콘도르 정보 검색 시스템

(Information Retrieval System : Condor)

박순철*, 안동언*
(Soon-Cheol Park, Dong-Un An)

요 약 본 연구는 다중어 질의어를 제공하는 대용량 정보검색 시스템, 콘도르에 대한 고찰이다. 이 시스템은 전북대학교, ㈜서치라인, 그리고 카네기멜론 대학교가 컨소시엄 형태로 개발하였다. 이 시스템의 질의처리는 확률 모델을 기반하고 있으며 최근 정보검색 시스템에서 제공하는 문서 클러스터링 기능을 제공하고 있다. 특히 시스템의 특징은 다중어 질의어를 처리하고 질의를 중심으로 온라인으로 문서를 클러스터링하고 요약하는 것이다. 본 시스템은 이미 국내의 3,000만개 웹페이지에 대한 테스트를 마쳤으며 그 안정성을 확보하고 있다.

핵심주제어 : 정보검색, 다중어 질의어 처리, 온라인 문서 요약, 계층적 문서 클러스터링

Abstract This paper is a review of the large-scale information retrieval system, CONDOR. This system was developed by the consortium that consists of Chonbuk National University, Searchline Co. and Carnegie Mellon University. This system is based on the probabilistic model of information retrieval systems. The multi-language query processing, online document summarization based on query and dynamic hierarchy clustering of this system make difference of other systems. We test this system with 30 million web documents successfully.

Key Words : Information Retrieval, multi-language query processing, online document summarization, hierarchy clustering

1. 서 론

정보사회의 발달이 가져오는 변화 가운데 하나는 정보의 급격한 증대이다. 최근에는 책, 신문과 같은 인쇄매체 만이 아니라 라디오, 텔레비전과 같은 음성 및 영상매체와 인터넷에서 제공되는 정보의 양은 과거와는 비교할 수 없이 급격히 늘어나고 있다. 한 연구 결과에 의하면 세계적으로 30만년동안 12엑사바이트(Exa Byte) 정보가 만들어 졌으며 향후 3년 내에 45엑사바이트 이상의 새로운 정보가 생성될 것이라고 한다. 또한 매년 2배 이상의 신규 정보량이 증가하고 있으며 대부분의 정보는 디지털화되고 있다[1]. 이러한 정보량의 증가와 디지털화는 더욱더 정보검색의 필요성을 강조하고 있다.

정보의 양이 급증한다는 것은 사용자들에게 필요한 정보가 다양하고 풍부해진다는 것을 뜻한다. 동시에 너무나 풍부해진 정보의 바다 속에서 사용자가 원하는 정보를 찾는 것이 더욱 복잡해지고 있다. 따라서 정보화시대에 가장 필요한 기술은 정보검색 기술이라고 할 수 있다. 정보검색 시스템의 목적은 대용량의 정보에서 사용자가 원하는 정보를 신속하고 정확하게 찾아주며 또한 사용자가 사용하기 편리하게 처리하는 기술이 요구된다.

지금까지 웹정보검색 시스템 중 세계 최고로 알려진 Google (http://www.google.com/)은 약 30억 개 이상의 웹사이트 정보를 가지고 있다. 그러나 검색기

* 전북대학교 전자정보공학부 부교수

능이 단순하여 사용자들이 검색 결과에서 원하는 정보를 찾기가 불편하다. 이러한 점을 보완하여 개발된 WiseNut (http://www.wisenut.com)와 Vivisimo (http://vivisimo.com/)는 문서 클러스터링 기능을 갖추고 있다. 그러나 Google이 갖는 대용량 검색기술을 따라가기는 힘들다. 외국에서 개발된 기존의 정보검색 시스템의 장단점을 반영하여 본 연구팀에서는 다양한 정보검색기능과 대용량 데이터를 동시에 처리할 수 있는 세계 최고의 검색 시스템인 콘도르(Condor)를 개발하였다.

콘도르는 동양권 언어를 이해하고 있는 국내 기술진들에 의해서 한국어, 중국어, 일본어 문서를 처리할 수 있는 색인기를 개발하여, 영어를 포함한 다중어 질의어 처리가 가능하게 하였다. 또한 주어진 질의를 중심으로 온라인으로 요약이 가능하며 다이나믹한 계층적 클러스터링 기능을 포함하고 있다. 이러한 콘도르의 기능은 기존의 시스템과 분명한 차이가 있으며 계속적으로 차세대 검색 기능을 개발하고 채용해 나갈 것이다.

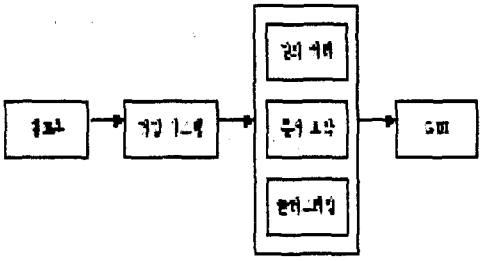
본 논문의 구성은 다음과 같다. 다음 2장은 콘도르의 전체 구조와 각 모듈에 대한 소개와 기술 내용이다. 3장은 구현 결과를 구현 화면 중심으로 한 설명이다. 마지막으로 4장에서는 본 연구에 대한 결론을 언급하겠다.

2. 콘도르의 구조

콘도르(Condor)의 원 뜻은 하늘을 높이 나는 캘리포니아의 독수리를 의미한다. 즉 높은 고공에서 정확히 먹이를 찾는 콘도르처럼 수많은 웹사이트에서 사용자의 질의에 따라 신속하고 정확하게 데이터를 검색해주는 시스템을 콘도르라고 이름하였다. 이 콘도르 정보검색 시스템은 전북대학교의 정보검색연구센터, ㈜서치라인, 카네기멜론 대학교의 언어기술연구소의 공동 프로젝트로 2001년부터 시작하여 약 1년 여 만에 구현된 시스템이다.

콘도르의 전체 구조는 크게 웹로봇, 저장 시스템, 질의 처리부, 문서 요약, 클러스터링 그리고 GUI (Graphic User Interface) 부분으로 구성된다. [그림 1]은 콘도르의 구조를 나타낸다.

각 모듈 별로 간단히 소개하면, 로봇과 저장 시스템은 일반 정보검색 시스템과 유사하다. 질의어 처리를



[그림 1] 콘도르의 구조

위한 검색 모델은 확률 모델을 따랐다. 또한 다중어 (한국어, 중국어, 일본어, 영어) 질의어 처리가 가능한 것도 콘도르의 특징이다.

문서 요약은 문서를 단순히 요약하는 오프라인 요약과 질의어에 포함된 용어를 중심으로 요약하는 온라인 요약을 포함한다. 특히 온라인 문서 요약은 현재까지 상용 정보검색 시스템에서는 구현된 적이 없다.

문서 클러스터링의 구조는 계층 구조를 갖는다. 이러한 구조는 사용자의 판단에 따라 정확한 문서 집단을 택할 수 있도록 유도한다. 또한 클러스터링의 수는 임으로 정한 고정적인 것을 기본으로 하나 경우에 따라서는 그 수가 동적으로 변하는 알고리즘을 택하여 클러스터링의 결과를 좀더 정확하게 유지하도록 했다.

다른 시스템들과 마찬가지로 콘도르의 GUI 환경은 검색 시스템에 필요한 기능들을 사용자가 손쉽게 이용할 수 있도록 설계했다. 특히 클러스터링의 계층 구조를 가시화 했고 아울러 요약에는 온라인(하늘색)과 오프라인 요약(검은색)을 구분하여 동시에 나타나도록 했다.

2.1 색인어의 가중치

색인어의 가중치는 정보검색 시스템에서 기본적으로 요구되는 것이다. 이러한 색인어 가중치는 검색 결과에 문서 순위를 정하는 것, 문서 요약, 문서 클러스터링 등에 사용이 된다. 가중치를 계산하는 방법은 일반적으로 Tf(term frequency)와 idf(inverse document frequency)를 이용한다[2]. 그러나 본 연구에서는 가중치의 계산에 용어의 특성을 포함시켜 각 용어가 갖는 가중치의 실제적인 정확도를 높였다. 식 (1)은 콘도르에서 사용한 용어의 가중치 공식이다. Tf와 idf 값은 Okapi[2]의 계산법을 따랐다. 용어의 특성을 나타내는 P값은 문자 폰트, 문자의 크기, 문자의 타입에 따라 임의로 정하여졌다. 식 (1)에서 가중치 Wij는 j문서에

나타나는 i번째 용어의 가중치이다.

$$W_{ij} = Tf_{ij} \cdot idf(w_{ij}) \cdot P(w_{ij}) \quad (1)$$

$$\begin{cases} Tf_{ij} = \dfrac{tf_{ij}}{tf+ij+0.5+1.5\times \dfrac{doclen_j}{avgdoclen}} \\ idf(w_{ij}) = \log\left(\dfrac{N-df_{ij}+0.5}{df_{ij}+0.5}\right) \\ P(w_{ij}) = \begin{cases} 2.0 : high \\ 1.5 : important \\ 1 : others \end{cases} \end{cases}$$

2.2 다중어 질의어 처리

콘도르는 다중어 질의어 처리가 가능하다. 특히 한국어, 중국어, 일본어는 그 언어를 이해하는 동양권 연구진이 개발한 색인기를 통하여 색인어를 추출하고 질의어 처리 과정을 거친다.

(1) 색인어 추출

색인어 추출[3]은 각 언어 고유의 언어 구조와 특성을 가지고 있기 때문에 한국어, 중국어, 일본어, 영어 각각에 대해서 서로 다른 방법을 적용하였다.

한국어는 형태소 접속정보를 이용한 형태소분석기를 사용하여 색인어를 추출하였다. 형태소 결합 정보를 가지고 있는 접속정보를 형태소 분석 엔진과 독립적인 지식베이스로 구축하여 유지보수가 용이하도록 하였다[4].

중국어는 사전과 코퍼스를 이용하여 글자들 사이의 상호정보를 이용하여 색인어를 추출하였다. 사전에 있는 단어를 이루는 글자들 사이의 결합도를 코퍼스의 글자들 사이의 결합도보다 높게 부여하였다. 코퍼스는 신화사 신문기사이며 사전은 13만 7천여 개의 단어로 이루어져 있다[5].

일본어는 사전을 사용하여 최장일치에 의해서 색인어를 추출하며 EDR의 형태소 접속정보를 이용한다[6]. 단어를 분리하기 위한 범위를 결정하는데 한자와 가다카나 문자를 사용하여 색인어 추출이 용이하도록 하였다.

영어는 Porter의 알고리즘을 구현한 스테머를 사용하여 색인어를 추출하였다[7].

(2) 다중어 질의어 처리

다중어 질의어를 처리하기 위해서는 우선 코드 문제를 해결하여야 한다. 모든 문서들이 유니코드로 작성되어 있다면 색인과 다중어 질의어 처리가 간단하겠지만 아직 유니코드로 작성된 문서가 많지 않다. 따라서 다중어 처리에 있어서 코드 선택이 우선 되어야 한다.

각 언어마다 여러 가지 코드들이 사용되기 때문에 기준 코드를 정하고 다른 코드들은 변환 모듈을 통하여 기준이 되는 코드로 변환하였다. 한국어에서는 확장형 완성형 코드를 사용하였다. 중국어에서는 간자체 코드(GB)를 기본으로 하고 번자체 코드(Big5)는 변환하였다. 일본어에서는 shift-JIS-code를 기본으로 하였다.

2.3 문서 요약

문서 요약[8, 9, 10, 11, 12]은 문서를 빠른 시간에 쉽게 이해할 수 있도록 도와준다. 이 콘도르에서 사용된 요약은 오프라인 방식으로 입력된 문서에 대하여 미리 요약을 하는 것과 사용자의 질의를 중심으로 한 온라인 문서 요약이 있다.

(1) 오프라인 문서 요약

콘도르의 오프라인 문서 요약은 색인어의 가중치를 중심으로 각 문장의 중요도를 결정한다. 즉 각 문장에 포함되어 있는 색인어들의 가중치를 합한 값에 의해서 문장의 중요도를 결정한다. 문서 요약 알고리즘은 식 (2)와 같다.

$$\arg\max_k \frac{\sum_{i=1}^{|passage|} Tf_{ij} \cdot idf(w_{ij}) \cdot P(w_{ij})}{|passage|} \quad (2)$$

$$\begin{cases} Tf_{ij} = \dfrac{tf_{ij}}{tf_{ij}+2} \\ idf(w_{ij}) = \max\left(M, \log\dfrac{N}{df_{ij}}\right) \\ P(w_{ij}) = \begin{cases} 2.0 : high \\ 1.5 : important \\ 1 : others \end{cases} \end{cases}$$

식 (2)에서 사용된 Tf, idf, P값의 정의는 식(1)에서의 정의와 유사하다.

또한 콘도르 문서 요약의 특징은 Maximal Marginal Relevance (MMR)[8, 13]를 추가한 것이다.

MMR의 특징은 문서 요약에서 적절한 문장을 선택함으로써 중복성을 감소시키는 것이다[3]. 식 (3)은 콘도르에서 사용된 MMR 알고리즘이다. 식 (3)에서 W값은 문장단위의 가중치를 의미한다.

$$\arg\max^{k}\{W_{passage_{new}} - \lambda \cdot \max sim(passage_{new}, passage_{olds})\} \quad (3)$$

(2) 온라인 문서 요약

콘도르의 두 번째 문서 요약 방법은 사용자의 질의에 포함되어 있는 용어를 중심으로 실시간으로 문서를 요약하여 그 결과를 보여주는 것이다. 즉, 질의에 나타나있는 용어들의 가중치를 임의의 비율로 증가시켜 본문에 있는 질의의 내용이 요약에 반영되도록 하였다. 식 (4)는 콘도르 온라인 요약에 적용한 질의에 포함된 용어의 가중치 계산 방법이다. j문서에 나타나는 i번째 용어의 원래 가중치 W_{ij}^{old}에 (1+b)를 곱하여 새로운 가중치 W_{ij}^{new}를 구하였다. 그 외 기본적인 알고리즘은 콘도르의 오프라인 문서 요약과 같다.

$$W_{ij}^{new} = (1+\beta) \cdot W_{ij}^{old} \quad (4)$$

2.4 문서 클러스터링

콘도르는 문서 검색 결과에 대하여 유사한 것끼리 모으는 문서 클러스터링[14] 기능을 포함한다. 문서 클러스터링 방법은 수정 K-Means 알고리즘이다[15]. 일반 K-Means 알고리즘은 클러스터의 수가 정적인데 반하여, 콘도르의 알고리즘은 그 수가 가변적이다. 즉 클러스터링하려고 하는 문서의 집단의 밀집된 정도에 따라 그 수가 동적으로 변하도록 했다.

(1) 기본적인 알고리즘

K-Means 알고리즘[16]은 이해하기 쉽고 구현이 간단하다. [그림 2]는 콘도르에서 사용된 K-Means 알고리즘이다.

1. choose k.
2. select k proto-centroids from a set of document d.
3. compute dist(d_i , c_j)
4. assign di to G_{C_j} by
   a. $\arg\min dist(d_i, c_j)$, $i=1,n$ $j=1,k$
   b. $d_i \in G_{C_j}$
      if $dist(d_i, c_j) < dist(d_i, c_l)$ for *all* $l=1,2,\ldots,k$ $(l \neq j,\ i=1,n)$
5. recomputed the centroids process
   $$c_j^{\rho} = \frac{1}{|c_j|} \sum_{i=1}^{|C|} d_i^{\rho}$$
6. check if $\max \delta(c_j, c_j^{new}) < \theta$
   then return else $ac_j = c_j^{new}$ goto 3.

[그림 2] K-Means 알고리즘

(2) 계층적 문서 클러스터링 알고리즘

콘도르의 클러스터링 구조는 계층적 구조이다. 계층적 구조를 위한 알고리즘도 K-Means 알고리즘을 기본으로 사용하고 있으며 top-down방식으로 계층을 만들어 나간다. [그림 3]은 콘도르에서 사용한 계층적 문서 클러스터링의 알고리즘이다.

1. select K centroids, c, in each cluster.
2. assign items to the nearest centroids, C_j.
3. compute the new centroids, c_j^{new}.
4. check if $\max \delta(c_j, c_j^{new}) < \theta$
   then return else $c_j = c_j^{new}$ goto 2.

[그림 3] 계층적 문서 클러스터링 알고리즘

(3) 레벨링

각 문서 클러스터에 이름을 짓는 것(레벨링)은 클러스터링에 대한 또 하나의 중요한 연구 분야이다. 콘도르에서 사용하는 레벨링 기법은 가장 단순한 기법을 사용했다. 즉, 클러스터를 대표하는 중심 벡터(centroid)에 포함되어 있는 용어 중 가중치가 상위에 있는 3개의 단어를 택하였다. [그림 4]는 레벨링 알고리즘을 보여주고 있다.

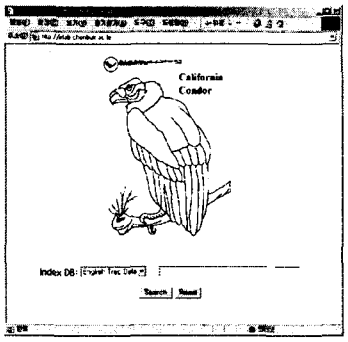
1. get the cluster centroid from each cluster.
2. sort the terms in the centroid by the weights of them.
3. take three terms whose weights are most high from the centroid.

[그림 4] 레벨링 알고리즘

3. 시스템 구현

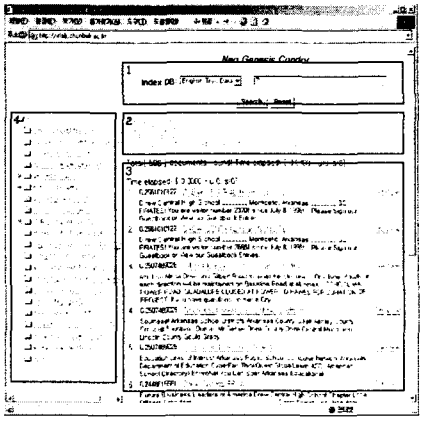
본 장에서는 구현된 화면 결과를 중심으로 설명하겠다. 콘도르는 리눅스 운영체제에서 운영되는 대용량 정보검색 시스템이다. 프로그램 코드 중 많은 부분이 저장 시스템에 관련된 부분이며 또 하드웨어 리소스를 사용하기 때문에 C언어로 구현하였다.

[그림 5]는 콘도르의 초기 화면이다. 콘도르가 비록 상용 시스템으로 개발되기는 하였지만 지속적인 개발을 위하여 테스트를 위한 시스템과 데이터를 서비스하고 있는 상용 시스템과 구분하였다. 따라서 초기 화면에서는 단순히 개발자 혹은 사용자가 정보검색을 어느 데이터베이스에서 할 것인지를 선택하기 위한 윈도우 박스를 추가했다. 물론 검색할 용어를 입력할 윈도우 박스도 제공한다.



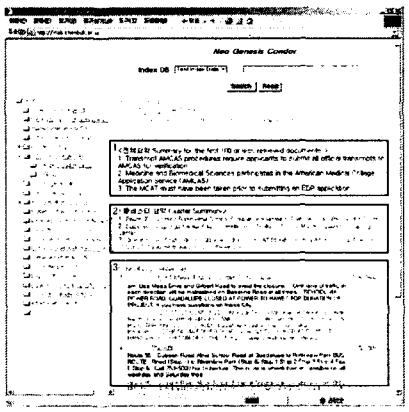
[그림 5] 초기화면

초기 화면에서 질의어가 입력되면 바로 [그림 6]과 같은 실행 화면이 나타난다. 화면 위의 1번 창은 초기 화면의 기능을 그대로 옮겨놓은 것이다. 즉 검색할 데이터베이스를 선택할 수 있는 박스와 검색할 용어를 입력할 수 있는 박스가 있다. 2번 창은 검색된 데이터의 통계 정보를 보여준다. 3번창은 검색된 데이터를 순위별로 정렬하여 문서의 제목, 위치, 요약, 그리고 미리보기 기능 등을 포함한다. 4번 창은 검색된 문서의 클러스터링 결과를 계층적 트리 구조로 가시화 하였다.



[그림 6] 실행 화면 1

사용자는 문서 클러스터를 통하여 좀더 정확한 결과 내의 검색을 할 수 있다. [그림 6]의 4번 창에 나타난 클러스터 트리에서 원하는 클러스터를 클릭함으로써 클러스터 내에 있는 문서들의 내용을 알 수 있다. [그림 7]은 클러스터를 클릭한 결과이다. [그림 7]의 1번 창은 모든 클러스터 내의 내용을 실시간으로 요약한 다중 문서 요약 결과이다. 2번 창은 현재 클릭한 클러스터에 대한 실시간 문서 요약이다. 3번 창의 내용 중 검은색의 문자열은 그 문서에 대한 오프라인 문서 요약이고 하늘색 문자열은 질의를 중심으로 요약한 온라인 문서 요약이다.



[그림 7] 실행 화면 2

콘도르는 다중어(한국어, 중국어, 일본어, 영어) 질의 처리가 가능하다. 다중어를 처리하기 위하여 우선 각 언어의 특징을 반영한 색인어 추출기를 개발하였고 각 언어의 기준 코드를 정하고 변환 모듈을 이용하여 기타 코드를 처리하였다.

이렇게 개발된 콘도르는 현재 국내 3,000만개의 웹 문서를 수집하여 테스트를 완료하였다.

4. 결론

콘도르는 학교와 산업체가 공동으로 개발한 대용량 정보검색 시스템이다. 특히 콘도르의 개발에는 정보검색분야의 최고 기술을 보유하고있는 카네기멜론 대학교의 언어연구소의 기술적 지원이 있었다. 본 논문은 기존 시스템과 차별이 있는 다중어 질의어 처리, 온라인 문서 요약, 계층적 문서 클러스터링을 위주로 작성하였다.

다중어 질의 처리는 한국어, 중국어, 일본어, 영어를 일관된 형태로 입력할 수 있다. 다중어 질의어 처리가 동양권 언어를 이해하는 사람들에 의해서 만들어졌기 때문에 정확한 결과를 사용자들에게 제공할 수 있을 것이다.

문서 요약은 오프라인과 온라인으로 이루어졌다. 두 경우 모두 MMR 기법을 적용하여 요약 결과에 중복된 의미를 제거하였다. 이렇게 함으로써 요약 결과가 훨씬 더 함축적이고 분명하다. 보통 문서 요약 시스템은 질의와 상관 없이 문서를 요약한다. 따라서 많은 경우 질의에 대한 요약 결과는 질의와 상관없는 내용이 나타나기도 한다. 그러나 질의에 포함되어 있는 용어를 중심으로 한 실시간 문서 요약은 기존 문서 요약 시스템이 갖는 단점을 해소할 수 있었다.

콘도르의 문서 클러스터링은 계층적 문서 클러스터링이다. 문서 클러스터링은 질의에 기반한 검색 결과에 따라 분류되기 때문에 실시간으로 이루어져야 한다. 또한 대량의 문서를 다루어야 하기 때문에 속도가 빠르고 유지보수가 용이한 K-Means 클러스터링 기법을 사용했다. 문서 클러스터링의 계층은 3단계의 깊이까지 가능하다. 그러나 경우에 따라 문서의 유사도에 따라 동적으로 계층 구조를 시스템이 결정하도록 하였다.

콘도르의 우수성을 유지하기 하기 위해서는 다양한 현장 테스트와 새로운 알고리즘의 개발이 필수적이다. 지금까지 국내의 3,000만개의 웹페이지 테스트가 완료되어 시스템의 안정성을 확보했다. 또한 콘도르를 계속적으로 보완하기 위하여 학교와 회사간의 산학협동이 이루어지고 있다. 현재 콘도르는 리눅스 운영체제에서 운영되고 있고 윈도우 운영체제에서도 운영할 수 있는 시스템도 개발하고 있다.

참 고 문 헌

[1] "스토리지 시스템 특집을 내며", 한국정보처리학회지, 제8권, 제4호, 2001.7.

[2] Rong Jin, Christos Falusos, and Alex G. Hauptmann. "Meta-scoring : automatically evaluating term weighting schemes in IR without precision-recall," In Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval, 2001.

[3] 김영택 외 공저, '자연언어처리', 생능출판사, 2001.9.

[4] 최유경, 안동언, 정성종, "정보검색용 다중 스레드 한국어 형태소 해석기", 2001년도 제13회 한글 및 한국어 정보처리 학술대회, 2001.10.

[5] Chih-Hao Tsai, Tsai's list of Chinese words, http://www.geocities.com/hao510/wordlist/

[6] EDR, EDR Electronic Dictionary Technical Guide, 1993.

[7] M.F. Porter, "An algorithm for suffix stripping," In Readings in Information Retrieval, Morgan Kaufmann Publishers, Inc., 1997.

[8] Jaime Carbonell and Jade Goldstein. "The use of MMR, diversity-based reranking for reordering documents and producing summaries." In Proceedings of ACM- SIGIR'98, Melbourne, Austral-ia, August 1998.

[9] 강상배, '한국어 문서의 통계적 정보를 이용한 문서요약 시스템 구현', 부산대학교, 전자계산학과, 석사학위 논문, 1998.2.

[10] Vibhu Mittal, Mark Kantrowitz, Jade Goldstein, and Jaime Carbonell, "Selecting Text Spans for Document Summarizes: Heuristics and Matrics," In Proceedings of the 16th National Conference on Artificial Intelligence, pages 467-473 1999.

[11] Jade Goldstein, Mark Kantrowitz, Vibhu Mittal, and Jaime Carbonell, "Summarizing Text Documents: Sentence Selection and Evaluation Metrics," In Proceedings of ACM-SIGIR'99, Berkeley, CA, August 1999.

[12] Therese Hand. "A Proposal for Task-Based Evaluation of Text Summarization Systems,"

In Proceedings of the ACL/EACL'97 Workshop on Intelligent Scalable Text Summarization, Madrid, Spain, July 1997.
[13] ACM Press, 2001.Jaime Carbonell and Jade Goldstein, "The use of MMR, diversity-based reranking for reordering documents and producing summaries," In Proceedings of the 21st ACM-SIGIR International Conference on Research and Development in Information Retrieval, Melbourne, Australia, 1998.
[14] Leuski and J. Allan. "Improving interactive retrieval by combining ranked lists and clustering," In Proceedings of RIAO'2000, pages 665-681, April 2000.
[15] 오형진, '클러스터 중심 결정 방법을 개선한 K-Means Algorithm의 구현', 전북대학교, 컴퓨터공학과, 석사학위 논문, 2002.8.
[16] http://nlp.korea.ac.kr/~bewise/research/KMeans.pdf

박 순 철 (Soon-Cheol Park)

1979년 인하대학교 (공학사)
1991년 미국 루이지아나주립대학 (전산학박사)
1991년 - 1993년 한국전자통신 연구원 근무
1993년 - 현재 전북대학교 전자정보공학부 부교수
(관심분야 : 정보검색, 영상정보검색, 데이터구조 및 알고리즘)

안 동 언 (Dong-Un An)

1981년 한양대학교 전자공학과 (공학사)
1987년 KAIST 전산학과 (공학석사)
1995년 KAIST 전산학과 (공학박사)
1995년 - 현재 전북대학교 전자정보공학부 부교수
(관심분야 : 정보검색, 한국어정보처리, 문서분류, 문서요약)