

복합명사의 통계적 처리에 대한 평가*

°남세진*, 이지연*, 신동욱*, 채미옥**

*충남대학교 컴퓨터공학과, **한국전자통신연구소 데이터베이스 연구실

The evaluation of statistic processing on korean compound nouns

°SeJin Nam*, JiYun Lee*, DongWook Shin*, MiOk Chae**

*Dept. of Computer Engineering, Chungnam National University

**Database Section, Electronics and Telecommunications Research Institute

요 약

한글을 대상으로 하는 검색 시스템의 경우 문서의 대부분을 차지하는 복합명사는 원칙적으로 단어와 단어 사이를 띄어 써야 하지만 붙여쓰기 또한 허용하므로 정보 검색 시스템에서는 이를 고려하여야 한다. 본 논문에서는 MIDASIR 정보검색 시스템에서 통계적인 정보를 이용하여 복합명사를 처리하는 방법을 구현하고 이를 실험을 통하여 평가하고자 한다. MIDASIR 은 크게 복합명사의 통계적인 정보를 이용하는 색인 부분과 확장 불리안 모델 및 벡터 공간 모델을 제공하는 검색 부분으로 이루어져 있다. 색인기에서는 복합명사를 처리할 뿐 아니라 고유명사와 같이 사전에 등록되지 않은 명사를 처리하는 작업을 하게 되며 검색 부분은 클래스 라이브러리로 구현되어 있어 임의의 검색 모델도 쉽게 추가 될 수 있도록 설계하였다. 본 연구에서는 KITSET 을 이용하여 불리안 모델 및 벡터 공간 모델에서의 성능을 실험을 통하여 평가하였으며, n-그램을 사용한 시스템과 비교 분석하였다.

1 서론

정보검색 시스템을 개발하는데 있어, 검색 대상이 되는 문서를 다루는 것은 전체 시스템의 수행 효율(efficiency)과 검색 효율(effectiveness)을 결정짓는 중요한 요소이다. 문서는 동영상이나 그림 화일과는 달리 크기가 몇 십 바이트에서 몇 백 메가 바이트까지 다양하며, 언어 자체의 고유한 특성으로 인해 처리하는데 고려할 점이 많이 있다. 본 논문에서는 한글의 특성을 고려한 정보 검색 시스템 MIDASIR 을 개발하는데 그 목적이 있다. MIDASIR 은 한국 전자통신연구소에서 개발한 MIDAS-III 라는 하부 저장 시스템[4]을 기반으로 하고 있는데, 이 저장 시스템은 다양한 크기의 문서를 저장하기 쉽도록 여러 가지 데이터 타입을 제공하고 있다. 또한 문서 이

외에도 그림이나 동영상, 그리고 사운드를 저장하기 위한 데이터 타입을 제공하고 있어 다양한 크기의 문서, 그림, 동영상, 사운드를 저장하기 쉽다. MIDASIR 은 MIDAS-III 위에 위치해 문서를 저장하고 색인하며, 또한 효율적으로 검색하기 위하여 설계되었다.

이러한 목적을 위하여 본 논문에서는 복합명사의 색인과 사전에 등록되지 않은 단어의 처리에 대한 방법을 제안하고자 한다. 한글에서 각 복합명사에 포함된 각 명사들은 띄어쓰기를 할 수도 있으며, 그렇지 않을 수도 있기 때문에 "정보검색"이라는 복합명사는 "정보 검색"이라고도 표기될 수 있다. 한글의 자유로운 표기법으로 인해 두 단어가 의미는 같지만 같은 단어로 인식하기란 쉽지 않다.

본 논문에서는 띄어쓰기 문제를 가지고 있는 복합명사 처리

* 본 연구는 정보통신부가 지원하는 정보통신연구개발사업 중 데이터베이스 서비스 시스템 개발 사업 수행 결과의 일부입니다.

와 사전에 등록되지 않은 고유명사와 같은 단어를 처리하는 효율적인 방법을 사용하였다. 또한 MIDAS/IR의 검색 부분을 확장 불리안 모델과 벡터 공간 모델에 공통인 기능을 가진 일반화(generic)된 클래스 라이브러리로 설계하여 각 모델을 쉽게 수행시킬 수 있으며, 실험할 수 있게 하였다.

본 연구에서 제안한 시스템의 평가를 위해 한국통신에서 개발한 KTSET을 데이터 집합으로 사용하였으며, 실험을 통해 복합명사의 통계적 처리방법[3]의 효율성을 실험하였다. 또한 MIDAS/IR의 검색 효율을 평가하기 위하여 다양한 가중치 방법을 사용하여 실험하였으며 실험 결과 n-그램 방법[6]을 사용한 다른 시스템과 성능을 비교하였다.

본 논문은 다음과 같이 구성되어 있다. 2장에서는 복합명사와 관련하여 한글 문서를 어떻게 색인할 것인가를 설명할 것이다. 3장에서는 벡터 공간 모델과 세가지 확장 불리안 모델이 이루는 클래스 라이브러리와 함께 검색 엔진에 대하여 설명한다. 4장에서는 데이터 집합으로 KTSET을 사용한 실험 결과와 다양한 가중치 방법과 복합명사 처리 방법을 사용한 MIDAS/IR 시스템의 성능을 분석하고자 한다. 그리고 마지막으로 5장에서는 결론과 향후 연구 방향에 대하여 말하고자 한다.

2 색인

MIDAS/IR은 역화일 구조를 가지는 색인 화일을 만든다. 일반적으로 역화일은 키워드의 집합과 문서 구별자, 문서를 가리키는 포인터, 그리고 가중치로 구성되어 있는 포스팅화일 그리고 문서가 실제로 저장되어 있는 문서화일로 구성되어 있다. MIDAS/IR에서는 B+ 트리를 사용하여 색인화일을 구성하며 CLOB(character large object)을 사용하여 포스팅화일과 문서 화일을 저장한다.

포스팅 화일내의 포스팅 정보는 문서 구별자와 가중치, 그리고 문서 내에서 단어 빈도 정보를 가지고 있다. 또한 빈도수 만큼의 위치 정보를 가지고 있는데 이 위치 정보는 $[P_i, S_i, W_i]$ 로 이루어져 있다. 여기에서 P_i, S_i, W_i 는 각각 단락 번호, 문장 번호, 그리고 단어가 위치한 단어 번호를 나타내는데, 이 정보는 사용자로 하여금 검색 과정에서 NEAR, ADJ, WITHPARA, WITHSENT와 같은 근접 연산자를 사용할 수 있도록 검색기에서 사용되어 진다.

다음은 한글 문서에서 어떻게 키워드를 추출하고 어떻게 가중치를 계산하는지에 대하여 설명한다. 또한 키워드를 추출하는 과정에서 형태소 분석을 하지 않고 통계적인 정보를 가지고 어떻게 복합명사를 처리하는지를 설명한다.

2.1 명사의 추출과 가중치

MIDAS/IR의 색인 과정에서는 문서 내의 명사만을 고려한다. 그래서 문서에서 한 토큰을 추출하면 먼저 단어 사전을 참조하여 등록된 단어인지를 확인하는데, 확인 과정에서는 최장 일치율을 이용한다. 최장 일치란 비교하는 토큰의 구성 단어들과 가장 길게 매칭되는 사전의 단어만을 추출하는 방법이다. 그러나 이렇게 사전을 이용해서 색인이 되는 명사를 추출하게 되면 사전에 등록되지 않은 전문용어나 고유명사는 색인으로 선택되지 않는 단점이 있어 토큰의 뒤에 붙는 조사를 검사하여 사전에 등록되어 있지 않은 색인으로 가정한다. 조사를 검사하기 위해선 미리 490개의 조사로 구성된 조사 사전을 만들어 사용하였다.

색인으로 채택된 단어는 가중치 값이 계산되어지는데, 가중치를 계산하기 위하여는 출현빈도, 장서빈도, 정규화를 고려하여야 한다. [표 1]은 각 구성 요소에 대해 잘 알려진 공식들을 나열 한 것이다[5]. 앞으로 가중치 방법을 표시하기 위해서는 각 구성요소에서 사용한 공식의 첫자를 조합한 것으로 대신 표시한다. 즉 mf 라 하면 출현빈도 값은 단어의 빈도수를 사용한 것이고, 장서빈도와 정규화는 각각 $\ln(N/n)$ 과 코사인 정규화를 사용한 것이다.

표 1. 가중치 부여 기법의 구성 요소

출현빈도	
b	1.0
n	tf 색인어의 출현빈도
a	$0.5 + 0.5 \frac{tf}{\max tf}$ 보강된 출현빈도
l	$\ln mf + 1.0$ 색인어 출현 빈도에 로그함수 적용

장서빈도	
n	1.0
t	$\ln Nn$ 색인어 출현 빈도와 역문헌 빈도의 곱 N:전체 문서들의 수 n:n은 그 색인어를 포함하고 있는 문서들의 수
정규화	
n	1.0
C	$\frac{1}{\sqrt{\sum \text{vector} Wi^2}}$ 유클리디안 벡터 길이를 이용한 코사인 정규화

본 MIDAS/IR에서는 Spark Jones[8]가 제안한 $tf \times idf$ 의 변형을 사용하였다. 즉 문서 D_j 내에서의 단어 T_i 의 중요도 W_{ij} 를 IDF_i 와 T_{ij} 의 곱에다 평준화(normalization) 값 C_j 로 나누어 계산하였다.

- IDF_i $\log_2(N/N_i) + 1$, N 은 문서 집합 내에 있는 전체 문서의 개수이며, N_i 는 단어 T_i 가 나타난 문서의 개수.
- T_{ij} 문서 D_j 에서 T_i 가 사용된 횟수.
- C_j 문서 내에서 최대의 가중치 값

가중치의 평준화를 위해 본 시스템에서는 코사인 평준화 방법을 사용하는 대신 보다 단순하면서도 비슷한 역할을 하는 한 문서 내의 최대 가중치 값을 평준화의 요소로 사용하였다. 이상으로 설명한 것들이 MIDAS/IR에서 사용하는 기본적인 가중치의 계산 방법이다. 우리는 이러한 방법을 표 1의 표기법에 의하여 n mm 방법이라 하였는데 여기서 처음의 n 은 단어의 tf (term frequency)이며, 두번째 t 는 위에서 계산한 IDF 값이다. 그리고 마지막으로 m 은 문서 내에서의 최대 가중치 값이다.

본 시스템에서는 위의 방법 외에도 [5]에 제시된 가중치 부여 방법들을 구현하였으며, n mm 방법과 함께 평가하였다.

2.2 복합명사의 처리

한국어의 경우 원칙적으로 단어별로 띄어쓰기를 원칙으로 하고 있으나, 고유 명사나 전문 용어의 경우 띄어쓰기를 할 경우 각 단어가 결합하여 가리키는 하나의 대상이 잘 파악되지 않는 단점이 있어 단위별로 띄어 쓰는 것을 허용하기도 한다. 또한 전문 용어의 경우도 마찬가지로 원칙적으로 단어별로 띄어 써야 하지만 붙여 쓰는 것 또한 허용하고 있어 "정보 검색"과 "정보검색"을 다 함께 사용할 수가 있다[11].

이러한 이유로 해서 한글을 처리하는 과정에서 위의 두 경우를 같은 것으로 인식하는 작업은 매우 중요하다. 특별히 정보검색에서 이런 복합명사를 적절하게 처리하게 되면 시스템의 검색 효율을 향상시킬 수가 있는데 이것은 문서 내에서 명사가 차지하는 개념적 중요도가 다른 품사보다 크며, 대부분 색인어로서 사용되기 때문이다.

또한 명사의 사용면에 있어서도 많은 부분에서 복합명사로 사용되기 때문이다. 만약 "정보 검색"의 각 구성 명사를 단순히 색인어로 사용하게 되면, 사용자가 "정보검색"이라는 복합명사를 사용한 질의어를 사용하였을 경우, "정보 검색"이 포함된 문서는 검색되지 않게 된다. 이보다 좀더 개선된 방법으로 위의 두 경우를 한 경우로 모두 바꾸어 처리할 수도 있지만 복합명사를 정확하게 분리해 내는 작업이 필요하게 되어 정확한 검색을 할 수가 없다.

이에 본 연구에서는 정확한 형태소 분석을 하는 대신 명사만을 고려한 색인과정을 고려하여 복합명사의 통계적인 정보 [3]를 사용하여 복합명사를 색인하고자 한다. 이 방법은 복합명사의 대부분이 일정한 형태를 가지며 사용된다는 점에 착안한 것이다. 복합명사의 형태에 대한 연구에 의하면 한글에서 거의 96%의 복합명사가 다음 5가지 경우 중 하나이다.

- type 1: 명사 + 명사 (58.4%)
- type 2: 명사 + '의' + 명사 (22.6%)
- type 3: 명사 + 명사 + 명사 (7.0%)
- type 4: 명사 + '의' + 명사 + 명사 (3.1%)
- type 5: 명사 + 명사 + '의' + 명사 (5.1%)

이러한 유형을 이용하여 복합명사를 분별해 냈다. 복합명사를 구별해내는 과정에서는 사전에 등록된 명사만을 사용하였으며, 사전에 등록되지 않은 단어의 경우에는 비록 그것이 명사일지라도 하나의 명사로 취급하고, 복합명사로는 취급하지

않았다.

위의 5 가지 유형에 해당되는 복합명사가 나타났을 경우의 색인어는 각 명사의 순서를 유지하며, 모두 붙어 있는 명사와 각 구성 명사들, 그리고 type3,4,5의 경우에는 전체 단어, 그리고 각 구성 명사 이외에도 명사를 두개씩 순서대로 나누어 저장하였다.

예를 들어 설명하자면, 문서 내에서 "정보 검색 시스템"이라는 복합명사가 있는 경우 색인기는 "정보", "검색", "시스템", "정보검색", "검색시스템", "정보검색시스템" 모두가 문서 내에 있는 것처럼 각 색인어에 대하여 빈도수를 증가시켜 가중치에 영향을 주도록 하였다. 이러한 방법으로 인해 사용자는 "정보 검색 시스템" 혹은 "정보검색 시스템"과 같은 질의어를 사용하더라도 복합명사 "정보검색시스템"이 있는 문서를 검색해 낼 수 있다.

3 검색 엔진

MIDASIR은 벡터 공간 모델 및 여러 가지 확장 불리언 모델[1]을 제공한다. 각 모델은 비록 관련 문서를 검색하는 접근 방법은 다르지만 서로가 많은 공통적인 수행 과정을 거쳐, C++의 상속 개념을 이용하여 하나의 클래스 라이브러리로 구현하였다[10]. 각 모델의 공통적으로 사용되는 데이터와 함수는 Retrieval 클래스에 정의하였으며, 각 모델은 Retrieval 클래스로부터 상속 받아 구현되었다. 그림 1은 MIDASIR에서 지원되는 클래스 구조다.

그림에서 점선으로 그려진 클래스는 단지 구조만을 제공하며 실제 구현은 상속 받은 클래스에서 이루어지는 추상 클래스이며, 그 외의 클래스들은 클래스의 완전한 구현이 이루어진 것들이다. 그림에서 볼 수 있듯이 MIDASIR은 벡터 공간 모델과 세 개의 확장 불리언 모델 MMM(Mixed Min Max), Paice, 그리고 Pnorm을 각각의 클래스에서 지원하고 있다. 벡터 공간 모델은 사용자의 질의어를 단어들의 벡터로 취급해서 얻은 질의어 벡터와 문서 벡터를 이용하여 유사도를 계산한다. 그리고 세 개의 확장 불리언 모델은 모두 같은 질의어 문법을 제공하는데, 질의어는 AND, OR, NOT 같은 불리언 연산자와 ADJ, NEAR, WITHPARA, WITHSENT 같은 근접 연산자와 피연산자인 단어들로 이루어진다.

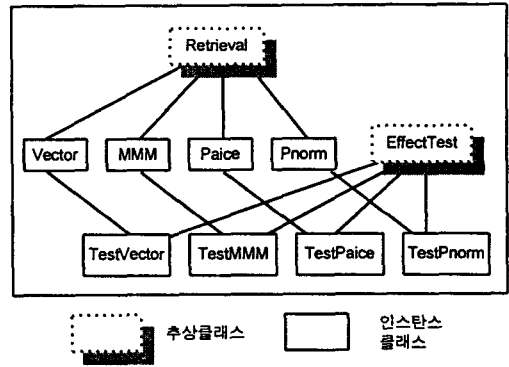


그림 1. 검색엔진에서 제공되는 클래스 계층.

검색모델을 위한 클래스들 이외에도 각 모델의 검색 효율을 평가하기 위하여 EffectTest 클래스를 제공한다. EffectTest 클래스는 질의어를 주는 일반 사용자보다는 검색 엔진의 성능을 향상시키며, 적당한 색인 방법에 관심을 갖는 시스템 개발자에게 사용되어질 것이다.

각 모델을 이렇게 상속 관계를 가진 클래스로 설계하므로 시스템의 사용자는 쉽게 새로운 검색 모델을 기존에 제공되는 클래스를 이용하여 추가 할 수 있다.

4 실험 및 분석

복합명사의 통계적 정보를 이용한 MIDASIR의 성능 평가는 검색모델중 확장 불리언 모델의 성능이 우수하다고 알려진 Pnorm 모델과 벡터공간을 사용하였으며, 테스트 문서 집합으로 KTSET을 사용하였다. KTSET은 한국통신에서 개발한 데이터 집합으로 컴퓨터와 관련된 1053 개의 문서와 50 개의 질의어로 이루어져 있다[9]. 시스템의 성능은 복합명사 처리를 한 경우와 그렇지 않은 경우에 MIDASIR가 여러 가지 가중치 방법에서 어떤 성능을 나타내는지를 살펴 이루어졌다. 색인과정에서 사용한 가중치 방법은 표2에서 보듯이 먼저 복합명사 처리를 하지 않고 단지 방법만을 ac 사용한 경우 N-그램을 사용한 다른 시스템과 비교하여 볼 때 약 4% 낮은 성능을 보였다.

그리고 여러 가지 가중치 방법을 사용하여 본 결과 가중치의 방법이 MIDASIR의 검색 효율에 영향을 주었는데, 표2에서는 표1에서의 각 구성요소의 조합으로 만들어진 가중치

기법을 이용하여 각각에 대해 성능을 평가, 정리하였다. 표에서 보는 바와 같이 복합명사를 처리하며 색인한 방법중에는 nm 방법을 평균 정확도가 53%로 가장 좋은 결과를 보였다. 또한 복합명사의 통계적인 정보가 얼마나 잘 적용되었는지를 확인하였다. 확인을 위해서는 컴퓨터 과학 분야의 전문가 2명이 KTSET 내에 있는 복합명사를 먼저 살펴보았는데, 15,292 개를 찾았다. 반면 MIDASIR의 색인에서는 12621 개의 복합명사를 검출하였다. 표3에서 각 타입에 대해서 얼마만큼의 명사들이 검출되었는지를 정리하였다. 마지막으로 복합명사 처리 모듈을 추가시킨 MIDASIR를 평가하였는데, 11-포인트 평균 정확도가 nm 방법에서는 약 3%의 차이를 내는 것을 볼 수가 있었다. 표2에서 볼 수 있듯이 복합명사를 처리하였는데도 성능의 개선이 없는 이유는 먼저 KTSET에서 포함된 불리안 질의어가 복합명사를 처리에 민감하게 결과가 나도록 구성되어 있지 않았기 때문이라 볼 수 있는데, 즉 복합명사를 처리함으로써 검색된 문서들은 많으나 검색된 문서의 대부분이 관련된 문서가 아니었기 때문에 정확도에 영향을 주었다.

표2 기중치 부여 방법별 11-포인트 평균 정확도

모델	명사 단위 색인		복합명사 처리 색인	
	확 장 불리안	백 테 공 간	확 장 불리안	백 테 공 간
Inc	0.5121	0.4576	0.4971	0.4694
atc	0.4602	0.4511	0.4740	0.4805
Itc	0.5176	0.5144	0.5118	0.4805
anc	0.4437	0.4205	0.4495	0.4248
ntm	0.5729	0.5339	0.5351	0.5241

(확장불리안 모델은 P-nom 모델 사용)

표3. 검색된 복합명사의 타입별 분류

복합명사타입	개수	비율(%)
타입 1	5944	47.1
타입 2	2049	16.2
타입 3	3160	25.1
타입 4	805	6.4
타입 5	662	5.2

합계	12621	100
----	-------	-----

5 결론

본 논문에서는 전자통신연구소에서 개발한 멀티미디어 하부 저장 시스템인 MIDAS-III를 기반으로 한 정보 검색 시스템 MIDASIR을 구현하였다. MIDAS-III는 동영상, 사운드, 이미지, 그리고 텍스트를 효율적으로 저장, 관리 할 수 있도록 다양한 데이터 타입을 제공한다. MIDASIR은 크게 인덱싱 부분과 검색 부분으로 이루어져 있는데, 색인 부분에서는 복합명사의 통계적인 정보를 이용하여 복합명사를 효율적으로 처리하며, 사전에 등록되지 않은 미 등록어를 처리하였다. 실험에 의하여 제안하는 복합명사의 처리 방법은 단순하면서도 형태소 분석 없이도 좋은 검색 효율을 보임을 알 수 있었다. MIDASIR의 검색 부분은 벡터 공간 모델과 3가지 확장 불리안 모델을 포함하는 클래스 라이브러리로 이루어져 있다. 이러한 모델은 서로간에 공유할 기능이 많으므로 구현 시 계층구조를 조정하여 공통의 기능을 가지는 추상 클래스를 정의하였다.

클래스 라이브러리를 제공하므로 사용자는 임의의 검색 모델을 사용하여 검색을 할 수 있으며, 새로운 검색 모델의 추가도 쉽게 할 수 있다. 실험을 위해서는 확장 불리안 모델 중 성능이 그 중 우수한 P-nom 모델과 벡터공간모델을 사용하였다.

본 논문에서는 MIDASIR의 성능을 평가하기 위하여 KTSET을 데이터로 사용하여 검색 엔진의 검색 효율을 측정하였는데 nm 기중치 기법을 사용하였을 때 가장 좋은 결과를 나타내었다. 또한 통계적인 방법으로 복합명사를 처리하는 모듈에서도 대부분의 복합명사를 찾아내었음을 확인하였다. 또한 이러한 실험 결과 기존의 n-그램을 사용한 방법보다 MIDASIR의 검색 효율이 더 좋게 나왔음을 알 수 있었으나, 기중치 기법이 atc 인경우에는 2%정도 낮은 성능을 보였다. 복합명사의 처리로 인한 시스템의 성능향상은 기대보다 낮게 나왔는데, 이런 정확도의 근소한 상승 내지 저하는 질의어가 데이터 집합 내의 복합명사 처리를 반영하지 못하는 것을 나타낸다고 볼 수 있으며, 또한 데이터 집합의 크기가 작은 것도 한 원인이라고 생각된다.

MIDASIR 시스템은 현재 MIDAS-III의 API를 통해 직접 사

용하는 응용 프로그램과 같이 구현되었는데, 앞으로 BADA-III 데이터 베이스 관리 시스템[12]과 밀접하게 연결되는 방향으로 진행될 것이다. 이에 두 시스템을 연동하는데 필요한 연구가 계속해서 이루어져야 할 것이다. 또한 정보 검색 시스템에 보내질 질의어와 데이터 베이스에 대한 질의어의 혼합을 위한 정보 검색 질의어와 SQL 과의 관계도 계속해서 연구되어야 할 것이다.

참고문헌

- [1] E. Fox, S. Bétrabet, M. Koushik, and W. Lee, "Extended Boolean Models," *Information Retrieval, Data Structures and Algorithms* W.B. Frakes and W. Baeza-Yates (eds), Prentice Hall, 1992.
- [2] W.B. Frakes and W. Baeza-Yates, *Information Retrieval : data structure and algorithms*, Prentice Hall, 1992.
- [3] 이창열 외, "자동 키워드 제작기 시스템 설계" 한글 및 한국어 정보처리 학술대회 93, 1993, pp. 71-77.
- [4] 이정기 외, "MIDAS-III 에 기반한 정보검색 하부구조의 설계" '95 추계 정보 과학회 학술대회, 1995, pp.259-262.
- [5] J.H. Lee, "Combining Multiple Evidence from Different Properties of Weighting Schemes," *SIGIR '95*, 1995, pp. 180-188.
- [6] J.H. Lee, "Using n-Grams for Korean Text Retrieval," *SIGIR '96*, 1996, pp.21-224.
- [7] 정상철, 신동욱, 시소리스 및 요약화일을 이용한 문서 검색시스템, '94 한글 및 한국어 정보 처리 학술대회, pp.400-408, 1994.
- [8] G. salton, M. McGill, "Introduction to modern information retrieval," McGraw-Hill Company, 1983.
- [9] 이준호 외, "정보 검색 연구를 위한 KRIST 데이터 컬렉션의 개발" 정보관리학회지 12 권 2 호 1995.
- [10] 이지연, 신동욱, 김준, "MIDAS-III정보검색시스템을 위한 C++클래스 라이브러리의 설계" '95 가을 정보과학회, pp.647-650, 1995.
- [11] 이 화승, 안병희, "한글 맞춤법 강의" 신구문화사, 1989.
- [12] M-O. Chae, K-H. Hong, M-Y. Lee, J. Kim, O-J. Cho, Y-G. Kim, and S-T. Jun, "Design of the Object Kernel of BADA-

III: An OODBMS for Multimedia Data Services," *International Workshop on Network and System Management*, Kyungju, Korea, pp.143-152, 1995.