

NFC 태그 정보를 이용한 검색 정보의 군집 시스템 모델

(Clustering System Model of Information Retrieval using NFC Tag Information)

박 선*, 김형균**, 심수정***

(Sun Park, HyeongGyun Kim, Su-Jeong Sim)

요약

NFC 스마트폰의 보급 증가는 앱과 연계하여 다양한 서비스를 제공하고 있으며, 단순한 인터넷 서비스를 개인화 서비스로 변화 시킬 것으로 예상되고 있다. 본 논문은 정보 접근을 위한 NFC 태그의 정보를 이용하여 유사정보를 활용할 수 있도록 검색 정보를 군집하는 시스템 모델을 제안한다. 제안된 모델은 NFC 태그에서 제공하는 정보를 이용하여 유사 정보를 검색할 수 있다. 또한 검색된 유사정보를 사용자가 활용할 수 있도록 주제별 군집할 수 있다.

■ 중심어 : NFC ; 문서군집 ; 정보검색 ; 개인화 서비스 ; 의미 특징

Abstract

The growth of the propagated NFC provides the various services with respect to internet applications, which it can be predicted from the simple internet services to the privated services. This paper proposes the clustering of information retrieval system model using NFC tag of access information for utilizing the similar information of the tag. The proposed model can search the similar information of the tag using the access information of NFC tag. In addition, it can cluster the similar retrieval information into topic cluster for utilizing users.

■ keywords : NFC ; document clustering ; information retrieval ; privated services ; semantic feature

I. 서론

스마트폰은 단순히 연락을 위한 휴대폰에서 이메일, 일정관리, 네비게이션, 오락, 멀티미디어 기기 등 다양한 서비스를 즐길 수 있는 다기능 휴대폰으로 발전하였다. 최근 NFC(near field communication) 기술이 스마트폰에 접목되면서 버스나 지하철의 교통카드 이용, 신용카드 및 멤버십 카드 이용, 상점의 물품정보 제공, 여행정보, 지역정보, 교통 등에 정보를 쉽고 빠르게 접근하여 이용할 수 있다.

NFC는 10Cm 이내에서 100Kbs~1Mbps의 전송속도를 가지는 근거리 통신으로 카드 에뮬레이션, 카드 리더, P2P(peer to peer)기능을 지원하고 있다. NFC의 주요 활용 가능 서비스로는 스마트 단말기간 접촉응용 서비스로 결제, 금융 이체, 개인 정보 교환 페어링 등을 지원하며, 개인정보 관리 응용 서비스로 개인인증, 원격통제, 관광안내등을 지원하고, 정보제공 및 각종 맞춤형 광고관련 응용서비스로 의료정보 제공, 주차정보 제공, 예약, 위치 정보 제공, 제품 정보제공, 콘텐츠 구매 지원, 소셜

네트워크 연계 서비스 등을 지원 할 수 있다[1].

스마트폰에 NFC의 지원으로 생활 곳곳에 NFC 태그에 부착될 것이며, 사용자들은 태그를 접촉함으로써 그들이 원하는 서비스나 정보를 제공받을 수 있다. 현재의 NFC 기술은 단순 지불, 지역이나 상품의 단순 정보제공에 국한되어 사용되고 있다. 이러한 단순 정보제공 서비스는 인터넷 검색만으로도 충분히 손쉽게 얻을 수 있으므로, 아직은 NFC나 NFC 태그를 활용한 새로운 서비스나 다양한 부가서비스는 많이 부족한 상태이다. NFC 태그 접촉을 통한 단순 정보만 이용되는 환경에서 사용자들이 NFC 태그를 터치할때 관련된 정보를 검색하여 이용할 수 있으면 NFC 서비스가 매우 핵심적인 생활 서비스로 등장할 가능성이 있다[1].

본 논문은 NFC 태그에 의해 접근되어 제공하는 정보와 관련된 정보를 검색하여 실제 필요한 유사정보를 활용할 수 있도록 검색된 정보를 군집하는 시스템 모델을 제안한다. 제안된 모델은 NFC 태그에서 제공하는 정보를 이용하여 유사 정보를 검색할 수 있으며, 검색된 유사정보를 사용자가 활용할 수 있도록 문서에 포함된 의미특징을 이용하여 주제별로 군집한다.

* 목포대학교 정보산업연구소

** 조선대학교

*** (주)동하테크

본 논문의 구성은 다음과 같다. 2장에서는 관련연구로 NFC를 이용한 활용서비스와 문서군집에 대하여 알아본다. 3장은 본론으로 제안된 NFC 태그정보를 이용한 검색정보의 군집 시스템 모델에 대하여 설명한다. 4장에서는 결론에 대하여 기술한다.

II. 관련연구

1. NFC

진동수는 NFC 기반의 전자상거래 비즈니스 모델을 연구하였다. 여기서 NFC와 비즈니스 모델, 사례연구방법론 주요 개념에 대한 문헌 연구를 통하여 대표적인 NFC 상거래 사례들을 선택하고, 해당 사례의 성공과 실패에 영향을 미치는 주요 요인들에 대하여 연구하였다[1]. 이정전과 홍성우는 NFC 기반 세렌디피티 시스템을 설계하였다. 세렌디피티는 사용자가 NFC 태그를 터치할때 의외의 가치를 발견하는 것으로, 이를 위해서 태그 터치 데이터베이스, 세렌디피티 사례 베이스 데이터베이스, 세렌디피티 규칙 패턴 데이터베이스, 사용자 프로파일 데이터베이스를 구성하여 세렌디피티를 시스템과 활용 시나리오를 제안하였다[2]. 최동원의 저자들은 클라우드 환경에서 NFC를 활용한 정보공유 서비스에 대하여 연구하였다. 이들은 스마트폰과 클라우드 서버를 동기화시키는 정보를 태그에 저장하고, 태그를 터치함으로써 클라우드 서버에 접근하여 자료를 동기화할 수 있는 방법을 제안하였다[3]. 박재영의 저자들은 NFC 모바일 폰을 활용한 도서관 시스템 구축방법을 제안하였다. 이들의 제안방법은 NFC의 읽기/쓰기 모드와 휴대폰의 고유번호를 이용하여 신분증을 대체하는 방법을 제시하였으며, NFC의 메시지 프로토콜을 활용하여 시스템 구성요소간의 통신 방법의 구성을 제안하였다[4]. 배경재와 박승진은 시각장애인의 정보접근성 향상을 위한 유비쿼터스 도서관 구축에 관한 연구를 하였다. 이들은 시각장애인의 고유 정보와 상황을 감지해서 네트워크에 쉽게 연결할수 있도록 NFC를 사용하였으며, 도서관 전용 휴대폰의 NFC가 PC 동글을 터치하여 자동 사용자 인증을 통한 콘텐츠를 다운받을 수 있도록 하였다[5].

2. 문서군집

외부지식 기반의 문서군집방법은 어휘사전인 워드넷(WordNet), 단어 관계 사전인 온톨로지(ontology, 공유된 개념화), 온라인 백과사전인 위키피디아(wikipedia) 등의 외부지식을 이용하여 문서내의 용어를 확장하여 군집의 성능을 향상시킨다. 이들방법은 정보손실문제, 온톨로지 구축 비용문제가

있다[6-10]. 내부지식 기반의 문서군집방법은 문서집합의 내부의 구조를 나타내는 의미특징(semantic feature)을 이용하여 군집의 주제에 적합한 문서들로 쉽게 군집할 수 있다. 그러나 문서집합의 구성 문서들이 유사한 특성을 갖거나, 극단적으로 다른 특성을 갖고 있으면 추출된 의미특징들의 문서집합의 내부 구조를 충분하게 반영할 수 없으므로 좋은 군집 결과를 얻기 힘들다[11-17]. 내외부 복합지식 기반방법은 외부지식 기반방법과 내부지식 기반방법의 장점을 이용한 복합적인 방법이다. 이 방법은 외부지식 기반방법이나 내부지식 기반방법에 비하여 좋은 군집효율을 보이나, 내부지식 기반방법과 마찬가지로 군집 내부의 자료특성에 영향을 많이 받는다[18, 19].

외부지식 기반의 문서군집 방법은 다음과 같다. Trappey의 저자들은 특허문서를 군집하기 위하여 퍼지에 기반을 둔 온톨로지 방법을 제안하였다[6]. Hu와 저자들은 위키피디아의 개념과 문서집합을 연결하여 위키피디아의 분류(category)를 문서군집에 이용할 수 있도록 제안하였다[7]. Huang와 저자들은 위키피디아를 기반으로 활성학습(active learning)하여 문서를 군집하는 방법을 제안하였다. 그러나 이들의 방법은 군집하기 이전에 사전작업으로 위키피디아를 학습해야한다[8]. Huang와 저자들은 위키피디아 기반의 개념의 표현을 이용하여 문서를 군집하는 방법을 제안하였다[9]. Kiran와 저자들은 연관규칙과 위키피디아를 이용한 계층적 문서군집방법을 제안하였다[10].

내부지식을 이용한 문서군집에 대한연구는 다음과 같다. 의미특징에 기반을 둔 문서군집의 최근 연구로 Li 이외 저자들은 문서군집과 관련된 군집의 하위 공간구조의 특징을 이용한 ASI(Adaptive Subspace Iteration) 알고리즘을 제안하였다[11]. Wang과 Zhang은 문서군집을 위하여 지역 레이블과 전역 레이블의 특징을 이용한 CLGR(Clustering with Local and Global Regularization) 알고리즘을 제안하였다[12]. Xu 이외 저자들은 비음수 행렬 분해(NMF, Non-negative Matrix Factorization)의 의미특징을 이용하여 문서를 군집하는 방법을 제안하였다[13]. 본 논문의 저자들은 이전에 의미특징에 기반 문서군집방법, 내외부 복합지식 기반방법인 의미특징과 외부지식기반의 문서군집방법을 제안하였다. 의미특징에 기반을 둔 방법으로는 의미특징과 군집의 응집도를 이용한 방법[14, 15], 의미특징과 퍼지관계를 이용한 방법[16], 마지막으로 주성분 분석과 퍼지연관을 이용한 방법[17]이 있다. 이들 방법은 의미특징에 기반을 두고 있기 때문에 구성 문서의 특성이 극단적으로 유사하거나 다르면 군집의 성능이 좋지 않을 수 있는 문제를 가지고 있다. 이전 저자들이 제안한방법의 문제점을 해결하기 위해서 내외부 복합지식 기반방법인 군집 주제의 유의어와 유사도를 이용한 문서군집방법[18]과 군집의 중요 용어와 위키피디아를 이용한 문서군집[19]을 제안하였다. 그러나 이 방법 역시 내부 자료의 구성에 따라서 군집의 성능에 영향을 미치는 문제

점을 가지고 있다.

III. 본 론

본 장에서는 제안 시스템 모델에 대하여 설명한다. 제안된 시스템 모델은 NFC 태그에 의해서 접근되는 정보를 이용하여 관련정보를 검색하고, 유사정보를 활용할 수 있도록 검색된 정보를 군집하는 시스템 모델이다. 본 논문에서 제안한 시스템 모델은 NFC 장치, 군집 서버로 구성된다.

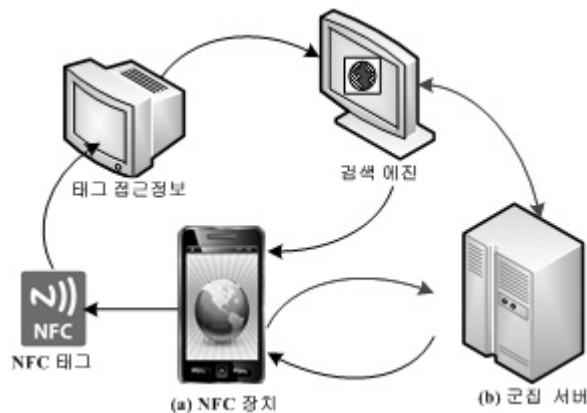


그림 1. 제안된 시스템의 구성도

1. NFC 장치

NFC 장치(스마트폰)는 크게 두가지 역할을 수행한다. 첫째 역할은 NFC 태그의 접근정보를 검색엔진에 전달하여 NFC 접근정보와 관련된 검색된 정보를 스마트폰에 보여준다. 두번째 역할은 유사정보를 활용할 수 있도록 검색정보를 주제별로 군집한 결과를 군집 서버로 부터 받아 화면에 출력한다.

NFC 장치는 다음과 같이 동작한다. 첫째, 사용자가 NFC 태그를 터치하면 터치 정보와 연결된 인터넷상의 NFC 태그 연결 정보를 스마트폰에 표시해준다. 이와 동시에 NFC 연결 정보를 전처리하여 검색어를 추출하고, 추출한 검색어와 외부 검색엔진을 이용하여 관련된 정보를 검색하여 사용자에게 검색결과를 검색엔진의 검색순위순으로 보여준다. 사용자가 검색된 결과인 검색문서를 만족하지 못하면, 검색엔진에서 검색된 결과를 군집서버에 보내어 주제별로 군집하여 스마트폰에 표시해준다.

NFC 장치의 위의 두가지 역할을 동작하는 관리프로그램은 자바와 안드로이드 프로그램을 기반으로 작성하여 현재 가장 많이 보급된 NFC 부착장비인 안드로이드 기반 스마트폰에서

사용할 수 있도록 한다.

2. 군집 서버

군집 서버는 검색엔진으로부터 검색된 NFC 접근정보와 관련된 검색정보를 주제별로 군집하여 사용자에게 제공한다. 이를 위해서 검색정보로부터 군집의 주제(즉, 분류명, 레이블)를 추출해야하며, 추출된 주제를 기준으로 접근정보를 군집해야 한다.

본 논문에서는 비음수 행렬분해[20]를 이용하여 주제출 추출하고 검색된 문서들을 군집한다. 비음수 행렬분해(NMF, non-negative matrix factorization)는 Lee와 Seung이 제안한 행렬분해방법으로 원본자료를 두 개의 의미 특징 행렬로 분해하여 부분정보의 선형 조합으로 원본자료를 표현할 수 있는 방법이다. 이 방법은 정보의 구조 안에 포함하고 있는 의미특징을 쉽게 파악할 수 있고, 차원 축소를 통하여서 대량의 정보를 효율적으로 표현할 수 있게 한다[20].

가. 군집 주제 추출방법

(1) 전처리

군집의 주제를 추출하기 위해서는 전처리로 검색된 정보(문서)를 이용하여 용어-문서 빈도행렬을 구축한다. 용어-문서 행렬을 구축하기 위해서는 문서로부터 용어들을 추출한 다음에 문서에서 불필요한 용어를 제거하는 불용어(stop-word) 작업과 용어들의 의미를 일치시키는 어근(stemming)추출 작업은 진행한다. 영어문서의경우 불용어 제거는 Rijsbergen의 불용어 목록[21]을 이용하여서 목록에서 정의하고 있는 무의미한 용어들을 제거한다. 어근추출은 Porter의 어근추출 알고리즘[26]을 이용하여서 영어의 파생어들을 가장 중심이 되는 용어인 어근으로 변환한다. 마지막으로 추출된 용어가 문장에서 출현하는 빈도를 이용하여 용어-문서 빈도행렬을 구축한다. 한글문서의 경우 형태소분석[22]을 통하여 명사만을 추출하여 용어-문서 빈도행렬을 생성한다.

(2) 군집 주제 추출

본 논문에서 제안한 군집의 주제를 추출하는 방법은 다음과 같다. 비음수행렬 분해 식(1)을 용어-문서 빈도행렬에 적용하여 비음수행렬 의미특징행렬 W 와 비음수 의미변수행렬 H 를 계산한다. 군집의 주제를 추출하기 위하여 비음수 의미특징 행렬 W 를 이용하여 각각의 군집 안에 포함된 의미특징의 값이 가장 큰 의미특징과 대응되는 용어를 군집의 주제로 선택한다. 여기서 의미특징 값이 크다는 것은 대응되는 용어가 중요하다는 것을 의미한다.

$$A \approx WH \quad (1)$$

비음수 행렬 식(1)의 목적은 행렬 A 를 비음수 $i \times r$ 행렬 W 와 비음수 $r \times j$ 행렬 H 로 분해하는 것이다. 여기서, A 는 i 개의 용어와 j 개의 문서로 구성된 $i \times j$ 행렬이고, r 은 의미특징행렬의 크기를 결정할 수 있는 의미특징의 개수이다. 본 논문에서 의미특징의 개수 r 은 문서의 군집의 개수를 설정하기 위해 사용된다. 즉, 사용자가 원하는 군집의 개수를 설정하면 된다. 예를 들어 사용자가 3개의 군집을 원하면 $r=3$ 으로 설정하면 된다. 또한 두 개의 비음수 의미 특징 행렬을 구별하기 위하여, 비음수행렬분해 알고리즘을 제안한 Lee와 Seung은 두 행렬 W 와 H 를 의미특징 행렬 W 와 의미변수행렬 H 로 각각 이름을 정의하였다[20].

$$J = \frac{1}{2} \|X - WH\| \quad (2)$$

여기서 $W = [w_{ij}]$ 이고 $H = [h_{ij}]$ 이며 $W = [W_1, W_2, \dots, W_d]$ 이다. W 와 H 의 원소 값을 갱신하기 위하여 목적함수 J 값이 수렴 허용오차 보다 작아지거나 지정한 반복횟수를 초과할 때까지 식(3)과 식(4)를 반복한다[20].

$$w_{ij} \leftarrow w_{ij} \frac{(XH^T)_{ij}}{(WHH^T)_{ij}} \quad (3)$$

$$h_{ij} \leftarrow h_{ij} \frac{(W^T X)_{ij}}{(W^T WH)_{ij}} \quad (4)$$

여기서 H^T 는 H 의 전치행렬이고, W^T 는 W 의 전치행렬이다.

나. 검색정보 군집 방법

본 논문에서는 검색정보를 주제별로 군집하기 위하여 Xu등이 제안한 군집방법[23]을 이용하여 다음과 같이 군집한다. 용어-문서 빈도행렬 A 에 식(1)을 이용하여 비음수 행렬 분해를 수행하여서 비음수 의미특징 행렬 W 와 비음수 의미 변수 행렬 H 를 얻는다. 식(5)와 식(6)을 이용하여 행렬 W 와 H 를 정규화한다. 행렬 H 를 이용하여 문서를 군집한다. 예를 들어, 만약 $x = \arg\max_i h_{ij}$ 이면 j 번째 군집의 j 번째 의미특징 요소에 대응되는 문서를 분류 x 에 할당한다.

$$w_{ij} \leftarrow \frac{w_{ij}}{\sqrt{\sum_i w_{ij}^2}} \quad (4)$$

$$h_{ij} \leftarrow h_{ij} \sqrt{\sum_i w_{ij}^2} \quad (5)$$

다. 제안방법의 사례 연구

본 장에서는 사례 연구를 통하여 제안방법의 군집 주제 추출 방법과 문서 군집방법에 대하여 설명한다.

표1은 NFC 태그의 접근정보를 이용하여 관련된 정보를 검색 엔진을 통하여 검색한 7개의 문서를 전처리하여 용어-문서 빈도행렬을 구성한 것이다. 표2는 표1의 용어-문서 빈도행렬에 의미특징(군집의 개수)을 3으로 설정하여 식(1)의 비음수 행렬 분해 하여 비음수 특징행렬 W 와 비음수 의미변수행렬 H 를 계산한 결과이다.

표 1. 7개의 검색 문서에 대한 용어-문서 빈도행렬

용어 \ 문서	d1	d2	d3	d4	d5	d6	d7
t1	0	0	5	4	1	0	0
t2	0	0	1	2	0	1	0
t3	0	0	0	0	4	3	7
t4	0	0	0	0	3	6	2
t5	2	4	0	0	0	0	0
t6	1	4	0	0	0	0	0

표2는 표1에 비음수 행렬분해 식(1)을 이용하여

표2(a)의 의미특징행렬 W 를 이용하여 군집의 주제를 추출한다. 즉, 첫번째 군집(st1)에서는 가장 큰 의미특징을 갖는 t3(8.263)가, 두번째 군집(st2)에서는 가장 큰 의미특징을 갖는 t2(6.475)가, 세번째 군집(st3)에서는 가장 큰 의미특징을 갖는 t5(4.449)가 각각의 군집을 대표하는 용어들로 군집의 주제로 추출된다.

표2(b)의 의미변수행렬 H 를 이용하여 검색문서를 군집한다. 즉, 첫번째 군집(st1)에는 문서 각각에서 가장 큰 의미특징 값을 갖는 문서 d5, d6, d7이, 두번째 군집(st2)에서는 문서 각각에서 가장 큰 의미특징 값을 갖는 문서 d3, d4가, 세번째 군집(st3)에서는 문서 각각에서 가장 큰 의미특징 값을 갖는 문서 d1, d2가 군집된다.

표 2. 표1의 비음수 행렬 수행 결과

(a) 비음수 의미특징 행렬 W

의미특징 용어	st1	st2	st3
t1	0	6.475	0
t2	0.407	2.052	0
t3	8.263	0	0
t4	6.242	0.091	0
t5	0	0	4.449
t6	0	0	4.094

(b) 비음수 의미변수 행렬 H

문서 의미특징	d1	d2	d3	d4	d5	d6	d7
st1	0	0	0	0	0.481	0.582	0.655
st2	0	0	0.747	0.651	0.132	0.039	0
st3	0.335	0.9347	0	0	0	0	0

표3은 제안방법을 이용하여 NFC 태그와 관련된 검색정보에 관련된 문서에 대한 군집의 주제 추출과 문서의 군집 결과를 보여준다.

표 3. 제안방법을 이용한 군집주제 및 군집 결과

군집 주제	군집 문서
t3	d5, d6, d7
t1	d3, d4
t5	d1, d2

IV. 결 론

본 논문은 NFC 태그에 의해 접근되어 제공하는 정보와 관련된 정보를 검색하여 실제 필요한 유사정보를 활용할 수 있도록 검색된 정보를 주제별로 군집하는 시스템 모델을 제안하였다. 제안된 시스템 모델은 다음과 같은 장점을 갖는다. 첫째, NFC 태그에서 제공하는 정보와 검색엔진을 이용하여 유사 정보를 검색하여 스마트폰 사용자에게 제공함으로써 사용자에게 다양한 유사한 서비스의 선택 기회를 부여할 수 있다. 둘째, 검색된 유사정보를 사용자가 활용할 수 있도록 문서에 포함된 의미특징을 이용하여 주제별로 군집하여 사용자에게 제공한다. 문서에 포함된 의미특징을 이용하여 군집의 주제별로 군집함으로써 사용자가 예상하지 못한 상황에서 사용자에게 유용한 정보를 제공할 수 있다.

Acknowledgement

본 연구는 지식경제부 및 한국산업기술평가관리원의 산업융합원천기술개발사업 (정보통신)의 일환으로 수행하였음. [10041057, 휴대단말 기반의 RFID 서비스 산업 활성화를 위한 모바일 RFID/NFC 융합형 기술 개발]

참 고 문 헌

- [1] 진동수, "NFC 기반의 전자상거래 비즈니스 모델에 관한 연구", *통상정보연구* 제13권 제4호, 81-100쪽, 2011년 11월.
- [2] 이경진, 홍성우, "NFC 기반 세렌티피티 시스템 설계", *지능정보연구* 제17권 제4호, 293-204쪽, 2011년 12월.
- [3] 최동원, 정순목, 윤승현, 이경민, 조성진, 김길재, 권기호, 전재욱, "Cloud 환경에서 NFC를 활용한 정보 공유 서비스", *HCI 2012 학술대회*, 2012년 1월.
- [4] 박재영, 김용강, 이정현, 최광훈, "NFC 모바일 폰을 활용한 도서관 시스템", *한국정보과학회 2011년 가을 학술발표논문집*, Vol.38, No.2(D), 43-45쪽, 2011년 11월.
- [5] 배경재, 박승진, "시각 장애인의 정보접근성 향상을 위한 유비쿼터스 도서관 구축에 관한 연구", *한국도서관 정보학회지*, 제37권 제2호, 273-290쪽, 2007년 2월.
- [6] A. J. C. Trappey, C. V. Trappey, F. C. Hsu, and D. W. Hsiao, "A Fuzzy Ontological Knowledge Document Clustering Methodology", *The Journal of IEEE Transaction On System, Man and Cybernetics*, vol. 39, no. 3, Jun. pp.806-814, 2009.
- [7] X. Hu, X. Zhang, C. Lu, E. K. Park, X. Zhou, "Exploiting Wikipedia as External Knowledge for Document Clustering", *Proceeding of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 397-406, 2009.
- [8] A. Huang, D. Milne, E. Frank, I. H. Witten, "Clustering Document with Active Learning using Wikipedia", *Proceeding of the 8th IEEE International Conference on Data Mining (ICDM'08)*, pp. 839-844, 2008.
- [9] A. Huang, D. Milne, E. Frank, I. H. Witten, "Clustering Document using a Wikipedia-based Concept Representation", *Proceeding of Advances in Knowledge discovery and data mining, LNCS 5476*, pp.628-636, 2009.
- [10] G. V. R. Kiran, K. Ravi Shankar, V. Pudi,

"Frequent Itemset based Hierarchical Document Clustering using Wikipedia as External Knowledge", Technical Report No: IIT/TR/2010/33, Wales, UK, 2010.

- [11] T. Li, S. Ma, M. Ogiwara, "Document Clustering via Adaptive Subspace Iteration", In proceeding of SIGIR'04, pp. 218-225, 2004.
- [12] F. Wang, C. Zhang, "Regularized Clustering for Documents", In proceeding of ACM SIGIR'07, pp. 95-102, 2007.
- [13] W. Xu, X. Liu, Y. Gon, "Document Clustering Based On Non-negative Matrix Factorization", Proceeding of Special Interest Group on Information Retrieval (SIGIR), pp. 267-274, 2003.
- [14] S. Park, D. U. An, B. R. Char, C. W. Kim, "Document Clustering with Cluster Refinement and Non-negative Matrix Factorization", In proceeding of ICONIP'09, pp. 281-288, 2009.
- [15] 박선, 김철원, "비음수 행렬 분해와 군집의 응집도를 이용한 문서군집", *한국해양정보통신학회 논문지*, 제13권 제12호, pp. 2603-2608, 2009.
- [16] 박선, 김경준, "비음수 행렬 분해와 퍼지 관계를 이용한 문서군집", *한국항행학회 논문지*, 제14권 제2호, pp. 239-246, 2010.
- [17] 박선, 안동언, "구성분 분석과 퍼지 연관성을 이용한 문서군집 방법", *한국정보처리학회 논문지*, 제17-B권, 제2호, pp. 177-182, 2010.
- [18] 박선, 김경준, 이진석, 이성로, "군집 주제의 유의어와 유사도를 이용한 문서군집 향상 방법", *전자공학학회논문지* 제48권 SP편 제5호, pp. 30-38, 2011.
- [19] 박선, 정민아, 이성로, "군집의 중요 용어와 위키 피디아를 이용한 문서군집 향상", *전자공학학회논문지*, 제49권 SP편 제2호, pp. 42-52, 2012.
- [20] D. D. Lee, H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol.401, 788-791, 1999.
- [21] W. B. Frakes, B. Y. Ricardo, "Information Retrieval : Data Structure & Algorithms," Prentice-Hall, 1992.
- [22] 한경남, 남경완, "한국어정보처리 입문 : 컴퓨터가 우리말을 이해하려면?", *커뮤니케이션북스*, 2007.
- [23] W. Xu, X. Liu, Y. Gon, "Document Clustering Based On Non-negative Matrix Factorization", Proceeding of Special Interest Group on Information Retrieval (SIGIR), 267-274, 2003.

저자 소개



박선

1996년 2월 전주대학교 전자계산학과 이학사
2001년 2월 한남대학교 정보통신학과 공학석사
2007년 2월 인하대학교 컴퓨터정보공학과 공학박사

2008~2009년 호남대학교 컴퓨터공학과

전임강사

2010년 전북대학교 인력양성사업단 박사후 과정

2010년 12월~현재 목포대학교 정보산업연구소 연구진
임교수

<주관심분야 : 정보검색, 데이터마이닝, 데이터베이스, 해양IT정보융합>



김형균

1998년: 조선대학교 전자계산전공 공학석사

2004년: 조선대학교 컴퓨터공학과 공학박사

2004년: 동강대학 컴퓨터정보과 초빙교수

2007년: 호남대학교 공학혁신센터 선임연구원

현재: 조선대학교 컴퓨터공학과 겸임교수

<주관심분야: 영상처리, 스마트폰 어플리케이션, 클라우드 컴퓨팅, 빅데이터, NFC>



심수정

1996.2월 호남대학교 컴퓨터공학과 학사 (졸)

1999.2월 전남대학교 전산학과 석사 (소프트웨어공학)

2003.8월 전남대학교 전산학과 박사과정 수료(자연어처리)

2012.3 호남대학교 초빙교수

2013.6 (주)동하테크 책임연구원

<주관심분야: 자연어 처리, 문서분류, 정보검색 등>