

문장 및 단어 중요도를 통한 한국어 문서 연관 이미지 검색

김남규, 강신재*
대구대학교 컴퓨터정보공학부

Relevant Image Retrieval of Korean Documents based on Sentence and Word Importance

Nam-Gyu Kim, Shin-Jae Kang*

Division of Computer & Information Engineering, Daegu University

요약 텍스트로만 이루어진 글에서 알지 못하는 단어가 나온다면, 글을 읽는 도중 집중이 되지 않고 내용을 이해함에 있어 어려움이 생긴다. 또한 이미 알고 있는 단어라도 아이들의 경우 경험이 적기 때문에 글에서 상황을 묘사하는 표현이 생소하거나 애매하다면 머릿속에 떠올리기 힘들다. 이에 본 논문에서는 글을 이해를 돕고 독자의 흥미를 증가시키기 위해서 글의 텍스트들을 분석하여 중요하다고 판단되는 내용을 선택하고, 이 내용과 가장 관련 있는 이미지를 웹에서 자동으로 가져와 연결하여 보여주는 시스템을 구현하고자 한다. 시스템의 구현은 글을 문단 단위로 나누어 글을 분석하고, 문단마다 중요한 문장을 선택한 후, 중요한 문장 내에서 이 문장을 가장 잘 표현할 수 있는 중요한 단어들을 선택하여 웹에서 연관 이미지를 검색하고, 검색된 이미지 결과를 이전에 나눈 각 문단마다 연결시켜준다. 실험으로 글에서 중요한 문장을 선택하는 방법과 문장 내 중요한 단어를 선택하는 방법을 제시하였다. 실험한 결과, 선택된 이미지 3개와 해당 중요 문장과의 연관 여부를 정확률로 평가하였을 때 60%의 성능을 얻을 수 있었다.

Abstract While reading text-only documents and finding unknown words, readers will become the focus disturbed and not be able to understand the content of the documents. Because children have little experience, it is difficult to understand correctly if the description in context is unfamiliar or ambiguous. In this paper, in order to help understand the text and increase the interest of the readers, we analyze the texts of documents and select the contents that are considered important, and implement a system that displays the most relevant images automatically from the web and links the texts and the images together. The implementation of the system divides the article into paragraphs, analyzes the text, selects important sentences for each paragraph and the important words that best represent the meaning of the important sentences, searches for images related to the words on the web, and then links the images to each of the previous paragraphs. Experiments have shown how to select important sentences and how to select important words in the sentences. As a result of the experiment, we could get 60% performance by evaluating the accuracy of the relation between three selected images and corresponding important sentences.

Keywords : Word Importance, Sentence Importance, TF-IDF, TextRank, Relevant Image Retrieval

1. 서론

어린 아이들이 독서를 할 때, 텍스트만으로 구성된 글

을 읽게 되면 쉽게 지루해지고, 생소한 표현이나 어려운 단어가 등장할 때에 글의 의미를 잘 이해하지 못하는 어려움이 발생한다. 다음은 생소한 표현과 어려운 단어에

본 논문은 2018학년도 대구대학교 학술연구비지원으로 수행되었으며, 2018년도 인문사회과학기술융합학회 하계학술대회에서 발표한 내용을 확장한 것임.

*Corresponding Author : Shin-Jae Kang(Daegu Univ.)

Tel: +82-53-850-6584 email: sjkang@daegu.ac.kr

Received December 5, 2018

Revised December 31, 2018

Accepted March 8, 2019

Published March 31, 2019

대한 예시이다.

예시1 : 붉은 검기가 검광을 뿌리자 허공에 매화 꽃잎이 하나둘 피어나기 시작했다. 그리고 검무가 점점 고조될수록 그에 비례해 매화향은 점점 더 짙어졌다.

예시2: 합지박만큼 크다는 그의 얼굴입니다.

위 예시에서 첫 번째 문장은 무협소설들에서 자주 보이는 무공들에 대한 설명 중 하나이다. 평소에 무협이라는 장르를 전혀 접해보지 못한 사람이라면 이런 생소한 표현에 대해 전투의 모습이 어떻게 진행되어 가는지 머릿속으로 상상하기가 매우 힘들다. 두 번째 문장은 통나무 속을 파서 큰 바가지 같이 만든 그릇, 합지박 대해 얼굴을 비유한 문장인데 앞뒤로 합지박에 대한 다른 묘사가 있었다더라도 합지박에 대해서 모른다면 글의 표현에는 한계가 있기 때문에 색, 크기, 질량, 질감 등을 상상하기 힘들어 그냥 넘어가버리거나 왜곡되게 생각하며 글을 읽어나간다. 본 연구에서는 이러한 문제를 해결하여 흥미를 더하고 이해를 돕기 위해 내용에 적합한 그림 이미지를 웹에서 자동으로 검색하여 보여주는 시스템을 구현하고자 한다.

2장에서는 본 연구의 배경이 되는 주요 문장 및 단어 추출 연구와 자연어처리의 중요한 연구 주제인 단어의 의미성 해소와 이미지 검색에 관한 기존 연구에 대해 정리하고, 3장에서는 대규모 언어 지식베이스나 학습 말뭉치 없이 문서 연관 이미지 검색 시스템을 구현하는 방법을 제안하고, 4장에서 실험결과를 제시하며, 5장에서 결론 및 향후 연구방향에 대해 서술한다.

2. 관련 연구

2.1 중요 문장 및 단어 선택

글에서 중요 문장 및 단어를 추출하는 방법에는 여러 가지가 있다. 구글의 PageRank를 텍스트에 적용시킨 TextRank 알고리즘[1]을 통해 중요 문장을 추출해낼 수도 있고[2-3], 정보검색 분야에서 많이 활용되고 있는 TF-IDF[4]를 이용해 중요 단어를 추출해 낼 수도 있다[5]. 기존 TextRank 알고리즘에 상호정보량 척도를 결합하여 군집기반에서 키워드를 추출하는 기법[6]과 폭소노미 시스템을 이용한 단어들 간의 의미적 관계를 분석해 중

요 단어를 추론하는 기법[7], 문맥 확장을 통한 다중문서 요약 기법[8], 의미와 맥락을 고려하여 문장 간 수반관계를 이용해 요약문의 의미와 개념적 연결성을 고려한 기법[9] 등이 있다.

2.2 동형이의어 처리

동형이의어 처리에 대해서는 고전적인 방법으로 각 단어의 사전적 뜻을 풀이를 이루는 단어들을 이용해 서로 같은 단어가 얼마나 있는지에 따라 사전적 의미를 결정하는 Lesk 알고리즘과 평균 상호정보량 및 복합명사 의미사전을 이용해 Lesk 알고리즘보다 성능을 향상시킨 기법[10], 의미중의성을 해소시키기 위한 효과적인 품사 집합을 찾는 연구[11], 단어 간 의존 구문분석을 위해 한국어 가이드라인과 언어분석의 평가를 위한 엑스브레인 언어분석 말뭉치를 제시한 연구[12] 등이 있다.

2.3 이미지 검색

국내에서는 ImageNet에서 제공하는 대용량의 이미지 데이터와 라벨을 CNN 딥러닝 기술로 학습시켜 인스타그램 이미지로부터 라벨 정보를 추출해 자동으로 태깅시켜 인스타그램의 기존 검색보다 더 높은 정확도를 산출해낸 연구[13]가 있고, 오토인코더의 인코더를 CNN으로 구성한 연구[14]가 있었으나, 문서와 연관된 이미지를 자동으로 검색하는 본 연구의 주제와는 다소 차이가 있다. 국외에서는 Meng[15]의 연구가 본 연구와 유사한 연구인데, 자연어처리 기법과 WordNet을 이용한 단어 추출 모듈, 키워드를 기반으로 한 웹 크롤러 디자인 및 이미지 다운로드 모듈, CNN을 기반으로 이미지 태깅 모듈, Keywords Overlap 알고리즘과 Score First 알고리즘과 TF-IDF를 통해 검색에 쓰인 문장과 추출된 이미지에 태깅된 결과를 비교하여 문장에 가장 근접한 이미지를 찾는 시스템을 제안하였다. TED 영어 연설문을 대상으로 실험한 결과 60.7%의 정확률을 보였다.

3. 연구내용 및 방법

전체 시스템의 절차는 그림 1과 같다. 문서가 입력 데이터로 들어오면 우선적으로 글을 문단 단위로 분할한다. 문단 단위로 나누는 이유는 글은 대체로 비슷한 내용의 문맥이 문단 단위로 구분되기 때문이다. 따라서 문단

을 나누어 이미지를 연결한다면 각 내용의 흐름에 맞게 적절한 이미지를 연결할 수 있을 것으로 보인다. 나뉜 문단은 한국어 정보처리를 위한 파이썬 패키지 `konlpy` 에서 Twitter 형태소 분석기를 이용해 명사만을 추출하며, 불용어와 명사, 형용사, 동사 외 도움이 되지 않는 품사를 가진 단어들은 전부 제거한 후, 중요 문장과 단어를 구분하기 위해 그림 2와 같이 TF-IDF 알고리즘과 TextRank 알고리즘을 통해 다양한 방법으로 문장과 단어의 중요도를 계산하였다. 계산된 중요도를 기반으로 먼저 문단 내 하나의 문장을 선택하게 되는데, 그 이유는 문단 전체에서 중요 단어를 선택하게 되면 여러 문장에 흩어져 출현한 단어들이 선택되므로 이는 하나의 장면을 잘 표현할 수 있는 이미지 검색을 위한 절의어로 사용하기에는 적합하지 않기 때문이다. 이후 문장 내 선택된 주요 단어를 사용하여 문단과 연관된 이미지를 검색하고, 검색된 상위 이미지를 해당 문단에 연결하여 주는 절차로 진행된다.

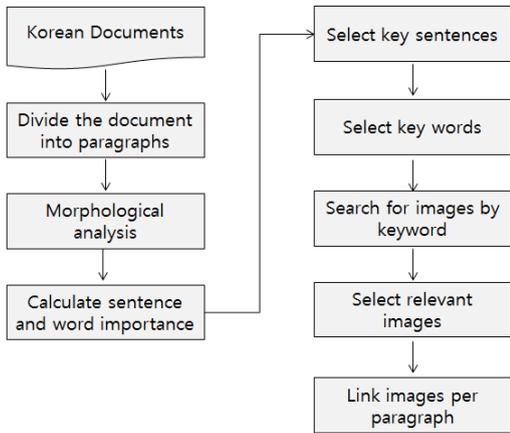


Fig. 1. Overall Process for Image Retrieval System

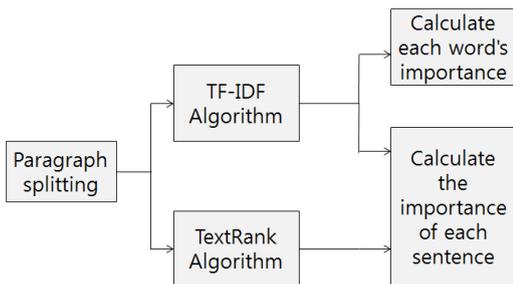


Fig. 2. Calculation of the importance of sentences and words through various algorithms

3.1 문장 및 단어 중요도 계산

문단을 분할한 후, TF-IDF 기반의 문장 중요도 및 단어 중요도를 구하기 위해 미리 TF-IDF를 계산하는 과정을 거친다.

$$TF(t, d) = f(t, d) \tag{1}$$

$$IDF(t, d, D) = \frac{D}{(d \in D: t \in d) + 1} \tag{2}$$

$$TF-IDF(t, d, D) = \log(TF(t, d) + 1) * \log(IDF(t, d, D)) \tag{3}$$

TF-IDF에서 TF(식 1)는 단어 빈도수로서 문서 d 내에서 단어 t 의 개수, 즉 문서 내에서의 단어의 출현 횟수를 나타내고, IDF(식 2)는 역 문서빈도로 단어 t 가 한번이라도 출현한 문서의 개수를 전체 문서의 개수 D 로 나눈 DF값의 분모와 분자를 바꾼 값이다. TF와 IDF의 곱 연산이 TF-IDF(식 3)로 문서 내의 단어의 출현 횟수가 높을수록, 전체 문서에서 한번이라도 출현한 횟수가 낮을수록 TF-IDF값이 높아지면서 단어의 중요도 수치가 높아진다.

$$TR(V_i) = (1 - d) + d \times \sum_{V_j \in In(V_i)} \frac{w_{ij}}{\sum_{V_k \in Out(V_j)} w_{jk}} TR(V_j) \tag{4}$$

TextRank 알고리즘을 적용하기 위해 위 (식 4) 왼쪽의 $TR(V_i)$ 는 문서에서 문장들을 각각 하나의 노드로 간주하고 그래프화하였을 때, 임의의 노드 V_i 가 다른 노드들로부터 얼마만큼의 영향을 받았는지를 계산했을 때, 노드 V_i 의 TextRank 중요도를 의미한다. $\sum_{V_j \in In(V_i)}$ 는 노드 V_i 로 향하는 다른 노드들 V_j 의 간선의 크기들의 합이고, w_{ij} 는 앞의 $\sum_{V_j \in In(V_i)}$ 를 얼마만큼 적용시킬지의 대한 가중치이다. $\sum_{V_k \in Out(V_j)}$ 는 노드 V_j 에서 다른 노드들 V_k 로 나가는 간선의 크기들의 합이다. w_{jk} 는 앞의 $\sum_{V_k \in Out(V_j)}$ 가 얼마만큼 적용시킬지의 대한 가중치이다. 여기에 $TR(V_j)$ 라는 노드 V_j 의 중요도를 곱하면 $TR(V_i)$ 의 TextRank 점수가 계산된다.

3.2 주요 문장 및 주요 단어 선택

주요 문장을 선택하는 방법으로 첫 번째는 TF-IDF가 가장 높은 단어를 포함하는 문장이 여러 개일 시 두 번째로 높은 단어를 추가로 포함한 문장을 선택하며, 그래도 여러 개가 선택되면 세 번째로 높은 단어를 포함하는 방식으로 TF-IDF 수치가 가장 높은 단어를 포함한 최종한 문장을 선택한다. 두 번째는 TF-IDF 상위 3개 단어를 동일한 가중치로 가장 많이 포함하고 있는 문장을 선택하는데, 만약 Top1 단어를 포함한 문장이 있더라도 Top2, Top3를 동시에 포함하고 있는 문장이 있다면 후자를 택한다. 포함하고 있는 단어의 개수가 같다면 Top4 단어를 추가로 고려하는 순으로 선택한다. 세 번째로 TextRank 알고리즘을 이용해 문장을 선택하는데, TextRank 알고리즘을 장 단위로 실행하고 문단별 가장 높은 TextRank 수치를 가진 문장을 선택한다.

주요 문장의 주요 단어 선택 방법은 문장 내의 단어들을 대상으로 TF-IDF 상위 3개부터 4개 단어를 선택하거나 문장 내 최상위 TF-IDF 단어를 기준으로 주변단어 1~2개를 함께 선택하는 방법을 취하였다.

3.3 연관 이미지 검색 및 선택

주요 단어를 이용해 구글 이미지 검색을 실행하여 나오는 이미지들의 URL 주소를 첫 번째부터 순서대로 가져온다. 구글 검색결과는질의어와의 연관정도에 따라 검색결과가 정렬되어 제시되므로 상위 3개의 이미지 검색 결과를 가져온다. 3개의 이미지 모두가질의어와 연관된 이미지라면 다양한 형태의 결과를 제시함으로써 문서의 이해도를 높일 수 있을 것이고, 동형이의어에 의해 상관없는 이미지가 포함되더라도 복수개의 이미지를 선택함으로써 정답 이미지가 포함될 확률을 높이고자 하였다. 선택된 3개의 이미지는 해당되는 문단의 옆에 제시하게 된다.

4. 실험 및 평가방법

제안한 각 방법에 따라 검색한 이미지 3개를 대상으로 중요문장과 연관이 있는가를 수작업으로 정확률을 확인하였다. 실험 데이터는 국립국어원 언어정보나눔터에서 “사과나무밭 달님”[16]이라는 12개의 단편을 모은 동화집을 사용하였으며, 동화집에 포함된 총 1,001개의 문

단 중 앞 20개의 문단을 대상으로 정확률을 평가하였다.

4.1 실험결과

표 1에 제시한 바와 같이 문장 선택 단계에서는 최상위 TF-IDF 단어를 포함한 문장을 선택한 경우와, TF-IDF 상위 3개 단어를 가장 많이 포함한 문장을 선택했을 때가 가장 높은 정확률을 보였고, 단어 선택 단계에서는 TF-IDF 상위 3개 단어, 최상위 TF-IDF 단어와 그 주변단어 1개, 최상위 TF-IDF 단어와 그 주변단어 2개를 이용하여 문맥정보를 추출한 경우가 60%로 가장 높은 정확률을 보였다. 정확률을 낮추는 주된 원인으로 동형이의어와 다의어를 완전하게 해결하지 못한 부분이 있었고, 형태소 분석 오류, 품사 태깅 오류로 인한 문제가 있었다. 실험 데이터가 상이해서 객관적으로 비교하기 어렵지만 [15]에서 TED 영어 연설문을 대상으로 개체명 인식 등 심층 자연어처리 기법을 적용하여 얻은 60.7%와 거의 대등한 성능을 보여, 한국어 단어 의미 전자사전이나 의미 정보가 태깅된 말뭉치 등 방대한 언어학적 정보 없이 얻은 결과이므로 본 연구의 의미가 있다고 볼 수 있다.

Table 1. Image accuracy by word and sentence selection algorithm

Word selection algorithm / Sentence selection algorithm	TF-IDF Top 3 words	One word around the top TF-IDF	Two words around the top TF-IDF	TF-IDF Top 3 words, Top TF-IDF 1 word & Surrounding 2 words
Sentence with top-level TF-IDF word	50%	40%	40%	60%
Sentence containing most of the TF-IDF top three words	45%	45%	45%	60%
TextRank algorithm	30%	40%	40%	45%

4.2 시스템 실행결과 예시

Korean Input Documents	Extracted Keywords & Retrieved Images
<p>마을 입구, 다 허물어져 가는 비각 추녀 밑이 아주머니의 거처였습니다. 낮 동안은 거의 이 양지바른 비각 문 앞에 쭈그리고 앉아 낮잠을 자기도 하고, 멍하니 앞산을 바라보면서 세월을 보내었습니다.</p>	<ul style="list-style-type: none"> • monumental house, entrance, residence 
<p>외양간 주인은 아주머니를 위해 보릿짚 몇 단을 더 넣어 주었습니다. 얼마 소와 송아지가 있는 외양간에 잘 때면 꼭 그들과 한 식구처럼 느껴졌습니다. 밤중에 문득 겨드랑이가 간지러워 눈을 떠보면 송아지가 주둥이를 내밀고 아주머니 겨드랑 사이를 파고 들어와 있는 것이었습니다.</p>	<ul style="list-style-type: none"> • cowshed, calf 
<p>탑아, 탑아, 이 꽃 줄까? "머리에 꽃이 봐, 응?" 동네 아이들은 아주머니에게 존댓말을 쓰지 않고 마구 이름도 부르고 놀려 주기도 했습니다. 그날도 아이들은 산에서 꺾어 온 진달래꽃을 가지고 탑이 아주머니에게 보여 온 것입니다.</p>	<ul style="list-style-type: none"> • tower, azalea, flower 

Fig. 3. Example of system results

그림 3에서 첫 번째 문단은 주요 단어로 선택된 “비각”과 “입구”, “거처”로 인해 연관 이미지가 검색된 예이고, 두 번째 문단은 주요 단어인 “외양간”과 “송아지”로 인해 연관 이미지가 검색된 결과이다. 세 번째 문단에서는 주요 단어 “진달래꽃”에 의해 제대로 검색이 되었지만, 다른 주요 단어인 “탑”은 실제로는 등장인물의 별칭인 고유명사인데 명사 “탑”으로 해석될 소지가 있는 경우이다. 이처럼 동형이의어에 의한 오류가 시스템의 성능에 영향을 미치므로 향후 해결해야 하는 과제이다.

5. 결론

본 논문에서는 문서의 중요 내용을 가장 잘 표현하는 키워드를 자동으로 추출하여 관련 이미지를 검색하고, 이를 문서의 연관 이미지로 추천하는 연구를 제안하였다. 문서를 대표하는 주요 문장을 선택하기 위해 정보검색 연구에서 많이 활용되고 있는 TF-IDF와 구글의 PageRank를 텍스트에 적용시킨 TextRank 알고리즘을 활용하였으며, 주요 문장 내 대표 단어를 선택하기 위해서는 상위 TF-IDF값을 갖는 단어의 주변 단어를 다양한 범위를 설정하여 문맥정보를 추출하여 활용하였다. 이미지 검색 단계에서는 단어의 의미 중심성 해소 문제가 성능에 큰 영향을 미치게 되는데, 이를 해결하기 위해서는 동형이의어를 포함한 모든 단어의 의미 목록과 관련 정

보가 정리된 대규모의 지식베이스가 구축되어 있거나, 단어의 의미 정보가 태깅된 방대한 양의 말뭉치가 있어 야 기계학습 등을 통하여 중심성 해소가 가능하지만, 본 논문에서는 이러한 자원이 없더라도 적절한 성능을 얻기 위해 대표 키워드 주변의 단어들을 대상으로 다양한 범위를 설정하여 문맥 정보를 수집하여 질의어를 확장한 것이 기존 연구와 차별화된 부분이라 할 수 있다. 실험결과 60%의 정확률을 얻을 수 있었으며, 이는 [15]에서 TED 영어 연설문을 대상으로 심층 자연어처리 기법을 적용하여 얻은 60.7%와 거의 대등한 성능으로, 실험 데이터가 상이해서 객관적으로 비교하긴 어렵지만 방대한 언어학적 정보 없이 얻은 결과이어서 의미가 있다고 볼 수 있다.

향후 연구로는 성능을 보다 개선하기 위해 단어 간의 존관계를 분석하여 의미적으로 관련된 주요 단어들을 검색에 함께 활용함으로써 연관 이미지 검색의 성능을 높이고자 한다.

References

- [1] R. Mihalcea, P. Tarau, “TextRank: Bringing Order into Texts”, *Proc. of EMNLP-04 and the 2004 Conference on Empirical Methods in Natural Language Processing*, July, 2004. Available From: <https://web.eecs.umich.edu/~mihalcea/papers/mihalcea.emnlp04.pdf>
- [2] J. P. Hong, J. W. Cha, “A Korean Important Sentence Extraction using TextRank Algorithms”, *Proc. of Korean Institute Of Information Scientists and Engineers*, Vol.36, No.1C, pp. 311-314, 2009. Available From: <http://www.dbpia.co.kr/Journal/ArticleDetail/NODE01219123#>
- [3] Y. S. Cho, K. H. Ahan, S. K. Kim, “A Study on Key Sentence Extraction by using Enriched TextRank”, *Proc. of KIIT Summer Conference*, 5, pp. 39-44, 2013. Available From: <http://www.dbpia.co.kr/Journal/ArticleDetail/NODE02172622>
- [4] C. D. Manning, P. Raghavan, H. Schütze, *Introduction to Information Retrieval*, pp. 100~123, Cambridge University Press, 2008. DOI: <https://doi.org/10.1017/CBO9780511809071.007>
- [5] E. D. You, G. H. Choi, S. H. Kim, “Study on Extraction of Keywords Using TF-IDF and Text Structure of Novels”, *Journal of Korean Institute Of Information Scientists and Engineers*, Vol.20, No.2, pp. 121-129, 2015. DOI: <https://doi.org/10.9708/jksci.2015.20.2.121>
- [6] H. M. Yoo, H. J. Kim, J. Y. Chang, “Automatic Construction of Reduced Dimensional Cluster-based Keyword Association Networks using LSI”, *Journal of*

Korean Institute Of Information Scientists and Engineers, Vol.44, No.11, pp. 1236-1243, 2017.
DOI: <https://doi.org/10.5626/jok.2017.44.11.1236>

- [7] J. U. Heu, Y. D. Joo, D. H. Lee, "Multi-Document Summarization Technique using Semantic Analysis between Tags", *Journal of Korean Institute Of Information Scientists and Engineers*, Vol.39, No.1, pp. 78-88, 2012. Available From: <http://www.koreascience.or.kr/article/JAKO201213956027474.page>
- [8] D. H. Kim, T. M. Cho, J. H. Lee, "Multi-document Summarization Using Context-based Sentence Expansion", *Proc. of Korean Institute of Intelligent Systems*, Vol.25, No.1, pp. 149-150, 2015. Available From: <http://www.dbpia.co.kr/Journal/ArticleDetail/NODE06277145>
- [9] Y. D. Kwon, N. R. Kim, J. H. Lee, "Document Summarization Considering Entailment Relation between Sentences", *Journal of Korean Institute Of Information Scientists and Engineers*, Vol.44 No.2, pp. 179-185, 2017.
DOI: <https://doi.org/10.5626/jok.2017.44.2.179>
- [10] J. Hur, M. G. Jang, "Homonym Disambiguation using Sense-Tagged Compound Noun Dictionary", *Proc. of KIIT fall Conference*, Vol.32, No.2 pp. 538-540, 2005. Available From: <http://www.dbpia.co.kr/Journal/ArticleDetail/NODE00951194>
- [11] C. H. Kwak, Y. H. Seo, C. H. Lee, "Efficient Part-of-Speech Set for Knowledge-based Word Sense Disambiguation of Korean Nouns", *Journal of the Korea Contents Association*, Vol.16, No.4, pp. 418-425, 2016.
DOI: <https://doi.org/10.5392/jkca.2016.16.04.418>
- [12] J. H. Lim, Y. J. Bae, H. K. Kim, Y. J. Kim, K. C. Lee, "Korean Dependency Guidelines for Dependency Parsing and Exo-Brain Language Analysis Corpus", *Proc. of the 27th Annual Conference on Human and Cognitive Language Technology*, pp. 234-239, 2015. Available From: http://ocean.kisti.re.kr/IS_mvpopo212L.do?ResultTotalCNT=55&pageNo=6&pageSize=10&method=list&acnCnt=&poId=sight&kjic=OOGHAK&sVnc=y2015m10a&id=&setId=&iTableId=&iDocId=&sFree=&pQuery=%28kjic%3AOOGHAK%29+AND+%28voliss_ctrl_no%3Ay2015m10a%29
- [13] H. W. Jang, S. S. Cho, "Automatic Tagging for Social Images using Convolution Neural Networks", *Journal of KIISE*, Vol.43, NO.1, pp. 47-53, 2016.
DOI: <https://doi.org/10.5626/jok.2016.43.1.47>
- [14] Y. J. Ahn, D. Y. Hong, K. S. Shim, "Image Clustering with CNN and AAE", *Proc. of the KISS conference*, Vol.44, No.2, pp. 245-247, 2017. Available From: <http://www.dbpia.co.kr/Journal/ArticleDetail/NODE07322113>
- [15] W. Meng, "A sentence-based image search engine", Masters Theses, Missouri university of science and technology, 2015. Available From: https://scholarsmine.mst.edu/cgi/viewcontent.cgi?article=8473&context=masters_theses
- [16] Online place to share language information, National Korean Language Institute, Available From: <https://ithub.korean.go.kr>. (accessed May, 18, 2018)

김 남 규(Nam-Gyu Kim)

[준회원]



•2013년 2월 ~ 현재 : 대구대학교
컴퓨터정보공학부 컴퓨터공학전공
(학사과정)

<관심분야>
자연어처리, 빅데이터

강 신 재(Shin-Jae Kang)

[정회원]



•1995년 2월 : 경북대학교 컴퓨터공
학과 (공학사)
•1997년 2월 : 포항공과대학교 컴퓨
터공학과 (공학석사)
•2002년 2월 : 포항공과대학교 컴퓨
터공학과 (공학박사)
•2002년 3월 ~ 현재 : 대구대학교
컴퓨터정보공학부 컴퓨터공학전공
교수

<관심분야>
자연어처리, 시맨틱웹, 빅데이터 처리