 크기의 한국어 코퍼스를 이용하여 실험한 결과, 2,097개의 어절과 3,488개의 형태소를 획득할 수 있었다. 획득된 2,097개의 어절의 출현 빈도의 합은 1천만 어절의 38.63%에 해당하였고 형태소 추출의 정확도는 99.87%를 보였다.

Abstract This paper proposes a automatic korean lexical acquisition system which reflects the characteristics of human language acquisition. The proposed system automatically builds two kinds of lexicon, full-form lexicon and decomposition using Korean corpus as its input. As the experimental results using Korean Sejeong corpus of which size is 10 million Eojeols, the system acquired 2,097 full-form Eojeols and 3,488 morphemes. The accumulated frequency of the acquired full-form Eojeols covers the 38.63% of the input corpus and accuracy of morpheme acquisition is 99.87%.

Key words : Lexical Acquisition, Computational Model, Corpus, Mental Lexicon

1. 서론

본 연구는 인지신경과학과 컴퓨터과학이 접목된 인지 신경계산학(cognitive neuro-computational study)의 한 분야라 할 수 있다[7][8]. 인지신경계산학은 대뇌에서 일어나는 인지 기능의 정보처리적 원리, 구조적 원리 등을 추상화를 통하여 계산주의적 모델을 개발하고 구현하며, 인지신경원리에 기반한 효율적이고 지능적인 컴퓨터 시스템에 응용하는 제반의 연구라 정의할 수 있다.

인지신경계산학의 필요성과 중요성은 다음과 같이 정리할 수 있다. 첫째, 인지신경기반의 모델을 분석하고 시뮬레이션함으로써 사람을 대상으로 한 행동 실험을 통해서 얻은 기존의 이론을 검증하고 심화 이해하는데 기여할 수 있다. 둘째, 계산주의적 모델 및 시스템의 변수 조작과 파라미터 변경을 통해서 사람 피험자를 이용한 행동 실험으로는 수행하기 힘든 예측 실험이 가능하다. 셋째, 계산주의적 모델에 인위적인 손상을 가하여 손상된

모델이 만들어내는 오류 관측 및 분석을 통하여 사람 피험자를 이용해서는 불가능한 손상 연구(lesion study)를 수행할 수 있다.

본 논문은 인간의 인지신경학적 언어정보처리 원리를 반영한 자동 어휘 획득 시스템(computational automatic lexical acquisition system)을 제안하고 대용량의 한국어 코퍼스를 이용한 어휘 획득 결과를 분석한다. 인간의 언어정보처리 원리를 반영한 자동 어휘 획득 시스템의 개발의 필요성은 다음과 같이 정리할 수 있다. 첫째, 인간의 어휘정보처리 과정의 이해를 위해서 어휘획득 과정에 대한 이해가 중요한데, 계산주의적 시스템은 대뇌속의 어휘 획득 과정에 대한 시뮬레이션 환경을 제공할 수 있다. 둘째, 언어정보처리 이해에 중요한 역할을 하는 어휘 획득 과 어휘 표상은 밀접한 관계를 가지고 있는데, 어휘 획득 연구는 어휘 표상 연구(mental lexicon representation)를 위해서도 매우 중요하다[8]. 셋째, 계산주의적 자동 어휘 획득 시스템은 인간의 어휘 획득 과정을 검증하고 예측 실험이 가능하게 하며 손상 연구를 통한 심도 있는 연구를 가능하게 할 수 있다. 넷째, 자동 어휘 획득에 대한 연구는 지능형 로봇 개발을 위한 자동 언어 처리 시스템의 어휘 정보처리 시스템 개발에 기여할 수 있다.

본 논문은 2007년도 한신대학교 학술연구비 지원에 의하여 연구되었음

¹한신대학교 컴퓨터공학부

*교신저자 : 임희석(limhs@hs.ac.kr)

2. 한국어 심성어휘 표상

심성어휘집(mental lexicon)은 인간이 언어처리를 하기 위하여 사용하는 대뇌속의 어휘집 또는 사전을 의미한다. 심성어휘집의 표상(mental lexicon representation)은 심성어휘가 대뇌 속에서 어떻게 조직화되고 저장되어 있는지를 의미한다[1, 2, 3, 9]. 심성어휘집의 구조와 심성어휘의 표상은 인간이 언어 생활을 하면서 자연스럽게 습득된 어휘 및 어휘 지식 획득의 결과이므로 언어 획득은 심성어휘집의 표상과 매우 밀접한 관계를 갖는다. 그러므로 인지신경기반 자동 어휘 획득 시스템은 인간의 심성어휘집의 표상을 반영할 수 있도록 설계되어야 한다.

심성어휘집 표상에 대한 연구는 결합 모델(full-list model), 분해 모델(decomposition model), 하이브리드(hybrid) 모델로 구분할 수 있다[4, 5, 6]. 결합 모델은 활용(inflexion)과 파생(derivation) 등으로 여러 개의 형태소로 결합된 단어나 어절이 언어 생활에서 사용되는 형태 그대로 심성어휘집에 저장되고, 단어 인식(word recognition) 또는 어절 인식 시 심성어휘집에 등록된 형태대로 탐색이 이루어진다는 모델이다. 예를 들면, “먹(동사)+었(선어말어미)+고(연결어미)”이나 “뚫(동사)+아(연결어미)” 등이 “먹었고”나 “도와”와 같이 어절 전체로 심성어휘집에 저장되고 탐색된다. 분해 모델은 결합 모델과는 달리 형태소와 같은 단어나 어절을 이루는 하위 단위의 어휘 정보가 심성어휘집에 저장되어 있으며, 단어 인식 시 입력된 단어를 심성어휘집에 저장된 단위로 분해하여 분해된 각 단위를 사전에서 탐색하고 탐색해서 얻은 정보를 통합하여 단어를 인식한다는 모델이다. 하이브리드 모델은 결합 모델과 분해 모델 모두가 심성어휘집의 표상과 탐색에 적용된다는 모델이다.

3. 인지신경기반 어휘 자동 획득 시스템

3.1 자동 획득 시스템의 개요

본 논문이 제안하는 어휘 자동 획득 시스템은 코퍼스를 입력받아 결합 모델의 어휘 사전과 분해 모델의 형태소 사전을 구축하는 하이브리드 모델의 어휘 자동 획득 시스템이다. 시스템의 어휘 학습을 위하여 사용하는 입력 코퍼스는 인간이 세상을 살아가면서 듣고 보게 되는 언어 생활을 모델링하는 것이다. 제안하는 시스템이 어휘를 획득을 위한 기본 가정은 [표 1]과 같다.

표 1. 어휘 자동획득 시스템의 기본 가정

- ① 언어 생활에서 자주 접하는 고빈도 어절은 효과적인 인식과 어절 생성을 위하여 어절 그 자체가 결합 모델의 심성어휘집에 저장된다.
- ② 특정 시점 또는 문맥에서 반복해서 지속적으로 출현하는 어절 또는 단어는 결합 모델의 심성어휘집에 저장된다.
- ③ 어절에서 고빈도로 발생하고 다른 문자열과 결합되어 다른 어절들을 생성하는 생산성이 높은 특정 문자열은 형태소로 추출되어 분해 모델의 심성어휘집에 저장된다.

제안하는 시스템을 이용하여 한국어 어휘를 자동 획득하는 과정을 간단히 설명한다. 코퍼스가 입력되면 전처리를 통하여 빈칸을 경계로 띄어쓴 어절이 어절 획득 모듈과 형태소 획득 모듈에 입력된다. 어절 획득 모듈에서는 입력된 어절의 빈도를 계산하여 어절 임계값이 결합 형태 임계값인 n 보다 큰 어절은 고빈도 어절로 간주하여 어절 사전에 저장한다. 또한 특정 어절 크기 내에서 반복해서 출현했는지를 어절 최근성 검증 모듈을 통하여 검증하여 최근성 임계값인 n 보다 최근성 빈도가 높은 어절도 어절 사전에 저장한다. 형태소 획득 모듈은 입력된 어절의 가능한 모든 음절 경계에서 머리문자열/꼬리문자열로 나누고 각 머리문자열과 꼬리문자열의 출현 빈도 및 동시 출현 문자열 정보를 추출한다. 특정 머리문자열 또는 꼬리문자열의 빈도가 임계값을 넘을 경우 해당 머리문자열과 같이 출현한 꼬리문자열의 첫 음절의 발생 엔트로피(순방향 엔트로피)와 꼬리문자열의 앞에 출현한 머리문자열의 마지막 음절의 발생 엔트로피(역방향 엔트로피)값을 계산한다. 계산된 양방향 엔트로피값의 변화가 형태소 가능성을 보일 경우 해당 머리문자열과 꼬리문자열을 형태소로 추출하여 형태소 사전에 저장한다.

본 논문이 제안하는 자동 어휘 획득 시스템에서 형태소는 어절을 두 개의 문자열로 나누어 앞부분의 문자열로 구성된 형태소 또는 형태소 열을 머리 형태소로 정의하고, 뒷부분의 문자열로 구성된 형태소 또는 형태소 열을 꼬리 형태소로 정의한다.

위에서 설명한 인지신경 기반 자동 어휘 획득 과정을 pseudo 알고리즘으로 나타내면 [그림 1]과 같으며 3.2절과 3.3절에서 결합 형태 획득 모델과 형태소 획득 모델에 대하여 자세히 설명한다).

1) 제안하는 모델이 한국어에 적용될 때 ‘결합 형태’는 어절을 의미하는 것이므로 논문의 이후부터 결합 형태를 ‘어절’로 표시한다.

```
// 코퍼스 사이즈 : 언어 생활 기간
// Input : 코퍼스
// Output : Fdic, Mdic
1. 코퍼스에서 빈칸을 경계로 추출한 문자열(W) 입력
2. W의 출현 빈도 정보 및 최근성 등록;
3. If (W의 출현 빈도 > 출현 빈도 임계값) then W를
   Fdic에 저장;
   else if ( W의 최근성 빈도 > 최근성 임계값 ) W를
   Fdic에 저장;
4. successor 정보 등록(head-tail정보 등록);
5. 특정 임계값 빈도를 갖는 head의 morpheme가능성을
   조사(양방향 엔트로피 검증);
6. Morpheme의 조건을 만족하는 Head, tail과 각
   category 정보를 Mdic에 저장;
```

그림 1. 인지신경기반 자동 어휘 획득 알고리즘

3.2 어절 획득 모듈

어절 획득 모듈의 어절 획득 원리는 빈도 정보를 고려한 획득과 최근성 정보를 고려한 획득 두 가지이다. 언어 생활에서 고빈도 어절은 저빈도 어절에 비하여 인식 속도가 빠른 빈도 효과(frequency effect)를 보인다[9]. 이러한 빈도 효과가 나타나는 이유 중 한 가지는 어절 전체가 분해되어 각각의 형태소가 인식되고 각 형태소를 조합한 어절이 인식되는 것이 아니라 어절 전체가 하나의 단위로 인식되기 때문이다. 이렇게 어절 전체가 사전에 저장되어야 하는 이유는 어절의 인식의 효율성에 있어서도 타당성을 갖는데, 고빈도 어절일수록 어절 전체를 사전에서 탐색하여 빠르게 인식하도록 하는 것이 언어 이해에도 효율적이라 할 수 있다. 고빈도 어절을 Fdic에 저장하여 어절 획득을 하는 것은 이러한 원리를 반영하는 것이다. 최근성 원리는 특정한 기간이나 특정한 시기에 집중적으로 반복하여 접하는 어절의 경우 결합 형태로 Fdic에 저장된다는 원리로 상대적으로 강한 자극과 입력을 받은 어절은 분해되어 이해되기 보다는 결합 형태로 이해되기 쉽기 때문이다.

어절 획득 모듈에서 사용되는 파라미터는 결합 형태 빈도 임계값(m)과 최근성 임계값(n) 두가지가 사용된다. 각 임계값은 획득할 언어의 특성에 따라 다르게 설정되어야 할 것이다. 두 파라미터값은 행동실험을 통하여 밝혀진 심성어휘집에 저장된 어절들의 평균 빈도 또는 최근성 빈도를 규명하여 사용하는 것이 이상적이나 본 논문은 m 로 500을 n 로 20을 사용하여 실험할 것이다. [그림 2]는 어절 획득 알고리즘을 슈도우 코드로 나타낸 것이다.

```
// 코퍼스 사이즈 : 언어 생활 기간
// Input : 코퍼스
// Output : Fdic
01 ft와 rt 값 설정;
02 tmp_dic을 NULL로 설정; //tmp_dic는 어절의 누적
   빈도를 저장하는 테이블
03 rct_dic를 NULL로 설정; // rct_dic는 어절의 최근
   성 정보를 저장하는 테이블
04 Context_Wsize = 0;
05 코퍼스에서 빈칸을 경계로 추출한 문자열(W) 입력;
06 if ( W가 tmp_dic 등록되어 있나?) then W의 빈도
   증가;
07 else W를 tmp_dic에 등록; W=1;
08 if ( Context Wsize = Max_Context_Wsize ) then
   Context_Wsize = 0;
09 else Context_Wsize 빈도 1 증가;
10 If (W의 출현 빈도 > 출현 빈도 임계값) then W를
   Fdic에 저장;
11 else if ( W의 최근성 빈도 > 최근성 임계값 ) W를
   Fdic에 저장;
12 successor 정보 등록(head-tail정보 등록);
13 특정 임계값 빈도를 갖는 문자열의 형태소 가능성을
   조사(양방향 엔트로피 검증);
14 Morpheme의 조건을 만족하는 Head, tail과 각
   category 정보를 Mdic에 저장;
```

그림 2. 어절 획득 알고리즘

3.3 형태소 획득 모듈

어린이의 언어 획득 과정을 살펴보면 언어 생활 초기에는 반복해서 접하게 되는 어절 전체가 기억되고, 기억된 어절을 활용하여 언어 생성을 하게 된다. 그러면서 언어 환경에 계속 노출되고 어절내에 반복적으로 자주 등장하는 문자열의 존재를 탐지하게 되고, 그 문자열이 어미 또는 조사 등과 같이 어울려 새로운 어절을 생성할 수 있음을 깨닫게 된다. 그 순간 그 문자열은 형태소로 심성어휘집에 등록되게 되며, 그 형태소와 결합할 수 있는 형태소에 대한 정보 및 부가 정보가 저장된다. 제한하는 시스템의 형태소 획득 기본 원리도 인간의 형태소 획득 원리와 동일하다. 코퍼스의 어절에서 자주 등장하는 문자열을 탐지하고 그 문자열이 형태소가 될 수 있음을 검증하여 형태소 사전인 Mdic에 저장한다. 특정 문자열이 형태소가 될 수 있음을 검증하는 방법은 특정 문자열의 후행 음절의 엔트로피(successor entropy)와 선행 음절 엔트로피를 이용하여 검증한다. 엔트로피를 이용한 특정 문자열의 형태소 검증을 예를 들어 설명한다.

예를 설명하기 이전에 본 논문에서의 어절에 대한 정의를 내린다. 본 논문은 어절을 두 부분의 형태소 결합열로 구분하여 앞부분의 형태소 결합열을 머리 형태소(head morpheme)라 정의하고, 뒷부분의 형태소 결합열을 꼬리 형태소(tail morpheme)라 정의한다. [그림 3]은 본 논문에서 가정하는 어절, 머리형태소 그리고 꼬리 형태소를 BNF(Backus-Naur Form)로 정의한 것이다. 그림 4에서 {unit}⁺과 {unit}⁰는 각각 해당 unit이 1번 이상, 0번 이상 반복됨을 의미한다.

| |
|---|
| 어절 ::= 머리형태소 꼬리형태소 |
| 머리형태소 ::= {실질형태소} ⁺ {형식형태소} ⁰ |
| 꼬리형태소 ::= {실질형태소} ⁰ {형식형태소} ⁺ |
| 실질형태소 ::= {음절} ⁺ |
| 형식형태소 ::= {음절} ⁺ |

그림 3. 어절, 머리 형태소, 꼬리형태소의 정의

[표 2]는 코퍼스에서 문자열 “아빠”로 시작되는 어절과 각 어절의 출현 빈도를 나타낸 것이라 가정한다.

표 2. “아빠”로 시작되는 어절 빈도

| 어절 | 빈도 |
|--------|----|
| 아버지께서도 | 8 |
| 아버지만 | 4 |
| 아버지께도 | 4 |
| 아버지 | 16 |
| 아버지라도 | 8 |
| 합계 | 40 |

[표 2]의 어절의 예에서 문자열 “아”와 “아버”, 그리고 “아버지”의 출현 빈도는 모두 40회로 다른 부분 문자열보다 매우 높은 출현 빈도를 보인다. 따라서 이들 문자열이 모두 머리 형태소의 가능성이 있다. 하지만 형태소는 그 특성상 다른 형태소와 결합하여 어절을 생성하므로 만약 어떤 문자열이 머리 형태소라면 그 형태소 다음에는 여러 가지 꼬리 형태소가 연결될 수 있을 것이다. 예를 들면, “아버지”라는 문자열 다음에는 출현할 수 있는 문자열은 “지”일 것이나 “아버지”라는 머리형태소가 될 수 있는 문자열 뒤에는 여러 가지의 꼬리형태소가 올 수 있다는 것이다. 따라서 본 논문은 고빈도로 출현한 특정 문자열이 머리형태소의 가능성이 있는지를 조사하기 위하여 머리형태소 뒤에 출현하는 음절의 발생 엔트로피를 이용하여 계산한다. 특정한 문자열 뒤에 출현할 수 있는 음절의 엔트로피는 [식 1]과 같이 정의한다.

$$Entropy(S) = - \sum p_i \log p_i, \text{ where } p_i = \frac{Freq(S, C_i)}{Freq(S)} \quad [식 1]$$

본 논문은 엔트로피값이 상승하는 위치의 문자열을 찾아 머리 형태소 후보로 결정하고 이 후보가 머리 형태소일 가능성을 조사한다. 엔트로피가 상승하는 문자열을 바로 머리 형태소로 결정하지 못하는 이유는 특정 문자열을 공유하는 다양한 형태소가 존재할 수 있기 때문이다. 예를 들어, 표 3에서 부분 문자열 “아”가 반복적으로 출현하고 “아” 다음 음절의 종류도 “이”, “가”, “기”, “침” 등으로 다양하게 나타나고 있다. 따라서 “아”의 엔트로피 값이 높은 값으로 계산되어 머리 형태소로 등록되는 오류가 발생할 수 있다. 만약 “아”가 머리 형태소일 경우 “이스크림도”, “가도”, “기를”, “침에”, 그리고 “이스크림을”이 꼬리 형태소이어야 하는데, 이들이 꼬리 형태소일 가능성이 없다면 부분 문자열 “아”도 머리 형태소가 될 수 없다고 할 수 있다.

표 3. “아”로 시작되는 어절의 예

| 어절 | 빈도 |
|--------|----|
| 아이스크림도 | 4 |
| 아가도 | 8 |
| 아기를 | 8 |
| 아침에 | 8 |
| 아이스크림을 | 4 |

본 논문은 머리 형태소일 가능성의 조사를 위하여 어절에서 머리 형태소 후보를 제외한 나머지 부분이 꼬리 형태소일 가능성이 있는지를 조사하는 방식을 사용한다. 꼬리 형태소의 가능성은 머리 형태소 후보를 선정하는 방식과 같이 엔트로피값을 이용하는데, 어절의 마지막부터 역방향으로 만들어지는 문자열의 빈도 정보를 이용하여 앞에 나올 수 있는 음절에 대한 엔트로피 값을 계산한다. 예를 들면, “아버지께서도”라는 문자열에서 문자열 “아버지”가 머리 형태소가 되기 위해서는 “께서도”가 꼬리 형태소일 가능성이 있어야한다. 따라서 이 문자열의 역순 문자열인 “도서께”의 다음에 출현하는 음절의 엔트로피가 높은 값이 될 것이다.

4. 실험 및 결과

본 논문이 제안하는 자동 어휘 획득 실험을 위해서는 어휘 학습을 위한 코퍼스(C)와 결합형태 추출을 위한 결

합 형태 빈도 임계값(ft), 최근성 임계값(rt)가 결정되어야 한다. 본 논문은 어휘 학습을 위한 코퍼스로 세종 코퍼스 1천만 어절을 사용하였으며, ft로 500, rt로는 200을 사용하였다.

어절 획득(결합형태) 모듈에 의하여 총 2,097개의 어절이 학습되었으며 학습된 어절의 빈도의 합이 학습 코퍼스 1천만의 38.63%에 해당하였다. 형태소 획득 모듈에 의해서 획득된 형태소의 개수는 머리형태소와 꼬리형태소 각각 1,744개로 총 3,488개의 형태소가 획득되었다. 올바른 형태소가 획득되었는지는 정확도 계산에 의하여 이루어졌으며 99.87%의 정확도를 보였다. 획득된 머리 형태소나 꼬리 형태소가 그림 3에서 제시한 형식에 맞을 경우 정확하게 획득되었다고 판정한다. 정확도는 [식 2]와 같이 전체 획득된 형태소 수에 대한 정확하게 획득된 형태소의 비율로 계산한다.

$$\text{형태소 획득 정확도} = \frac{\text{정확하게 획득된 형태소수}}{\text{획득된 형태소수}} \quad [\text{식 2}]$$

5. 결론

본 논문은 인간의 언어 학습 원리를 반영한 자동 어휘 획득 시스템을 제안하였으며, 이를 한국어에 적용하여 실험하였다. 1천만 어절 크기의 한국어 코퍼스를 이용하여 실험한 결과 전체 코퍼스에서 빈도의 합이 38.63%를 차지하는 2,097개의 어절이 학습되었으며, 3,488개의 형태소가 99.87%의 정확도로 추출되었다.

심성어휘집이나 전자 사전은 어휘뿐만 아니라 품사 정보, 구문 정보, 그리고 의미 정보 등 그 어휘가 가지는 언어 지식으로 구성된다. 본 논문이 제안하는 모델은 이 중 어휘만을 획득하는 모델인데, 어휘뿐만 아니라 언어 지식을 획득할 수 있는 모델로 발전하여야 할 것이다. 언어 지식 중 특정 형태소와 연결될 수 있는 형태소들의 정보를 나타내는 분류(품사)정보는 해당 형태소와 연결될 수 있는 형태소들의 분포 정보를 이용한 자동 분류 방법을 이용하여 획득될 수 있을 것이며, 이에 대한 연구를 수행 중에 있다.

참고문헌

[1] Bradley, A. D. Lexical representation of derivational relation. In M. Aronoff and M. L. Kean(Eds.), *Juncture*, 37-55. Cambridge, MA: MIT Press, 1980.

[2] Caramazza, A., Laudanna, A., Romani, C. Lexical access and inflectional morphology. *Cognition*, Vol. 28, pp. 207-332, 1988.

[3] Foster, K. I. Accessing the mental lexicon. In R. J. Wales, E. Walker (Eds.), *New approaches to language mechanisms*, pp. 257-287, 1976.

[4] Jung, J., Lim, H., Nam, K., *Morphological Representations of Korean compound Nouns*. *Journal of Speech and Hearing Disorders*, Vol. 12, pp. 77-95, 2003.

[5] Lim, H., Nam, K., Hwang, Y., *A Computational Model of Korean Mental Lexicon*, *Lecture Notes in Computer Science*, Vol. 3480, pp. 1129-1136, 2005.

[6] Nam, K., Seo, K., Choi, K., *The word length effect on Hangeul word recognition*. *Korean Journal of Experimental and Cognitive Psychology*, 9, 1-18, 1997.

[7] McLeod, P., Plunkett, K., Rolls, E. T., *Introduction to Connectionist Modelling of Cognitive Processes*, Oxford University Press, 1998.

[8] O'Reilly R. C., Munakata, Y., *Computational Explorations in Cognitive Neuroscience Understanding the Mind by Simulating the Brain*, The MIT Press, London, England, 2000.

[9] Taft, M., *Reading and the mental Lexicon*. Hillsdale, NJ : Erlbaum, 1991.

임희석(Heui-Seok Lim)

[중신회원]



- 1992년 2월 : 고려대학교 컴퓨터학과(이학사)
- 1994년 2월 : 고려대학교 컴퓨터학과(이학석사)
- 1997년 9월 : 고려대학교 컴퓨터학과(이학박사)
- 1997년 9월 ~ 1999년 2월 : 삼성종합기술원 HCI Lab.

• 1999년 3월 ~ 현재 : 한신대학교 컴퓨터정보소프트웨어학부 부교수

<관심분야>

자연어처리, 인공지능, 인지신경계산학, 정보검색