문장군집의 응집도와 의미특징을 이용한 포괄적 문서요약

박선* · 이연우** · 심천식*** · 이성로****

Generic Document Summarization using Coherence of Sentence Cluster and Semantic Feature

Sun Park* · Yeonwoo Lee** · Chun Sik Shim*** · Seong Ro Lee****

이 논문은 2012년도 정부(교육과학기술부)의 재원으로 한국연구재단의 대학중점연구소 지원사업으로 수행된 연구임(2009-0093828), 본 연구는 지식경제부 및 정보통신산업진흥원의 대학 IT연구센터 지원사업의 연구결과로 수행되었음(NIPA-2012-H0301-12-2005), 이 논문은 2012년도 목포대학교 중형조선산업 지역혁신센터(RIC)에 의하여 지원되었음.

요 약

지식 기반의 포괄적 문서요약은 문장집합의 구성이 요약 결과에 영향을 받는다. 이러한 문제를 해결하기 위해서 본 논문은 의미특징에 의한 군집과 문장군집의 응집도를 이용하여 포괄적 문서요약을 하는 새로운 방법을 제안한다. 제안 방법은 비음수행렬분해에서 유도되는 의미특징을 이용하여 문장을 군집하고, 문서의 내부구조를 잘 표현하는 문장군집들로 문서의 주제 그룹을 분류할 수 있다. 또한 문장군집의 응집도와 재군집에 의한 군집의 정재를 이용하여 중요한 문장을 추출함으로써 요약의 질을 향상시킬 수 있다. 실험결과 제안방법은 다른 포괄적 문서요약 방법에 비하여 좋은 성능을 보인다.

ABSTRACT

The results of inherent knowledge based generic summarization are influenced by the composition of sentence in document set. In order to resolve the problem, this papser propses a new generic document summarization which uses clustering of semantic feature of document and coherence of document cluster. The proposed method clusters sentences using semantic feature deriving from NMF(non-negative matrix factorization), which it can classify document topic group because inherent structure of document are well represented by the sentence cluster. In addition, the method can improve the quality of summarization because the importance sentences are extracted by using coherence of sentence cluster and the cluster refinement by re-cluster. The experimental results demonstrate appling the proposed method to generic summarization achieves better performance than generic document summarization methods.

키워드

포괄적 문서요약, 의미특징, 문장군집의 응집도, 비음수 행렬분해, kmeans 군집

Key word

generic summarization, semantic feature, coherence of sentence cluster, NMF(non-negative matrix factorization), kmeans clustering

* 정회원 : 목포대학교 정보산업연구소 연구전임교수
(교신저자, sunpark@mokpo.ac.kr)
** 정회원 : 목포대학교 정보통신학과 부교수
*** 정회원 : 목포대학교 조선공학과 조교수
**** 정회원 : 목포대학교 정보전자공학과 교수

접수일자 : 2012. 07. 04
심사완료일자 : 2012. 08. 08

Open Access http://dx.doi.org/10.6109/jkiice.2012.16.12.2607

Ⅰ. 서 론

클라우드 컴퓨팅과 정보량의 증가에 의한 빅데이터의 출현으로 대량의 정보에 대한 분석기술의 필요성이 점차 증가 하는 추세에 있다. 빅데이터는 다양한 비정형 데이터에 의해 구성되고 있으나 아직까지도 문자에 의한 문서들이 가장 중요한 정보의 축적수단으로 사용되고 있다. 특히 빅데이터 내의 정보를 효율적으로 분석 및 관리하기 위해서는 효율적인 정보요약 기술이 필요하며, 이 때문에 문서요약기술의 필요성이 다시 증가하는 추세에 있다.

문서 요약은 문서의 기본적인 내용을 유지하면서 문서의 량을 줄이는 작업이다. 문서의 요약은 포괄적 문서요약(generic document summarization)과 질의 기반의 문서요약(query-based document summarization)으로 구분된다. 포괄적 문서요약은 독자가 문서를 다 읽지 않고도 문서의 내용을 파악할 수 있도록 문서를 대표할 수 있는 중요한 정보 및 주제 별로 요약하는 방법이다. 질의 기반의 문서요약은 문서에 포함된 여러 가지 주제로부터 사용자의 질의와 관련된 내용만을 요약하는 방법이다. 대상 문서를 기준으로 하나의 문서로 부터 요약하는 경우는 단일문서요약, 하나의 주제가 여러 개의 문서인 신문기사와 같은 다중문서로 부터 요약하는 경우를 다중문서요약 이라고 한다. 이외에도 웹상의 사용자 로그나 사용자의 흥미와 관련된 특별한 정보를 기준으로 문서를 요약하는 개인화 문서요약이 있다[1].

포괄적 문서요약에 대한 최근 연구들은 대표적으로 자연어처리 기반방법, 외부지식 기반방법, 내부지식 기반방법으로 구분할 수 있다[2, 3, 4, 5, 6, 7]. 자연어처리 기반의 요약방법은 어휘 분석 및 문법 분석을 이용하여 포괄적 문서요약하는 방법이다. 이 방법들은 높은 요약 성능을 보이나 자연어 처리를 위하여 복잡한 단계와 처리 비용이 많이 든다[2].

외부지식 기반의 문서요약 방법은 외부지식인 워드넷을 이용하여 요약하는 방법으로 외부지식을 이용하여 관련 용어를 확장하여 동음이의어나 유의어를 포함한 문장을 요약하여 요약 주제의 범위를 확장할 수 있다. 그러나 요약의 범위를 확장하기 위해서는 외부지식을 전처리해야하는 비용문제를 가지고 있다[3].

내부지식 기반방법은 문서의 내부 특성을 나타내는 의미특징을 이용한 방법으로서 쉽게 포괄적 문서요약의 특성을 나타내는 주제들을 추출하여 좋은 요약 결과를 얻을 수 있다. 그러나 문장집합의 구성 문장들이 유사한 특성을 보이거나, 극단적으로 다른 특성을 갖고 있으면 추출된 의미특징들의 문장집합의 내부 구조를 충분하게 반영할 수 없으므로 좋은 요약 결과를 얻기 힘들다[4, 5, 6, 7].

본 논문은 내부지식 기반의 포괄적 문서요약 방법의 제한 사항을 해결하기 위해서 의미특징을 이용한 문장군집과 군집의 응집도를 이용하여 포괄적 문서요약 방법을 제안한다. 제안방법은 문서의 주제를 나타내는 의미특징을 이용함으로써 문서에 포함된 각기 다른 주제를 잘 표현할 수 있다. 또한 문장 군집의 응집도를 이용함으로써 내부지식 기반의 포괄적 문서요약 방법이 원본문서의 내부구성에 제한받는 문제를 해결할 수 있다.

본 논문의 구성은 다음과 같다. 제2장은 관련연구로 기존 포괄적 문서요약방법에 대해서, 제3장은 제안한 포괄적 문서요약 방법을, 제4장에서는 실험 및 평가에 대해 기술한다. 마지막으로 제5장에서는 결론을 맺는다.

Ⅱ. 관련연구

다음은 현재 많이 연구되고 있는 포괄적 문서요약방법에 대한 관련연구이다. Moens외 저자들은 단일문서와 다중문서를 위해서 자연어처리방법을 이용한 포괄적 기술에 대하여 제안하였다. 이들은 단일 및 다중 문서의 요약 기술을 위해서 복합적인 문서를 사용하였으며, 전처리로 어휘 분석, 가설검정, 영문법분석을 이용하여 문장을 추출하였다. 추출된 문장으로부터 중요한 문장을 선택하기 위해서 문장의 주제들과 문서 내의 용어 분포를 계산하여 계층적 주제 트리를 구축하였다. 이후 자연어 처리의 3단계 문장 압축 방법을 이용하여 중복되는 내용을 탐색하여 문서를 요약한다[2].

이들의 방법은 문서요약 평가대회인 DUC(document understanding conference)에서 중상위의 성능을 보였으나 여러 단계의 자연어처리 방법과 통계학적 방법을 이용하기 때문에 처리비용이 많이 든다.

Bellare외 저자들은 외부지식인 워드넷을 사용한 포괄적 문서요약 방법을 제안하였다. 이들의 방법은 워

드넷을 이용하여 문서와 관련된 용어를 이용하여 그래프로 만들고, 관련용어들을 순위화한다. 이후 문서와 관련된 의미 있는 문장을 추출하여 주성분분석을 이용하여 문서를 요약한다[3]. 이들의 방법 또한 외부 사전인 워드넷을 전처리해야 하기 때문에 처리 비용이 많이 든다.

Gong과 Liu는 관련척도와 잠재의미분석을 이용한 두 가지 포괄적 문서요약 방법을 제안하였다. 관련척도는 전체문장집합과 문장 간의 유사도를 이용하여 유사도가 높은 문장을 중요문장으로 추출하여 요약하는 방법을 제안하였으며, 잠재의미분석을 이용한 방법은 문장집합을 특징값분해하여 문서의 내부구조를 나태나는 의미 특징값을 이용하여 문서를 요약하는 방법을 제안하였다[4]. 잠재의미분석을 이용한 문서요약방법은 의미특징인 고유값이 양수와 음수 값을 갖기 때문에 의미특징을 직관적으로 이해할 수 없으며, 의미가 적은 문장을 추출할 수 있다[5]. Zha는 문장군집과 상호강화 원리를 이용하여 주제어 추출과 문서요약을 동시에 진행하는 방법을 제안하였다. 이들의 방법은 문서를 가중 무방향 그래프와 가중 양방향 그래프로 모델링한 후에 스펙트럼 그래프 군집방법을 이용하여 주제 그룹으로 군집한다. 이후에 문장과 용어(terms)로부터 상호강화 원리를 이용하여 특징점수를 계산하고, 이를 이용하여 문장들을 주제그룹들로부터 문서를 요약하였다[5]. 이들의 방법 역시 문서 내부의 특징을 이용하기 때문에 문서의 구조에 요약결과가 많은 영향을 받는다. 본 논문의 저자들은 이전에 비음수행렬분해를 이용하여 문서의 의미특징인 비음수 의미특징행렬과 의미가변 행렬을 이용한 포괄적 문서요약을 제안하였다. 이 방법은 의미특징의 값은 높으나 전체 문서들에서 별로 중요하게 나타나지 않는 문장들이 선택될 수 있는 문제를 가지고 있다[6, 7].

Ⅲ. 포괄적 문서요약을 위한 제안 방법

본 장에서는 비음수행렬분해의 의미특징을 이용한 문장군집과 군집의 응집도를 이용하여 포괄적 문서요약의 성능을 향상시키는 방법을 제안한다.

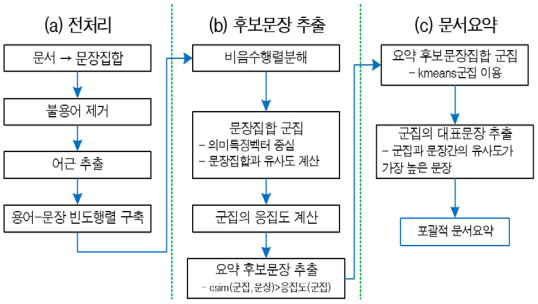

그림 1. 제안 포괄적 문서요약 방법
Fig. 1 Proposed generic document summarization method

제안 방법은 전처리 단계, 후보문장 추출단계, 문서요약 단계로 구성된다. 그림1은 본 논문에서 제안한 포괄적 문서요약의 개요이다.

3.1. 전처리 단계

그림1(a)의 전처리 단계는 주어진 문서를 문장으로 분해하여서 용어-문장 빈도행렬을 구축한다. 용어-문장 행렬을 구축하기 위해서는 문장으로 부터 용어들을 추출한 다음에 문장에서 불필요한 용어를 제거하는 불용어(stop-word) 작업과 용어들의 의미를 일치시키는 어근(stemming)추출 작업은 진행한다. 본 논문에서는 문서처리에 국제적으로 많이 사용되는 평가 자료를 이용하기 때문에 영어문서의 전처리를 기준으로 설명한다. 불용어 제거는 Rijsbergen의 불용어 목록[8]을 이용하여 목록에서 정의하고 있는 무의미한 용어들을 제거한다. 어근추출은 Porter의 어근추출 알고리즘[8]을 이용하여 영어의 파생어들을 가장 중심이 되는 용어인 어근으로 변환한다. 마지막으로 추출된 용어가 문장에서 출현하는 빈도를 이용하여 용어-문장 빈도행렬을 구축한다.

본 논문에서 행렬 X의 j번째 열벡터는 X_{*j}로, i번째 행벡터는 X_{i*}로, i번째 행과 j번째 열의 원소는 X_{ij}로 표시한다. 본 논문에서 사용되는 용어-문장 빈도행렬은 다음과 같이 표기된다. 용어-문장 빈도 행렬 A는, j번째 문장 A_{*j}는 용어빈도 벡터 $A_{*j} = [A_{1j}, A_{2j}, \cdots, A_{nj}]^T$로 표현되고, 행렬 A의 요소인 A_{ij}는 j번째 문서에서 i번째 용어를 나타낸다.

3.2. 후보문장 추출단계

그림 1(b)의 후보문장 추출단계는 비음수행렬분해, 문장집합 군집, 군집의 응집도 계산, 요약 후보문장 추출로 구성된다. 세부방법은 다음과 같다. 첫째, 용어-문장 빈도행렬을 식(2)와 식(3)을 이용하여 비음수행렬분해 하여서 문장집합의 내부특징을 표현하는 의미특징 벡터를 추출한다. 둘째, 추출된 의미특징벡터를 군집의 중심으로 설정하고, 문장집합에 식(4)의 코사인유사도를 이용하여 문장집합을 군집한다. 셋째, 각각의 문장군집에 식(5)를 이용하여 군집의 응집도를 계산한다. 마지막으로 군집과 군집에 속하는 문장의 유사도를 계산하고, 계산된 유사도가 군집의 응집도 보다 크면 요약후보문장 집합에 포함시킨다.

비음수행렬분해는 주어진 0과 양의 행렬로부터 비음수 인수를 찾아내는 행렬분해 알고리즘이다[9]. 비음수행렬분해 알고리즘은 다음과 같이 계산된다. 문장집합이 k개의 군집으로 구성된다고 가정할 때, 행렬 X를 식(2)의 목적 함수가 최소 값을 갖도록 식(3)을 반복하여서 식(1)과 같이 비음수의미특징행렬(NSFM, non-negative semantic feature matrix) W와 비음수의미변수행렬(NSFM, non-negative variable matrix) H로 분해한다.

$$X \approx WH \tag{1}$$

$$J = \frac{1}{2} \| X - WH \| \tag{2}$$

여기서 행렬 X는 $m \times n$개의 원소로 구성되는 행렬이며, 행렬 W는 $m \times k$개의 원소로 구성되고, 행렬 H는 $k \times n$개의 원소로 구성된다.

$$w_{ij} \leftarrow w_{ij} \frac{(XH^T)_{ij}}{(WHH^T)_{ij}}, \; h_{ij} \leftarrow h_{ij} \frac{(W^TX)_{ij}}{(W^TWH)_{ij}} \tag{3}$$

여기서 H^T는 H의 전치행렬이고, W^T는 W의 전치행렬이다.

코사인 유사도를 나타내는 식은 다음과 같다[8].

$$csim(A_{*a}, A_{*b}) = \frac{\sum_{j=1}^{m} A_{ja} \times A_{jb}}{\sqrt{\sum_{j=1}^{m} A_{ja}^2} \times \sqrt{\sum_{j=1}^{m} A_{jb}^2}} \tag{4}$$

여기서, A_{*a}와 A_{*b}는 행렬 A의 a번째와 b번째 열벡터로 문서에 포함된 각각의 문장을 나타낸다.

문장군집내의 군집의 응집도를 계산하기 위하여서 식(5)를 이용한다. 군집의 응집도는 $c(C^n)$는 n번째 문장군집 C^n에 포함된 문장간의 유사도를 기반으로 다음과 같이 계산한다[10].

$$c(C^n) = \frac{\sum_{A_{*i} \in C^n} \left(\sum_{A_{*j} \in C^n - \{A_{*i}\}} csin(A_{*i}, A_{*j}) \right)}{|C^n|} \tag{5}$$

여기서 $|C^n|$는 n번재 문장군집 C^n에 포함되어 있는 전체 문장의 개수이며, 군집의 응집도는 식(4)을 이용하여서 임의의 서로 다른 문장 A_{*i}, A_{*j}의 유사도 $csin(A_{*i}, A_{*j})$의 합을 군집에 포함된 문장의 전체 개수로 나눈 값이다.

3.3. 문서요약 단계

내부지식 기반의 문서요약 방법은 문서에 포함되는 문장의 구성에 따라서 추출되는 의미특징들이 문서의 내부구조를 충분히 반영할 수 없는 문제를 가지고 있다. 이러한 문제를 해결하기 위해서 본 논문에서는 의미특징을 반영하여 문장집합을 군집하고, 군집의 응집도를 이용하여 요약 후보문장집합을 구성한 다음 충분히 문서의 주제를 반영할 수 있도록 요약후보문장집합을 재군집하여서 중요한 문장을 추출한다. 즉, 문장군집을 재군집하여 문서의 내부구조를 정재 함으로써 문서의 내부구조가 의미특징에 편향되게 반영되는 것을 최소화 시킬 수 있다.

그림 1(c)의 문서요약 단계는 요약 후보문장집합 군집, 군집의 대표문장 추출 단계로 구성되며 세부 내용은 다음과 같다. 첫째, 요약 후보문장집하에 $k$means 군집방법[11]을 이용하여 요약문장의 개수만큼 군집한다. 마지막으로, 각각의 군집과 군집에 포함된 문장 간에 식(4)의 코사인유사도를 계산하여 유사도가 가장 높은 문장을 추출하여 문서를 요약한다.

Ⅳ. 실험 및 평가

본 논문에서 사용된 평가 자료는 Reuters-21578 문서집합[12]과 야후코리아 뉴스[13]의 기사를 무작위로 선

택하여 사용하였다. Reuters-21578 문서집합은 본 논문에서 설명된 전처리방법을 사용하였으며, 야후코리아 뉴스는 한글형태소 분석 도구[14]를 사용하여 용어만 추출하여 용어-문장 빈도행렬을 구성하였다. 제안방법을 비교하기 위하여서 세 명의 평가자가 수동으로 요약한 요약문과 포괄적 문서요약 방법들의 요약문에 대한 평가척도를 계산하였다. 다음 표1은 평가 자료에 대한 특성을 나타낸다.

표 1. 평가자료의 속성
Table. 1 Property of the test data set

문서의 속성	Reuters	야후 코리아
문서 수	500	500
10문장 이상인 문서 수	321	401
문서당 평균 문장 수	12.5	15.4
최소 문장 수	3	5
최대 문장 수	50	134

성능 평가척도는 정보검색에서 주로 사용되는 정확률(P, precision), 재현율(R, recall), F-measure(F)를 이용하였다[8, 14].

$$(R) = \frac{|S_{man} \cap S_{sum}|}{|S_{man}|} \tag{6}$$

$$(P) = \frac{|S_{man} \cap S_{sum}|}{|S_{sum}|} \tag{7}$$

$$(F) = \frac{2RP}{R+P} \tag{8}$$

여기서 S_{man}, S_{sum}은 각각 평가자와 문서요약 방법에 의해서 요약된 문장이다.

실험은 그림2와 그림3과 같이 내부지식 기반의 요약방법들인 KM, RM, LSA, MRP, NMF, SC을 구현하여 요약결과에 대한 성능을 비교 평가하였다. SC는 본 논문에서 제안한 방법으로 의미특징(semantic feature)과 군집의 응집도(coherence)를 이용하여 문서를 요약하는 방법이다. KM은 *k*means 군집방법을 이용하여 문서를 요약하는 방법이며[11], RM과 LSA은 Gong과 Liu가 제안한 방법으로 각각 관련척도(relevance measure)와 잠재의미분석(latent semantc analysis)을 이용하여 문서를 요약하는 방법이다[4]. MRP는 Zha가 제안한 방법으로 문장군집과 상호강화원리(mutual reinforcement principle)를 이용하여 문서를 요약하는 방법이다[5]. NMF는 저자들의 이전 제안방법으로 비음수행렬분해를 이용하여 문서를 요약하는 방법이다[6, 7].

그림 2에서 보는 것과 같이 Reuters-21578를 이용한 평가 결과에서는 제안 방법(SC)의 평균 재현율(R), 정확율(P), F-measure(F)가 KM에 비하여 19%, 21%, 15%가 높으며, RM에 비해서는 21%, 16%, 19%가 높고, LSA에 비해서는 11%, 11%, 12%가 높고, MRP에 비해서는 9%, 5%, 6%가 높고, NMF에 비해서는 6%, 2%, 3%가 높다.

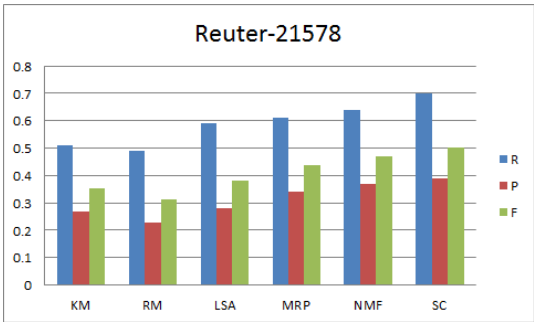

그림 2. Reuters-21578을 이용한 요약결과 비교
Fig. 2 Comparison summarization results using Reuters-21578

그림 3에서 보는 것과 같이 야후 코리아를 이용한 평가 결과에서는 제안 방법(SC)의 평균 재현율(R), 정확율(P), F-measure(F)가 KM에 비하여 19%, 1%, 13%가 높으며, RM에 비해서는 27%, 13%, 18%가 높고, LSA에 비해서는 2%, 11%, 14%가 높고, MRP에 비해서는 14%, 7%, 9%가 높고, NMF에 비해서는 3%, 6%, 7%가 높다.

Reuters와 야후 코리아의 평균성능평가결과 제안방법인 SC이 가장 좋은 결과를 나타내며, 그 다음으로 NMF, MRP, LSA, KM, RM 순으로 평가 결과를 보이고 있다. 이는 단순히 문장의 유사도에 의한 KM 방법보다 군집에 의해 문서의 주제를 추출하는 KM 방법이 더 좋은 성능을 보이고 있다. 또한 KM보다 문서의 내부 의미특징을 나타내는 LSA, NRP, NMF 방법이 더 좋은 성능

을 보이고 있으며, 특히 문서의 내부구조와 내부 의미특징을 재군집하여서 정재하는 본 논문의 제안방법인 SC가 가장 좋은 결과를 보이고 있다. 이것으로 문서의 속성에 많은 영향을 받는 내부지식을 이용한 요약방법은 문서의 내부 특성을 정재함으로써 요약을 효율을 향상시킬 수 있음을 알 수 있다.

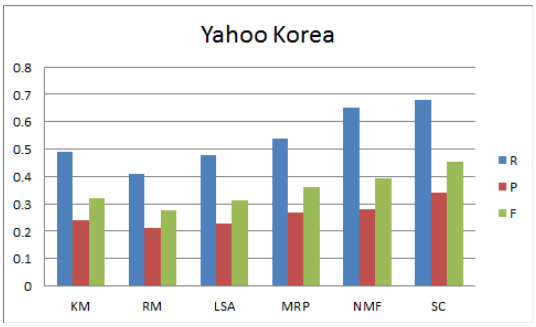

그림 3. 야후코리아 기사를 이용한 요약결과 비교
Fig. 3 Comparison summarization results using Yahoo Korea News

V. 결 론

본 논문은 내부지식 기반의 포괄적 문서요약 방법의 성능을 향상하기 위해서 의미특징을 이용한 문장군집과 군집의 응집도를 이용하여 포괄적 문서요약방법을 제안하였다. 제안방법은 의미특징을 이용하여 문장을 군집함으로써 문서의 주제를 잘 나타내는 의미 있는 문장을 추출할 수 있다. 또한 문장군집의 응집도를 계산하여 요약후보문장을 추출하고 재군집합으로서 의미특징이 원본문서의 내부구성에 제한받는 문제를 완화하여 문서요약의 성능을 향상 하였다. 실험결과 기존 내부지식을 이용한 포괄적 문서요약 방법에 비하여 더 좋은 평가 결과를 보였다.

참고문헌

[1] I. Mani, M. T. Maybury, "dvances in Automatic Text," The MIT Press, 1999.

[2] M. F. Moens, R. Angheluta, J. Dumortier, "Generic technologies for single-and multi-document summarization," Information Processing and Management 41, pp.569-586, 2005.

[3] K. Bellare, A. D. Sarma, A. D. Sarma, N. Loiwal, V. Mehta, G. Ramakrishnan, P. Bhattacharyya, "Generic Text Summarization using WordNet," In proceeding of LREC 2004, 2004.

[4] Y. Gong, X. Liu, "Generic Text Summarization Using Relevance Measure and Latent Semantic Analysis," In proceeding of ACM SIGIR'01, pp.19-25, 2001.

[5] H. Zha, "Generic Summarization and Keyphrase Extraction Using Mutual Reinforcement Principle and Sentence Clustering," In proceeding of ACM SIGIR'02, pp.113-120, 2002.

[6] S. Park, "Generic Summarization Using Non-negative Semantic Variable," Lecture Notes in Comupter Science 5226, Springer, pp.1052-1058, 2008.

[7] 박선, 이종훈, "의미특징의 포괄적 중요도를 이용한 포괄적 문서요약", 한국항행학회논문지, 제12권 제5호, pp.41-47, 2008.

[8] W. B. Frankes, B. Y. Ricardo, "Information Retrieval : Data Structure & Algorithms," Prentice-Hall, 1992.

[9] D. D. Lee, H. S. Seung, "Algorithms for non-negative matrix factorization," In Advances in Neural Information Processing Systems, vol. 13, pp.556-562, 2001.

[10] 주길홍, 이원석,"효율적인 문서검색을 위한 레벨별 불용어 제거에 기반한 문서 클러스터링", 컴퓨터교육학회 논문지 11권3호, 2008. 05

[11] J. Han, M. Kamber, "Second Edition Data Mining Concepts and Techniques", Morgan Kaufman, 2006.

[12] http://kdd.ics.uci.edu/databases/reuters21578/reuters21578.html, 2012.

[13] http://kr.news.yahoo.com/, 2012.

[14] S. S. Kang, "Information Retrieval and Morpheme Analysis," HongReung Science Publishing Co., 2002.

저자소개

박선(Park Sun)

1996년 전주대학교(학사)
2001년 한남대학교(석사)
2007년 인하대학교(박사)
2008년~2009년 호남대학교 컴퓨터공학과 전임강사
2010년 전북대학교 전기전자정보 인력양성사업단 박사후과정
2010년 12월~현재 목포대학교 정보산업연구소 전임연구교수
※관심분야 : 정보검색, 데이터마이닝, 데이터베이스, 해양IT정보융합

이연우(Yeonwoo Lee)

1994년 고려대학교(석사)
2000년 고려대학교(박사)
2000년~2003년 영국 Edinburgh 대학교 Research Fellow
2004년~2005년 삼성종합기술원
2005년~현재 국립목포대학교 정보공학부, 부교수
※관심분야 : 해상무선통신, e-Navigation, Cognitive Radio, 4G 이동통신

심천식(Shim Chun Sik)

1995년 인하대학교(학사)
1997년 인하대학교(석사)
2003년 인하대학교(박사)
2008년 9월~현재 목포대학교 조선공학과 교수
※관심분야 : 조선IT정보융복합

이성로(Lee Seong Ro)

1987년 고려대학교(학사)
1990년 한국과학기술원(석사)
1996년 한국과학기술원(박사)
1997년 9월~현재 목포대학교 정보전자공학과 교수
※관심분야 : 디지털통신시스템, 이동 및 위성통신 시스템, USN/텔레미틱스응용분야, 임베디드시스템