Journal of Korea Multimedia Society Vol. 20, No. 3, March 2017(pp. 533-542)
https://doi.org/10.9717/kmms.2017.20.3.533

# 한국어 번역 소설에서 인물명 명사구의 동일인물 공통참조 클러스터링 방법

박태근†, 김승훈††

# A Method for Clustering Noun Phrases into Coreferents for the Same Person in Novels Translated into Korean

Taekeun Park†, Seung-Hoon Kim††

## ABSTRACT

Novels include various character names, depending on the genre and the spatio-temporal background of the novels and the nationality of characters. Besides, characters and their names in a novel are created by the author's pen and imagination. As a result, any proper noun dictionary cannot include all kinds of character names. In addition, the novels translated into Korean have character names consisting of two or more nouns (such as "Harry Potter"). In this paper, we propose a method to extract noun phrases for character names and to cluster the noun phrases into coreferents for the same character name. In the extraction of noun phrases, we utilize KKMA morpheme analyzer and CPFoAN character identification tool. In clustering the noun phrases into coreferents, we construct a directed graph with the character names extracted by CPFoAN and the extracted noun phrases, and then we create name sets for characters by traversing connected subgraphs in the directed graph. With four novels translated into Korean, we conduct a survey to evaluate the proposed method. The results show that the proposed method will be useful for speaker identification as well as for constructing the social network of characters.

**Key words:** Literature, Character Name, Noun Phrase, Coreferent, Name Set

## 1. 서 론

최근, 영어 소설에서 인물명을 추출하고, 발화자를 인식 (Speaker Identification)하며, 인물의 성별, 나이, 성향, 감정을 파악하고, 인물간 소셜 네트워크를 분석하는 것을 목표로 하는 연구들이 진행되었다 [1-3]. 특히, [3]에서는 어린이에게 부모와 유사한 방식으로 책을 읽어주는 지능화된 TTS (Text-To-Speech) 시스템 개발을 위하여, 소설 분석 기술을 연구하였다. 지능화된 TTS 시스템이란 책에 등장하는 인물의 성별, 나이, 성향 등에 가장 적절한 목소리로 감정을 실어서 책을 읽어주는 시스템을 의미한다.

이와 같은 연구의 필수적인 전반부 작업은 소설에 등장하는 인물명 명사구를 추출하고, 추출한 명사구들을 동일인물 공통참조(Coreferents)로 클러스터링하는 것이다. 영어 소설에서 인물명은 "Mr. Sherlock

※ Corresponding Author: Taekeun Park, Address: (16890) 152, Jukjeon-ro, Suji-gu, Yongin-si, Gyeonggi-do, Korea, TEL: +82-31-8005-3162, FAX: +82-31-8021-7221, E-mail: tkpark@dankook.ac.kr
Receipt date: Dec. 12, 2016, Revision date: Jan. 10, 2017
Approval date: Jan. 17, 2017
† Dept. of Applied Computer Engineering, Dankook University
†† Dept. of Applied Computer Engineering, Dankook University
(E-mail: edina@dankook.ac.kr
※ This research was supported by Ministry of Culture, Sports and Tourism (MCST) and Korea Creative Content Agency (KOCCA) in the Culture Technology (CT) Research & Development Program 2016.

Holmes", "Mr. Holmes", "Sherlock Holmes", "Sherlock", "Holmes"와 같이 명사구 형태인데, 추출한 명사구들을 공통참조로 클러스터링 한다는 것은 이들을 하나의 이름 집합 (Name Set)으로 묶는 것을 의미한다 [1].

대소문자를 구분하는 영어 소설에서 인물명 명사구를 추출하는 도구 [4]는 이미 제공되고 있으나, 대소문자를 구분하지 않는 한국어 소설에서 인물명 명사구를 추출하는 것은 상대적으로 어렵기 때문에 인물명 명사구를 추출하고 동일인물 공통참조로 클러스터링하는 연구는 진행되지 않았다. 다만, 한국어로 창작되거나 번역된 소설에서 명사구가 아닌 단일 단어 형태의 인물명 추출 연구는 최근 진행되었다 [5, 6]. CPFoAN (Connective and Possessive Forms of Animate Nouns) [6]은 [5]의 발전 형태이며, 비록 "셜록 홈즈"라는 명사구를 추출하지 못하지만, 단어 "셜록", "홈즈", "씨(Mr.)" 등을 추출할 수 있다. CPFoAN에서는 "이사벨라"와 같은 고유 명사를 인물명 (Character Name)이라 하고 "아버지"와 같은 일반 명사를 등장인물 (Character Nominal)이라 한다. 그러나 관련연구를 서술하는 2장을 제외한 본 논문의 나머지 부분에서는 CPFoAN에 의해 추출된 인물명과 등장인물을 모두 인물명이라 통일하여 사용한다.

본 논문에서는, 한국어로 번역된 소설에서 CPFoAN에 의해 추출된 인물명을 활용하여, 인물명 명사구들을 추출한 뒤, 이 명사구들을 동일인물 공통참조로 클러스터링하는 방법을 제안한다. 제안하는 방법은 인물명으로 추출된 하나의 단어를 중심으로 특정 조건을 만족하는 앞뒤의 단어를 추가하는 방식으로 인물명 명사구를 추출하고, 추출된 명사구를 이용하여 방향성 그래프 (Directed Graph)를 구축한 뒤, 노드 분할과 연결 서브그래프 (Connected Subgraph) 순회를 통하여 동일인물 이름 집합 (Name Set)을 생성한다. 단어 $w_i$가 CPFoAN에 의해 추출된 인물명이거나 인물명을 포함하고 있다 (즉, 조사와 함께 사용된다)고 가정하자. 제안하는 방법은 $w_i$를 포함하는 모든 문장에서 $w_i$의 앞에 있는 두 개의 단어 ($w_{i-2}$ 및 $w_{i-1}$)와 $w_i$의 뒤에 있는 단어 ($w_{i+1}$)로부터 인물명 명사구를 추출한다. 단어 $w_i$가 이미 추출된 이름이거나 성인 경우, 단어 $w_{i+1}$은 성, 직업/직함 (Title), 또는 관계 (Relation)를 나타내는 단어일 수 있다. 또한, 단어 $w_i$가 이미 추출된 성, 직업/직함 (Title), 또는 관계 (Relation)를 포함하는 단어인 경우, 단어 $w_{i-1}$은 이름이거나 성일 수 있다. 단어 $w_{i-2}$는 $w_{i-1}$의 인물명 해당 여부 검토에 활용된다. 그 다음, 제안하는 방법은 추출된 명사구를 활용하여 방향성 그래프 $G$를 구축한다. 예를 들어, "셜록 홈즈"와 "홈즈 씨"가 인물명 명사구로 추출된 경우, 방향성 그래프 $G$에서 노드 "셜록"은 노드 "홈즈"와 방향성 링크 <"셜록", "홈즈">로 연결되고, 노드 "홈즈"와 노드 "씨"는 방향성 링크 <"홈즈", "씨">로 연결된다. 제안하는 방법은 구축된 방향성 그래프 $G$에서 필요시 노드 분할 뒤, 분할 후의 그래프 $G'$에 속한 연결 서브그래프를 동일인물 이름 집합으로 생성한다. 예를 들어, 방향성 그래프에 <"셜록", "홈즈">와 <"홈즈", "씨">라는 방향성 링크로 "셜록", "홈즈", "씨"라는 세 개의 노드가 연결 서브그래프를 구성하고 있다면, {"셜록", "홈즈", "셜록 홈즈", "홈즈 씨"}를 동일인물 이름 집합으로 생성한다. 동일인물 이름 집합은 소설의 등장인물간 대화 분석에 필수적일 뿐만 아니라, 인물간 소셜 네트워크 추출에서 동일인물의 노드 중복 방지에 필수적이다.

본 논문의 구성은 다음과 같다. 2장에서 관련 연구들을 소개하고, 3장에서 한국어로 번역된 소설에서 인물명 명사구를 추출하는 방법과 동일인물 공통참조 클러스터링 방법에 대하여 서술한다. 4장에서는 제안하는 방법이 네 권의 소설로부터 추출한 동일인물 공통참조 클러스터링 결과에 대하여 분석한 뒤, 마지막으로 5장에서 결론 및 향후 연구 방향을 정리한다.

## 2. 관련 연구

한국어로 작성된 소설 또는 문헌 국역본에서의 개체명 추출 내용을 포함하는 연구는 많지 않다. [7]에서는, 19세기에 작성된 문헌의 국역본에 대하여 개체명 추출의 필요성을 역설하기는 하였으나, 연구 내용에서 개체명 추출은 수동으로 이루어졌다.

[8]에서는, 한국어로 번역되거나 창작된 세 권의 소설에 대하여 등장인물간 소셜 네트워크의 구축을 위하여 인물명을 추출하기는 하였으나, 인물명 추출의 모든 단계를 자동화하지는 못했다. 구체적으로, [8]은 [9]에서와 같이 KAIST HanNanum 형태소 분

석기를 기본적으로 활용하면서, 이에 추가로 조사 목록 및 고유명사 사전을 활용하여 개체명을 자동 추출한 뒤, Wikipedia와 잘 알려진 인명사전에 대하여 추출된 인물명을 수동으로 교차 확인하는 작업을 수행하였다.

이상의 연구들과는 달리, [5]는 한국어에서 유정명사와 결합하여 사용되는 조사 목록(예: '-에게' 포함 13개 조사)과 꼬꼬마 형태소 분석기 [10]를 이용하여, 한국어로 번역되거나 창작된 소설에서 인물명과 등장인물을 자동으로 추출하였다. [5]에서는 유정명사와 결합하여 사용되는 조사를 유정조사라 불렀으며, 유정조사를 이용하는 간단한 아이디어만으로도, 한국어로 창작되거나 번역된 80권의 소설로부터 81.88%의 F-measure를 얻어낼 수 있음을 보였다. 하지만, 유정조사의 사용에 대한 작가의 성향, 즉 작가의 필체에 따라, 유정조사만을 사용하는 방법 [5]은 80권의 책 중에서 두 권의 소설에 대하여 40% 수준의 매우 낮은 재현율 (Recall)을 가질 수 있음이 발견되었다.

이러한 문제점을 해결하고 성능을 향상시키기 위하여, CPFoAN [6]이 제안되었다. CPFoAN은 [5]에 의하여 발견된 인물명 및 등장인물과 연결형 형태(Connective Form)로 등장하거나, 이미 발견된 세 명의 인물명의 소유격 형태 (Possessive Form)와 동일한 형태로 등장하는 후보단어를 인물명 및 등장인물로 추가 추출한다. [5]에서 사용한 동일한 80권의 책에 대하여, CPFoAN의 F-measure는 90.36%로 증가하였으며, Recall은 최저 60% 이상으로 증가하였다. 또한, 등장비율이 1% 이상인 인물명에 대한 정확률, 재현율, F-measure는 각각 97.84%, 96.19%, 97.01%로 매우 높게 나타났다.

그러나, CPFoAN은 명사구가 아닌 단어 형태의 인물명 및 등장인물을 추출하는 방법이기 때문에, 본 논문에서는 CPFoAN에서 추출되는 인물명 및 등장인물을 활용하여, 한국어로 번역된 소설에서 인물명의 명사구를 추출한 뒤, 이 명사구들을 동일인물 공통참조로 클러스터링하는 방법을 제안하고자 한다.

## 3. 제안하는 방법

본 논문에서 제안하는 방법은 두 단계로 구성된다. 첫 번째 단계는 CPFoAN에서 추출되는 인물명을 활용하여, 소설로부터 인물명 명사구를 추출하는 단계이다. 예를 들어, CPFoAN에 의해 추출된 "홈즈"를 활용하여, 제안하는 방법의 첫 번째 단계는 "홈즈"와 함께 등장하는 "셜록 홈즈"라는 명사구와 "홈즈 씨"라는 명사구 등을 추출한다. 두 번째 단계는 추출된 명사구들과 CPFoAN에 의해 추출된 인물명들로 방향성 그래프 $G$를 구축하고, 노드 분할 절차를 거친 뒤, 분할 후의 그래프 $G'$에 속한 연결 서브그래프를 동일인물 이름 집합으로 생성하는 것이다.

### 3.1 [단계 1] 인물명 명사구 추출 방법

제안하는 인물명 명사구 추출 방법은, 단어 $w_i$가 CPFoAN에 의해 추출된 인물명이거나 인물명을 포함하고 있을 때, $w_i$를 포함하는 모든 문장에서 $w_i$의 앞에 있는 두 개의 단어 $w_{i-2}$ 및 $w_{i-1}$과 $w_i$의 뒤에 있는 단어 $w_{i+1}$를 검사하여 인물명 명사구를 추출한다. Fig. 1의 Algorithm 1은 세 개의 단어 $w_{i-2}$, $w_{i-1}$ 및 $w_i$로부터 인물명 명사구를 추출하는 알고리즘이고, Fig. 2의 Algorithm 2는 두 개의 단어 $w_i$와 $w_{i+1}$로부터 인물명 명사구를 추출하는 알고리즘이다.

Algorithm 1은 네 가지 입력으로부터 인물명 명사구 집합 $S_{np}$를 생성한다. 첫 번째 입력 정보는 번역된 소설 텍스트이고, 나머지 세 개의 입력 정보는 이전에 추출한 단일 단어 형태의 인물명 집합인 $S_{cn}$과 조

Algorithm1: Extract noun phrases with previous two words

Input: A novel translated into Korean,
$S_{cn}$: set of known character names,
$S_{jx}$: set of words appearing with one of {'은', '는', '이', '가'},
$S_{p\&s}$: set of postpositions and symbols in [5].
Output: $S_{np}$: set of extracted noun phrases.

1. for each $w_{cn} \in S_{cn}$ {
2. for each line $L$ including word $w_i$ for $w_i = w_{cn} + p$ and $p \in S_{p\&s}$ in the novel{
3. extract previous two words, $w_{i-2}$, $w_{i-1}$, of $w_i$ in line $L$;
4. if ($w_{i-1} \in S_{cn}$) insert "$w_{i-1}$ $w_{cn}$" into $S_{np}$; //case 1
5. else if ($f_m(w_{i-1})$ is ambiguous && $w_{i-1} \in S_{jx}$) //case 2
6. insert "$w_{i-1}$ $w_{cn}$" into $S_{np}$;
7. else if ($w_{i-2}$ == null && $f_m(w_{i-1})$ is ambiguous ) { //case 3
8. if (*freq*("$w_{i-1}$ $w_{cn}$") > *threshold*($w_{cn}$) )
9. insert "$w_{i-1}$ $w_{cn}$" into $S_{np}$;
10. } // end of else if
11. else if ($f_m(w_{i-2})$ is ambiguous && $f_m(w_{i-1})$ is ambiguous ) {//case 4
12. if (*freq*("$w_{i-1}$ $w_{cn}$") > *threshold*($w_{cn}$))
13. insert "$w_{i-1}$ $w_{cn}$" into $S_{np}$;
14. } // end of else if
15. } // *end of for*
16. } // *end of for*

Fig. 1. Algorithm for extracting noun phrases with previous two words.

Algorithm2: Extract noun phrases with next one word

Input: A novel translated into Korean,
$S_{cn}$: set of known character names,
$S_{p\&s}$: set of postpositions and symbols in [5].
Output: $S_{np}$: set of extracted noun phrases.

1. for each $w_{cn} \in S_{cn}$ {
2. for each line $L$ including word $w_i$ for $w_i = w_{cn}$ without a postposition in the novel{
3. extract next one word $w_{i+1}$ of $w_i$ in line $L$;
4. if ($f_m(w_{i+1})$ is ambiguous ) {
5. for ( $k = 2$; $k <= w_{i+1}.length()$ ; $k$++) {
6. $freq_k$ = the appearance frequency of "$w_i$ $w_{i+1}.substring(0, k)$" in the novel;
7. } // end of for
8. $min$ = minimum value of $freq_k$ for $2 \le k \le w_{i+1}.length()$;
9. $max$ = maximum value of $freq_k$ for $2 \le k \le w_{i+1}.length()$;
10. if ($min$ != $max$) {
11. $l$ = largest $k$ for $2 \le k \le w_{i+1}.length()$ where $freq_k$ is max;
12. insert "$w_i$ $w_{i+1}.substring(0, l)$" into $S_{np}$;
13. } // end of if
14. } // end of if ($f_m(w_{i+1})$ is ambiguous )
15. } // end of for
16. } // end of for

Fig. 2. Algorithm for extracting noun phrases with next one words.

사 {'은', '는', '이', '가'}의 집합 $S_{jx}$ 및 [5]에서 인물명 등장빈도 계산을 위하여 정의한 조사와 기호의 집합 $S_{p\&s}$이다. Algorithm 1이 처리를 완료하면, 집합 $S_{np}$에 새로운 명사구가 추가되었는지를 확인하고, 추가된 명사구로부터 각 단어를 분리하여 집합 $S_{cn}$에 삽입한 뒤, 새로운 명사구가 추출되지 않을 때까지 Algorithm 1을 반복적으로 호출한다.

Algorithm 1은 다음의 네 가지 경우에 두 단어가 연결된 명사구 "$w_{i-1}$ $w_{cn}$" (for $w_{cn} \in S_{cn}$, $w_i = w_{cn} + p$ and $p \in S_{p\&s}$)을 집합 $S_{np}$에 추가한다. Line 4의 첫 번째 경우는, $w_{i-1}$가 이미 발견된 인물명의 집합 $S_{cn}$에 속해 있는 경우이다. 예를 들어, "해리"와 "포터"가 이미 $S_{cn}$에 존재하는 경우, "해리 포터"가 $S_{np}$에 추가된다.

Line 5의 두 번째 경우는, $w_{i-1}$의 형태소 분석 결과가 하나의 단어가 아닌 것으로 나오고 (즉, $f_m(w_{i-1})$의 값이 "*ambiguous*") $w_{i-1}$이 $S_{jx}$의 원소인 경우이다. 예를 들어, "스미스"가 이미 $S_{cn}$에 존재하고 "자카리아스"가 $S_{jx}$의 원소이면서, 형태소 분석기가 "자카리아스"를 "자(명사)+카(명사)+리아스(명사)"와 같이 하나 이상의 단어로 분석하는 경우, "자카리아스 스미스"가 $S_{np}$에 추가된다.

Line 7과 11의 세 번째와 네 번째 경우는, 두 번째 경우와 같이 $f_m(w_{i-1})$의 값이 "*ambiguous*"이면서, $w_{i-2}$가 존재하지 않거나 $f_m(w_{i-2})$의 값도 "*ambiguous*"인 경우이다. 이 경우, $w_{i-1}$가 인물명 명사구에 포함될 가능성을 추정하기 위하여, 명사구 "$w_{i-1}$ $w_{cn}$"의 등장빈도를 계산한다. 계산된 등장빈도가 임계치 계산함수 $threshold(w_{cn})$보다 큰 경우, 명사구 "$w_{i-1}$ $w_{cn}$"는 $S_{np}$에 추가된다. 임계치 계산함수 $threshold(w_{cn})$는 $w_{cn}$이 등장하는 문장의 수를 $n$이라고 했을 때, $\lfloor n/50 \rfloor$으로 계산된다. 예를 들어, "루핀"이 이미 $S_{cn}$에 존재하고 "리무스"는 한 번도 주어로 사용된 적이 없지만, $f_m$("리무스")의 값이 "*ambiguous*"이면서 "리무스 루핀"의 등장빈도가 $threshold$("루핀")보다 크면, "리무스 루핀"은 $S_{np}$에 추가된다.

Fig. 2의 Algorithm 2는 세 가지 입력으로부터 인물명 명사구 집합 $S_{np}$를 생성한다. Algorithm 2가 처리를 완료하면, 집합 $S_{np}$에 새로운 명사구가 추가되었는지를 확인하고, 추가된 명사구로부터 각 단어를 분리하여 집합 $S_{cn}$에 삽입한 뒤, 새로운 명사구가 추출되지 않을 때까지 Algorithm 2를 반복적으로 호출한다.

Algorithm 2는 $w_i$가 집합 $S_{cn}$에 존재하고 $f_m(w_{i+1})$의 값이 "*ambiguous*"인 (Line 4) 경우, $w_i$와 $w_{i+1}$의 부분문자열로 구성된 명사구 "$w_i$ $w_{i+1}.substring(0, l)$" (for $2 \le l \le w_{i+1}.length()$)을 집합 $S_{np}$에 추가한다. 형태소 분석기가 단어 $w_{i+1}$를 하나 이상의 단어로 해석하였기 때문에, $w_{i+1}$의 어느 부분문자열이 인물명이고 나머지가 조사인지 판단할 수 없다. 따라서 $2 \le k \le w_{i+1}.length()$인 $k$에 대하여, 부분문자열을 포함하는 명사구 "$w_i$ $w_{i+1}.substring(0, k)$"의 등장빈도 $freq_k$를 계산한다. 모든 $k$에 대하여 등장빈도 $freq_k$의 값이 모두 같다면, 인물명을 조사와 구별하여 추출할 방법이 존재하지 않는다. 그러나 만일 $freq_k$의 값이 서로 다르다면, 최대값 $\max(freq_k)$인 가장 큰 $k$값이 $l$이라 할 때, 명사구 "$w_i$ $w_{i+1}.substring(0, l)$"를 집합 $S_{np}$에 추가하면 된다. 예를 들어, $freq_5$("헤르미온느 그레인저가")=1, $freq_4$("헤르미온느 그레인저")=4, $freq_3$("헤르미온느 그레인")=4, $freq_2$("헤르미온느 그레")=4인 경우, "헤르미온느 그레인저"를 집합 $S_{np}$에 추가한다.

### 3.2 [단계 2] 동일인물 공통참조 클러스터링 방법

집합 $S_{cn}$은 단일 단어 형태의 인물명을 원소로 하

기 때문에, 집합 $S_{cn}$으로 그래프를 생성하면 모든 노드가 다른 노드와 전혀 연결되지 않은 그래프가 생성된다. 그러나 Algorithm 1과 2에 따라 생성된 명사구 집합 $S_{np}$를 이용하면 방향성 그래프 $G$를 생성할 수 있다.

Fig. 3은 방향성 그래프 $G$에 존재 가능한 연결 형태를 보여준다. 인물명이 집합 $S_{cn}$에만 존재하고 집합 $S_{np}$의 명사구에 포함되지 않으면, 그 인물명은 Fig. 3 (a)와 같이 연결되지 않은 노드로 존재할 것이다. 만일, "메로프 곤트"와 "곤트 씨"가 집합 $S_{np}$에 포함되어 있다면 세 개의 노드 "메로프", "곤트", "씨"는 그래프 $G$에 Fig. 3 (b)와 같이 일자형 연결 형태로 표현될 것이다. 그리고 집합 $S_{np}$에 "론 위즐리", "지니 위즐리", "퍼시 위즐리"가 존재한다면, 네 개의 노드 "론", "지니", "퍼시", "위즐리"는 그래프 $G$에 Fig. 3 (c)와 같이 역삼각형 연결 형태로 노드들이 연결될 것이다. Fig. 3 (c)와는 반대로, 집합 $S_{np}$이 "다아시 부인", "다아시 양"과 "다아시 씨"를 포함하고 있다면, 네 개의 노드 "다아시", "부인", "양", "씨"는 그래프 $G$에 Fig. 3 (d)와 같이 삼각형 연결 형태로 노드들이 연결될 것이다. Fig. 3 (e)는 노드들의 관계 표현에서 사이클이 발생하는 경우로서, 집합 $S_{np}$에 "장관 루퍼스", "루퍼스 스크림저", "스크림저 씨", "스크림저 장관"이 포함된 경우 발생하는 연결 형태이다.

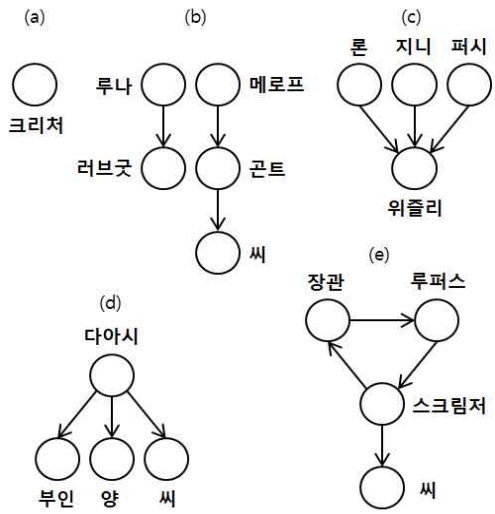


Fig. 3. Connection types in the directed graph generated by Scn and Snp.(a) Not connected, (b) Linear, (c) Inverted triangle, (d) Triangle, (e) Cycle.

이상의 연결 형태 중에서 Fig. 3 (c)의 연결 형태에서는 아래쪽 노드를 분할할 필요가 있다. 왜냐하면, 역삼각형 연결 형태에서 아래쪽 노드가 영어식 인물명의 성 (Family Name)인 경우 위쪽 노드는 인물명의 이름 (Given Name)이고, 아래쪽 노드가 직업 또는 직함 (Title, 예: 교수, 씨, 부인)인 경우 위쪽 노드는 인물명의 성이며, 아래쪽 노드가 관계 (Relation, 예: 이모, 이모부)인 경우 위쪽 노드는 인물명의 이름이기 때문이다. 따라서 Fig. 4와 같이 노드를 분할하여야 한다.

Fig. 4에서와 같이 분할된 노드는 회색으로 표시되는데, 회색은 인물명으로 단독 사용될 수 없다는 것을 의미한다. 즉, "덤블도어" 또는 "덤블도어 교수"는 특정 인물을 지칭할 수 있지만, "교수"만으로는 특정 인물을 지칭할 수 없다는 것을 의미한다. 이와 유사하게, 분할 노드가 아니더라도 주어로 사용된 적이 없는 그래프 $G$의 노드는 노드 분할 후의 그래프 $G$에서 모두 회색 노드로 표시된다. Fig 4 (b)는 역삼각형 연결 형태와 삼각형 연결 형태가 혼합된 경우에 노드 분할 예를 보여준다.

역삼각형 연결 형태와 유사하게 삼각형 연결 형태

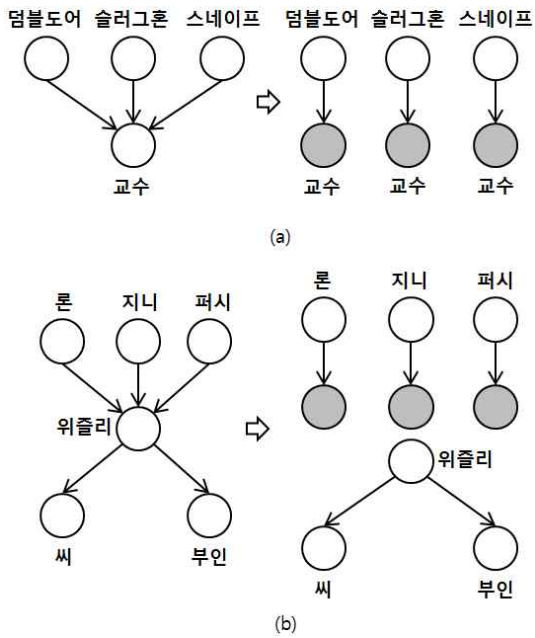


Fig. 4. Node decomposition. (a) in the type of inverted triangle connection, and (b) in the mixed type of triangle and inverted triangle connection (grey color represents a dependent node).

도 분할이 필요하다. 그러나 삼각형 연결 형태를 분할하기 위해서는, 직업 또는 직함 (Title)과 관계 (Relation)를 나타내는 단어들에 대한 의미 분석이 필요하다. 예를 들어, Fig. 4(b)와 같이 “위즐리 씨”와 “위즐리 부인”이 삼각형 연결 형태로 존재하는 경우, “씨”와 “부인”이 각각 남성과 여성에게 사용된다는 의미를 알고 있어야만 분할하는 것이 적절하다는 판단을 내릴 수 있다. 이에 반하여, “덤블도어 교수”와 “덤블도어 교수님”이 삼각형 연결 형태로 존재하는 경우, “교수님”이 “교수”의 높임말이라는 의미를 알고 있어야만 분할하지 않는 것이 적절하다는 판단을 내릴 수 있다. 이러한 의미 분석은 본 논문의 범위에 포함되지 않기 때문에, 본 논문에서는 삼각형 연결 형태에서는 노드 분할을 수행하지 않는다. 대신에, 삼각형 연결 형태로부터 생성된 이름 집합 (Name Set)은 추후 분할이 필요한 집합으로 분류한다.

따라서 본 논문의 동일인물 공통참조 클러스터링 방법에서는 역삼각형 연결 형태 노드 분할 후의 방향성 그래프 $G$에 속한 연결형 서브그래프들을 순회하면서 각각의 이름 집합을 생성한다. 예를 들어, Fig. 4 (b)의 노드 분할 후, {“론”, “론 위즐리”}, {“지니”, “지니 위즐리”}, {“퍼시”, “퍼시 위즐리”}, {“위즐리”, “위즐리 씨”, “위즐리 부인”}와 같이 4개의 이름 집합이 생성된다.

## 4. 실험 및 결과

### 4.1 실험 대상 소설

한국어로 번역된 소설에서 인물명 명사구의 동일인물 공통참조 클러스터링 방법을 실험하기 위하여, 본 논문에서는 네 권의 소설로 실험을 진행한다. 첫 번째 소설을 “조앤 K. 롤링”의 “해리 포터와 혼혈 왕자”이고, 두 번째 소설은 “제인 오스틴”의 “오만과 편견”이며, 세 번째 소설은 “아주마 나오미”의 “탐정은 바에 있다”이며, 네 번째 소설은 “나관중”의 “삼국지 (제2권)”이다. 앞의 두 권의 소설은 영어식 인물명에 대한 실험을 진행하기 위하여 선택되었으며, 세 번째 소설은 일본식 인물명에 대한 실험을 진행하기 위하여 선택되었다. 마지막 소설은 한국과 중국식 인물명에 대한 실험을 진행하기 위하여 선택되었다.

### 4.2 실험 결과 분석

Table 1은 “해리 포터와 혼혈 왕자”에 대하여 제안하는 방법이 생성한 동일인물 공통참조 클러스터링 결과를 보여준다. Table 1은 두 개의 행으로 구성되는데, 첫 번째 행은 동일인물 이름 집합으로 추정되는 것들을 나열하고 있고, 두 번째 행은 삼각형 연결 형태 또는 사이클 연결 형태를 포함하는 서브그래프를 순회하면서 생성된 이름 집합들을 보여준다.

Table 1의 첫 번째 행에서 1) {“해리”, “포터”, “해리 포터”} 이름 집합은 소설 본문에서 “해리”, “포터”, “해리 포터”가 모두 동일인물 공통참조로 볼 수 있음을 의미한다. 이에 반하여, 2) {“론”, “론 위즐리”}는 “론”과 “론 위즐리”만이 동일인물 공통참조이며, 단독으로 사용되는 “위즐리”는 “론”이나 “론 위즐리”와는 다른 인물일 수 있음을 의미한다.

Table 1 첫 번째 행의 6) {“해그리드”}와 관련하여, 소설에는 “루베우스 해그리드”가 존재하기 때문에 Algorithm 1의 세 번째 경우에 해당하여 “루베우스”를 추출하기는 하였으나, 소설 전체에서 오직 한 번 등장하기 때문에 등장빈도 규칙에 의해 명사구로 $S_{np}$에 삽입되지 못하였고, 그 때문에 이름 집합 {“해그리드”}에 “루베우스 해그리드”가 추가되지 않았다.

이와 유사하게, Table 1 첫 번째 행의 27) {“딘”}과 관련하여, 소설에는 “딘 토마스”가 존재하기는 하지만 “토마스”가 형태소 분석기에 의해 명사로 분석되기 때문에 Algorithm 2에 의해 명사구로 추출될 수 없었고, 그 때문에 이름 집합 {“딘”}에 “딘 토마스”가 추가되지 않았다. 만일, 형태소 분석기가 자신의 사전에 등록된 외국 이름이나 외래어를 명사가 아니라 외래어 등으로 표기한다면, 본 논문에서 제안하는 방법에 의해 더 많은 인물명 명사구가 추출될 것으로 기대된다.

앞서 언급한 바와 같이, Table 1 두 번째 행의 이름 집합들은 직업 또는 직함 (Title)과 관계 (Relation)를 나타내는 단어들에 대한 의미 분석에 대한 연구를 수행한 다음, 추가 분류할 필요가 있다.

Table 2는 “오만과 편견”으로부터 생성된 동일인물 공통참조 클러스터링 결과를 보여준다. Table 1과는 달리, Table 2의 첫 번째 행에 속한 이름 집합들은 대부분 하나의 원소만 가진다. 대신에, Table 2의 두 번째 행의 이름 집합은 Table 1의 이름 집합에 비하

Table 1. Name sets extracted from "*Harry Potter and the Half-Blood Prince*" written by *Joan K. Rowling*

| | Name sets |
|---|---|
| For a person | 1) {해리, 포터, 해리 포터}, 2) {론, 론 위즐리}, 3) {헤르미온느, 그레인저, 헤르미온느 그레인저}, 4) {말포이, 말포이 학생}, 5) {지니, 지니 위즐리}, 6) {해그리드}, 7) {볼드모트}, 8) {수상}, 9) {통스}, 10) {퍼지, 코렐리우스 퍼지, 퍼지 씨}, 11) {프레드}, 12) {네빌}, 13) {오그든}, 14) {루나, 러브굿, 루나 러브굿}, 15) {필요}, 16) {루핀, 리무스 루핀}, 17) {케이티}, 18) {조지}, 19) {모핀}, 20) {어머니}, 21) {크리처}, 22) {맥클라건, 코맥 맥클라건}, 23) {빌}, 24) {플뢰르}, 25) {나시사, 나시사 말포이}, 26) {마법사}, 27) {딘}, 28) {라벤더, 브라운, 라벤더 브라운}, 29) {톰}, 30) {도비}, 31) {자비니, 블레이즈 자비니}, 32) {메로프, 곤트, 메로프 곤트, 곤트 씨}, 33) {필치, 아구스 필치, 필치 씨}, 34) {헵시바, 헵시바 스미스}, 35) {고일}, 36) {집요정, 호키, 집요정 호키}, 37) {크레이브, 빈센트 크레이브}, 38) {그레이백, 펜리 그레이백}, 39) {시무스, 피니간, 시무스 피니간}, 40) {퍼시, 퍼시 위즐리}, 41) {머틀, 모우닝 머틀}, 42) {린느}, 43) {먼던구스}, 44) {뱀}, 45) {주인, 주인 아가씨}, 46) {패르바티, 패틸, 패르바티 패틸}, 47) {거미}, 48) {벅빅}, 49) {늑대인간}, 50) {팬시, 파킨슨, 팬시 파킨슨}, 51) {드멜자}, 52) {로밀다, 베인, 로밀다 베인, 베인 마법약}, 53) {그롭}, 54) {벨비}, 55) {엄브릿지, 돌로레스 엄브릿지, 엄브릿지 교수}, 56) {어니, 맥밀란, 어니 맥밀란}, 57) {두들리}, 58) {셀레스티나, 와베크, 셀레스티나 와베크}, 59) {아마커스}, 60) {나이젤러스, 피니어스 나이젤러스}, 61) {자카리아스, 자카리아스 스미스}, 62) {본즈, 수잔 본즈, 본즈 부인}. 63) {로즈메르타, 아씨오 로즈메르타, 로즈메르타 부인}, 64) {그리핀도르, 래번클로나 그리핀도르, 그리핀도르 학생}, 65) {호그와트, 호그와트 학생}, 66) {코울 부인}, 67) {핀스 부인}, 68) {올리밴더 씨}, 69) {드레이코 말포이}, 70) {말킨 부인}, 71) {폼프리 부인}, 72) {4학년 학생}, 73) {메리쏘우트 교수님}, 74) {티베리우스 씨}, 75) {히그스 씨}, 76) {루시우스 말포이}, 77) {로날드 위즐리}, 78) {폴리주스 마법약} |
| To be checked and divided | 1) {덤블도어, 덤블도어 교수님, 덤블도어 교수}, 2) {슬러그혼, 슬러그혼 교수, 슬러그혼 교수님}, 3) {스네이프, 스네이프 교수, 스네이프 교수님}, 4) {마법약, 마법약 교수, 왕자, 마법약 왕자}, 5) {벨리트릭스, 레스트랭, 벨라트릭스 레스트랭, 벨라트릭스 아가씨, 이모, 벨라트릭스 이모}, 6) {장관, 루퍼스, 장관 루퍼스, 스크림저, 루퍼스 스크림저, 스크림저 씨, 스크림저 장관}, 7) {버논, 이모부, 버논 이모부, 더즐리, 버논 더즐리, 더즐리 씨}, 8) {부엉이, 피그위존, 부엉이 피그위존, 헤드위그, 부엉이 헤드위그}, 9) {트릴로니, 트릴로니 교수, 트릴로니 교수님}, 10) {위즐리, 위즐리 부인, 위즐리 씨}, 11) {맥고나걸, 맥고나걸 교수님, 맥고나걸 교수}, 12) {프랭크 교수, 프랭크 교수님}, 13) {스프라우트 교수님, 스프라우트 교수}, 14) {디펫 교수, 디펫 교수님}, 15) {플리트윅 교수, 플리트윅 교수님} |

Table 2. Name sets extracted from "*Pride and Prejudice*" written by *Jane Austen*

| | Name sets |
|---|---|
| For a person | 1) {캐서린}, 2) {아버지}, 3) {어머니}, 4) {샬롯}, 5) {키티}, 6) {삼촌}, 7) {아시, 아시 씨}, 8) {메리}, 9) {피츠윌리암, 대령, 피츠윌리암 대령, 대령 부인}, 10) {아내}, 11) {하인}, 12) {가정부}, 13) {아가씨}, 14) {주인}, 15) {룽 부인}, 16) {엘라이저 양}, 17) {앤즐리 부인}, 18) {그랜틀리 양}, 19) {젠킨슨 부인}, 20) {니콜즈 부인} |
| To be checked and divided | 1) {엘리자베스, 베네트, 엘리자베스 베네트, 베네트 부인, 베네트 양, 베네트 씨, 엘리자베스 양}, 2) {조지아나, 조지아나 양, 다아시, 조지아나 다아시, 부친, 다이시 부친, 다아시 양, 다아시 군, 다아시 씨}, 3) {제인, 언니, 제인 언니, 제인 양}, 4) {캐롤라인, 빙리, 캐롤라인 빙리, 빙리 군, 빙리 씨, 빙리 부인, 빙리 양}, 5) {위컴, 위컴 군, 위컴 씨}, 6) {콜린즈, 콜린즈 씨, 콜린즈 부인}, 7) {딸, 마리아, 딸 마리아, 버어그, 딸 버어그, 경, 버어그 경, 경 부인, 버어그 양} 8) {가디너, 가디너 부인, 가디너 씨}, 9) {허스트, 허스트 씨, 허스트 부인} |

여 더 많은 원소를 가진다. 이러한 결과가 나온 이유는 "오만과 편견"이 "해리포터와 혼혈 왕자"와 다른 장르의 소설로서 가족 간의 관계를 나타내는 호칭이 많이 사용되었기 때문이다.

앞서 언급한 바와 같이, 본 논문에서는 "경"과 "부인"과 같은 직업 또는 직함 (Title)의 의미 분석을 수행하지 않았기 때문에, Table 2 두 번째 행의 7) {"딸", "마리아", "딸 마리아", "버어그", "딸 버어그", "경", "버어그 경", "경 부인", "버어그 양"} 이름 집합으로부터 "경 부인"과 같은 잘못된 명사구가 포함되는 결과가 발생하였다. 이러한 오류로부터 직업 또는 직함 (Title)과 관계 (Relation)를 나타내는 단어들에

Table 3. Name sets extracted from "*The Detective is in the Bar*" written by *Azuma Naomi*

| | Name sets |
|---|---|
| For a person | 1) {먼로, 먼로 씨}, 2) {하루, 하루 씨}, 3) {레이코, 레이코 씨}, 4) {아저씨}, 5) {하라다, 하라다 씨}, 6) {다카다}, 7) {마스터}, 8) {마쓰오, 마쓰오 씨}, 9) {히로}, 10) {아줌마}, 11) {기리하라, 기리하라 씨}, 12) {쇼코}, 13) {꼬마}, 14) {오카모토, 오카모토 씨}, 15) {미에코}, 16) {니시다}, 17) {쓰레기}, 18) {아이다}, 19) {꼬맹이}, 20) {점장}, 21) {마마}, 22) {치프}, 23) {에미코}, 24) {호객꾼}, 25) {직원}, 26) {사쿠라이}, 27) {케이키치, 케이키치 씨}, 28) {세이코 씨}, 29) {곤노 씨}, 30) {미무라 씨}, 31) {스미카즈 씨} |
| To be checked and divided | 1) {쇼지 씨, 마사하루, 쇼지 마사하루, 마사하루 씨} |

대한 의미 분석 연구가 필요함을 확인할 수 있다.

Table 3는 일본어 소설 "탐정은 바에 있다"으로부터 생성된 동일인물 공통참조 클러스터링 결과를 보여준다. 일본에서도 영어식 이름과 같이 성과 이름을 명사구로 표현하기는 하지만, 그와 같이 표현하는 빈도수가 앞서 살펴본 세 권의 영어 소설에 비하여 월등히 작음을 알 수 있다. 예를 들어, Table 3 두 번째 행의 1) {"쇼지 씨", "마사하루", "쇼지 마사하루", "마사하루 씨"} 이름 집합으로부터 "쇼지 마사하루"와 같은 성과 이름을 명사구로 표현하는 경우가 단 한 번 추출되었다. 소설에 따라 달라질 수는 있겠지만, Table 3으로부터 일본어 소설에서는 성과 이름을 명사구로 표현하기 보다는 "먼로 씨", "하루 씨", "레이코 씨"와 같이 직업 또는 직함 (Title)을 나타내는 단어를 더 많이 사용함을 알 수 있다. 그런데, 영어 소설과는 달리, 이름 다음에도 "씨"를 붙이는 특성을 가지고 있으며, 남성과 여성을 구분하지 않고 "씨"를 붙이는 특성을 가지고 있다. 이러한 특성은 추후 직업 또는 직함 (Title)과 관계 (Relation)를 나타내는 단어들에 대한 의미 분석 연구에서 참고할 필요가 있다.

Table 4는 중국어 소설 "삼국지 (제2권)"에 대하여 제안하는 방법이 생성한 동일인물 공통참조 클러스터링 결과를 보여준다. 중국식 이름은 한국식 이름과 마찬가지로 성과 이름이 공백 없이 붙어서 사용되며 두 글자 또는 세 글자로 구성되는 특성을 가진다. 따라서 "삼국지 (제2권)"에서는 이름 집합의 대부분이 하나의 원소를 가지는 것이 일반적이며, Table 4도 그러한 결과를 보여주고 있다. 예외적인 경우로, Table 4 두 번째 행의 1) {"오의", "손권", "오의 손권", "오의 군사"} 이름 집합과 같은 결과를 얻게 된 이유는, "오의"라는 인물명이 소설 내에 존재하는데 "오나라의 손권"을 "오의 손권"이라고 작가가 표현함으로써, "오의 손권"이 Algorithm 1의 첫 번째 경우에 의해 추출되었기 때문이다. 그 외, Table 4 두 번째 행의 2) {"부하", "감녕", ..., "장수 한호"}와 같은 이름 집합이 생성된 이유는 "부하"와 "장수"를 직업

Table 4. Name sets extracted from "*Romance of the Three Kingdoms (part 2)*" written by *Lou Guanzhong*

| | Name sets |
|---|---|
| For a person | 1) {조조, 조조 군사}, 2) {현덕}, 3) {공명, 공명 군사}, 4) {주유, 주유 군사}, 5) {장비}, 6) {마초}, 7) {황충}, 8) {노숙}, 9) {조운}, 10) {관우}, 11) {부친, 유장, 부친 유장}, 12) {장합}, 13) {유비}, 14) {위연}, 15) {방통}, 16) {부인}, 17) {병사}, 18) {장로}, 19) {황개}, 20) {조홍}, 21) {조인}, 22) {한수}, 23) {사자}, 24) {허저}, 25) {장간}, 26) {좌자}, 27) {엄안}, 28) {유기}, 29) {하후연}, 30) {장송}, 31) {장임}, 32) {서황}, 33) {관로}, 34) {감택}, 35) {장로}, 36) {방덕}, 37) {법정}, 38) {유표}, 39) {정보}, 40) {능통}, 41) {냉포}, 42) {제갈량}, 43) {양송}, 44) {장소}, 45) {우금}, 46) {마대}, 47) {사나이}, 48) {양군}, 49) {뇌동}, 50) {하후상}, 51) {군선}, 52) {이전}, 53) {왕필}, 54) {손건}, 55) {정봉}, 56) {헌제}, 57) {서서}, 58) {무사}, 59) {채중}, 60) {채화}, 61) {조자룡}, 62) {태사자}, 63) {한현}, 64) {등현}, 65) {마량}, 66) {악진}, 67) {유봉}, 68) {관평}, 69) {김의}, 70) {가후}, 71) {간옹}, 72) {양수}, 73) {진식}, 74) {조안}, 75) {조휴}, 76) {주태}, 77) {맹달}, 78) {순욱, 고문관인 순욱}, 79) {조식}, 80) {교국로}, 81) {송겸}, 82) {장숙} |
| To be checked and divided | 1) {오의, 손권, 오의 손권, 오의 군사}, 2) {부하, 감녕, 부하 감녕, 부하 한호, 장수, 부하 장수, 하후돈, 장수 하후돈, 황조, 장수 황조, 장수 한호} |

또는 직함 (Title)으로 인식하지 못하였기 때문이다. 이는 추후 직업 또는 직함 (Title)과 관계 (Relation)를 나타내는 단어들에 대한 의미 분석 연구에 의해 해결될 것으로 예상된다. 이와는 달리, Table 4 첫 번째 행의 78) {"순욱", "고문관인 순욱"} 이름 집합에서 "고문관인 순욱"이 명사구로 추출된 이유는 형태소 분석기의 오류 때문이다. 형태소 분석기가 "고문관"은 명사로 인식하지만 "고문관인"은 "고문(명사)"와 "관인(명사)"로 나누어 인식하기 때문에 "고문관인"이 형태소 분석이 불가능한 단어로 판단되어 Algorithm 1에 의해 이름 집합에 추가된 것이다. 명사구 "고문관인 순욱"은 소설에서 주어로 오직 한 번 등장할 뿐만 아니라 단어 "고문관인"이 단독 주어로 사용되지 않기 때문에, 발화자 인식 및 인물간 소셜 네트워크 구축에 영향을 미치지 않을 것으로 예상되지만, 해당 연구에서 영향을 분석할 필요가 있다.

이상의 결과로부터 제안하는 방법이 소설에 등장하는 인물명 명사구에 대하여 동일인물 공통참조 클러스터를 적절하게 생성함을 확인하였다. 이와 더불어, 직업 또는 직함 (Title)과 관계 (Relation)를 나타내는 단어들에 대한 의미 분석 연구가 추후 진행되어야 함을 확인하였는데, 그러한 연구는 발화자가 대명사인 경우 대명사가 의미하는 인물명을 추론하는 Anaphora Resolution Problem [1, 3]의 해결을 위해서도 필수적이기 때문에, 소설에서의 발화자 인식 연구와 더불어 진행될 필요가 있다. 이와 같이 생성된 동일인물 공통참조 클러스터는 소설의 특정 부분에서 어떤 인물들이 동일 공간에 존재하며 어떤 인물이 다른 어떤 인물에게 발화 (Utterance)를 하는지를 분석하는데 필수적인 역할을 할 뿐만 아니라, 소설로부터 추출되는 인물간 소셜 네트워크에서 동일인물이 여러 개의 다른 노드로 표현되는 것을 방지하는데 큰 역할을 할 것으로 기대된다.

## 5. 결 론

본 논문에서는 한국어로 번역된 소설에서 인물명 명사구를 추출하고, 추출된 명사구로부터 동일인물 공통참조를 클러스터링하는 방법을 제안하였다. 제안하는 방법은 인물명으로 알려진 단어를 포함하여 앞, 뒤 총 4단어로부터 인물명 명사구를 추출하고, 추출한 명사구로부터 방향성 그래프를 구축한 뒤, 연결 서브그래프를 순회함으로써 동일인물 공통참조 클러스터를 생성하였다. 제안하는 방법의 실험을 위하여 두 권의 영어 소설과 한 권의 일본어 소설, 한 권의 중국어 소설의 한국어 번역본을 사용하였다. 네 권의 번역 소설을 대상으로 하는 동일인물 공통참조 클러스터링 실험을 통하여, 언어별 소설의 특성을 파악하게 되었고, 직업 또는 직함 (Title)과 관계 (Relation)를 나타내는 단어들에 대한 의미 분석 연구가 진행되어야 함을 확인할 수 있었다. 향후에는 향상된 동일인물 공통참조 클러스터링 방법을 활용하여, 한국어 소설에서 발화자를 인식하는 연구를 진행하고자 한다.

## REFERENCE

[1] D.K. Elson and K.R. McKeown, "Automatic Attribution of Quoted Speech in Literary Narrative," *Proceedings of the 24th AAAI Conference on Artificial Intelligence*, pp. 1013-1019, 2010.

[2] D.K. Elson, N. Dames, and K.R. McKwown, "Extracting Social Networks from Literary Fiction," *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pp. 138-147, 2010.

[3] E. Iosif and T. Mishra, "From Speaker Identification to Affective Analysis: A Multi-Step System for Analyzing Children' Stories," *Proceeding of the 3rd Workshop on Computational Linguistics for Literature*, pp. 40-49, 2014.

[4] Stanford CoreNLP-A Suite of Core NLP Tools, http://nlp.stanford.edu/software/corenlp.shtml, (accessed Nov., 28, 2016).

[5] T. Park and S. H. Kim, "A Character Identification Method Using Postpositions for Animate Nouns in Korean Novels," *Journal of Information Technology Services*, Vol. 15, No. 3, pp. 115-125, 2016.

[6] T. Park and S.H. Kim, "A Character Identification Method Utilizing Connective and Possessive Forms of Animate Nouns in Novels Translated into or Written in Korean," *IEICE Transactions on Information and*

*Systems*, 2016.

[7] E.Y. Lee, "Named Entity Detection and Relation Extraction in the Personal Chronology of the 19th Century," *Journal of EONEOHAG*, Vol. 53, pp. 141-162, 2009.

[8] G.M. Park, S.H. Kim, and H.G. Cho, "Analysis of Social Network According to the Distance of Character Statements," *Journal of the Korea Contents Association*, Vol. 13, No. 4, pp. 427-439, 2013.

[9] B.H. Back, I. Ha, and B.C. Ahn, "An Extraction Method of Sentiment Information from Unstructured Big Data on SNS," *Journal of Korea Multimedia Society*, Vol. 17, No. 6, pp. 671-680, 2014.

[10] D.J. Lee, J.H. Yeon, I.B. Hwang, and S.G. Lee, "KKMA: A Tool for Utilizing Sejong Corpus Based on Relational Database," *Journal of KIISE: Computing Practices and Letters*, Vol. 16, No. 11, pp. 1046-1050, 2010.

박 태 근

1991년 포항공과대학교 컴퓨터공학과(학사)
1993년 포항공과대학교 컴퓨터공학과(석사)
2004년 포항공과대학교 컴퓨터공학과(박사)

1996년~2000년 SK Telecom 중앙연구원 선임연구원
2000년~2001년 3Com Korea 과장
2001년~2002년 Ericsson Korea 차장
2004년~현재 단국대학교 응용컴퓨터공학과 교수
관심분야 : 데이터컴퓨팅, IoT, 가상화서비스, 분산서비스

김 승 훈

1985년 인하대학교 전자계산학과(학사)
1989년 인하대학교 전자계산학과(석사)
1998년 포항공과대학교 컴퓨터공학과(박사)

1989년~1990년 ETRI
1991년~1993년 POSDATA
2001년~현재 단국대학교 응용컴퓨터공학과 교수
관심분야 : 데이터컴퓨팅 및 네트워킹, IoT, 분산시스템