

# 주요 검색 포털들의 통합 검색 서비스 비교 평가\*

## Comparative Evaluation of the Unified Search Services Provided by Major Korean Search Portals

박 소 연(Soyeon Park)\*\*

이 준 호(Joon-Ho Lee)\*\*\*

### 〈 목 차 〉

I. 서론	IV. 연구 결과
II. 선행 연구	1. 컬렉션 매칭
III. 연구 방법	2. 전체 적합도 평가 결과
1. 자료 수집	3. 컬렉션별 적합도 평가 결과
2. 평가 방법	VI. 결론

### 초 록

국내 검색 포털들이 제공하는 다양한 서비스들 중 가장 대중적이고 인기가 있는 서비스는 통합 검색 서비스이다. 따라서 급변하는 국내 인터넷 환경에서 검색 포털들의 경쟁력을 강화하기 위해서는 우수한 통합 검색 서비스를 이용자에게 제공하는 것이 필수적이다. 본 연구에서는 국내 주요 검색 포털들인 네이버, 다음, 야후, 엠파스의 통합 검색 서비스를 결과의 적합도 측면에서 비교, 분석하였다. 통합 검색 서비스 평가 시 본 연구에서는 실제 이용자들이 입력한 실제 질의들을 활용하였다. 또한 질의에 대한 검색 결과에서 이용자들이 실제 조회한 문서를 기록한 클릭 로그를 참고하였다. 본 연구의 결과는 향후 통합 검색 서비스의 개선에 활용되고, 이용자가 우수한 통합 검색 서비스 선택 시 참고자료로 활용될 수 있을 것으로 기대된다.

키워드: 통합 검색 서비스, 검색 포털, 컬렉션 매칭, 질의 로그, 클릭 로그 분석

### ABSTRACT

This study aims to perform an evaluation of unified search services provided by major Korean search portals, Naver, Daum, Yahoo-Korea, and Empas. These unified search services are evaluated in terms of the relevance of search results. In conducting this study, real queries that real users submitted were used. This study also utilized click logs that consist of documents users clicked and viewed. The results of this study can be implemented to the development and improvement of portal's unified search services. Users can refer to the results of this study in choosing unified search services from search portals.

Keywords: Unified Search Services, Search Portals, Collection Matching, Query Log, Click Log Analysis

\* 본 연구는 2007년도 NHN(주)의 지원을 받았음.

\*\* 덕성여자대학교 문헌정보학과 부교수(sypark@duksung.ac.kr) (제1저자)

\*\*\* 송실대학교 IT대학 컴퓨터학부 부교수(joonho@computing.ssu.ac.kr) (공동저자)

• 접수일: 2008년 2월 22일 • 최초심사일: 2008년 2월 25일 • 최종심사일: 2008년 3월 21일

## I. 서론

국내 검색 포털들은 통합 검색, 지식 검색, 블로그, 카페, 뉴스, 디렉토리, 쇼핑, 전문 자료 등 다양한 서비스들을 제공하는데, 그 중에서 이용도가 가장 높은 서비스는 통합 검색 서비스이다(이준호, 박소연, 권혁성, 2003). 통합 검색을 이용하는 경우, 이용자가 다양한 컬렉션을 동시에 검색할 수 있으므로 편리하고 효율적이라는 장점이 있다. 통합 검색 서비스는 2000년 8월 네이버가 인터넷 검색 업계에서 세계 최초로 개발한 이후 국내 타 검색 포털들도 구현, 제공하고 있는 서비스이다(임원기, 2007). 해외 검색 포털들의 경우, 2007년 6월 구글이 통합 검색과 기본 개념이 유사한 “Universal Search”를 도입, 공개한 것을 제외하면 통합 검색 서비스의 도입 및 개발에 소극적인 상황이다. 비즈니스 위크지가 구글이 NHN에게서 배워야 할 점으로 통합 검색 서비스를 주목했을 정도로(Moon and Woyke, 2006), 통합 검색 서비스는 국내에서 가장 대중적이고 인기가 있는 서비스라고 할 수 있다. 따라서 급변하는 정보 환경에서 검색 포털들의 경쟁력을 강화하기 위해서는 우수한 통합 검색 서비스를 이용자에게 제공하는 것이 필수적이다. 한편 웹 검색 분야에서 다양한 연구가 수행되어 왔지만, 국내외 선행 연구들 중 포털들이 제공하는 통합 검색 서비스에 대한 평가를 실제적으로 수행한 연구는 드문 실정이다. 또한 이용자들의 실제 질의에 근거하여 통합 검색 서비스를 비교, 평가한 연구도 존재하지 않는 실정이다.

이에 본 연구에서는 국내 주요 검색 포털들인 네이버, 다음, 야후, 엠파스의 통합 검색 서비스의 적합도를 비교, 평가하고자 한다. 이를 위하여 첫째, 통합 검색 서비스의 평가 방법을 설정하고, 이 방법에 근거하여 통합 검색 서비스를 비교, 평가하고자 한다. 둘째, 성격이 유사한 컬렉션들의 적합도를 개별적으로 비교, 분석하여 각 포털별로 상대적으로 적합도가 우세하거나, 취약한 컬렉션을 파악하고자 한다. 본 연구에서는 검색 포털들의 통합 검색 서비스 평가 시 실제 이용자들이 입력한 실제 질의를 활용하고자 한다. 또한 이용자들이 실제로 조회한 문서들로 구성된 대규모 클릭 로그에 근거하여 통합 검색 서비스의 결과를 평가하고자 한다. 본 연구의 결과는 향후 통합 검색 서비스의 개선에 활용될 수 있을 것으로 기대된다. 즉 본 연구의 결과는 포털 업체들의 효과적인 통합 검색 서비스 콘텐츠 구축 및 효율적인 검색 알고리즘 개발에 기여할 것으로 기대된다. 또한 본 연구의 결과는 이용자가 우수한 통합 검색 서비스 선택 시 참고자료로 활용될 수 있을 것이다.

## II. 선행 연구

국내 웹 검색에 관한 연구는 전산학, 경영학, 문헌정보학, 심리학, 신문방송학 등 다양한 분야에서 수행되고 있다. 이 중 검색 포털들이 제공하는 특정한 서비스의 평가에 대한 연구는 최근에 주

로 문헌정보학 분야에서 수행되어 있으며, 이들은 대체로 지식 검색 서비스의 평가에 집중되어 있다. 즉, 박주범, 정동열(2004)은 네이버, 야후, 엠파스의 지식 검색 서비스를 대상으로 실험을 통하여 응답률, 응답소요시간과 같은 간단한 성능을 비교하였다. 이들은 또한 온라인 설문조사를 통하여 이용자들의 전반적인 특성, 정보이용행태 등을 조사하고, 지식검색 서비스를 개선하기 위한 방안을 지식의 효과성과 서비스의 효율성이라는 측면에서 제시하였다. 최상희(2005)는 네이버 지식인의 질의응답문서들을 대상으로 클러스터링 기법과 수작업 주제 분류의 주제 표현 성능을 비교, 평가하였다. 장혜란, 이은태(2006)는 네 포털들의 지식 검색 서비스를 대상으로 질문의 수, 유형, 응답률, 적시성, 정확성, 적합도를 측정, 분석하였다. 박소연, 이준호, 전지운(2006)은 지식 검색 데이터베이스를 구성하는 지식 문서의 세부적인 평가 기준을 제시하고, 이에 근거하여 지식 문서의 실제 평가를 수행하였다. 이들은 또한 국내 주요 검색 포털들의 백과사전 서비스를 결과의 적합성, 최신성, 멀티미디어 제공 측면에서 분석, 평가하였다(박소연, 이준호, 2006).

한편 국외 검색 포털들의 경우 개별 컬렉션별로 분리하여 검색 결과를 제공할 뿐 통합 검색 서비스를 제공하고 있지 않으므로, 통합 검색 서비스와 관련된 연구를 찾아보기 어렵다. 국내 통합 검색에 관한 연구는 주로 특정한 검색 시스템이나 환경 또는 특정한 주제 분야의 검색 시스템을 대상으로 수행되어 왔다(김재원, 최승기, 송명원, 윤병남, 2002; 안영수, 박순영, 정원일, 배해영, 2003; 이수상, 2006). 따라서 국내의 관련 선행 연구들 중에서 이용자의 실제 질의에 근거하여 검색 포털들이 제공하는 통합 검색 서비스를 실제로 평가한 연구는 드문 실정이다. 이에 본 연구에서는 국내 주요 검색 포털들인 네이버, 다음, 야후, 엠파스의 통합 검색 서비스들에 대한 실제 평가를 수행하고자 한다.

### Ⅲ. 연구 방법

#### 1. 자료 수집

본 연구에서는 국내 주요 검색 포털들인 네이버, 다음, 야후, 엠파스를 대상으로 이들이 제공하는 통합 검색 서비스 서비스를 분석, 평가하였다. 4개의 포털들을 선택한 이유는 첫째, 이들의 인지도와 대중성 때문이며, 둘째, 이들이 통합 검색 서비스를 적극적으로 지원하고 있기 때문이다. 즉, 이들은 NCSI(National Customer Satisfaction Index), KCSI(Korean Customer Satisfaction Index) 등의 각종 고객만족도 조사 검색 포털 부문에서 공통적으로 상위 5위권에 포함되고 있다. 또한 이들은 웹 사이트 평가 및 트래픽 분석업체인 인터넷 매트릭스(<http://www.metrixcorp.com>)의 방문자 수 조사에서도 지속적으로 상위 5위에 포함되고 있다.

네이버, 다음, 야후, 엠파스의 통합 검색 서비스 서비스를 평가하기 위하여 본 연구에서는 이용자들이 직접 입력한 실제 질의를 활용하였다. 즉 2007년 8월 10일 이용자들이 네이버 통합검색 창에 입력한 전체 질의 중에서 700개의 질의를 무작위로 선정하였다. 질의 선정 이전에 로그 정제 작업을 통하여 로봇이나 프로그램이 자동으로 입력하는 질의, 편법을 이용하여 클릭 횟수를 조작하는 클릭 어뷰즈(click abuse), 클린 검색에 의해 접속이 차단된 성인성 질의 등을 제외하였다. 2007년 8월 10일의 질의들을 선택한 이유는 이 날짜가 주중의 평일을 대표할 수 있는 날짜이며, 연구를 수행하던 시점에서 최신의 질의들을 구할 수 있는 날이기 때문이다.

이용자들이 실제 입력한 질의에 대한 세부 정보가 대외비로 간주되어 확보되기 어려운 현실에서 네이버에 이용자들이 입력한 통합 검색 질의를 선택한 이유는 다음과 같다. 첫째, 국내 검색 포털 분야에서의 네이버의 위상 때문이다. 네이버는 최근 국내 검색 포털들 중 시장 점유율 조사, 방문자 수 조사, 고객 만족도 조사 등에 있어서 지속적으로 1위를 차지하고 있다. 즉, 네이버의 방문자 수 차나 시장 점유율 수치가 모두 전체 시장의 70% 이상을 차지하므로<sup>1)</sup>, 네이버에 입력된 질의들이 전체 국내 웹 이용자들이 입력한 질의들에 대한 대표성을 지니고 있다고 판단되었다. 둘째, 2007년 8월 10일의 포털별 일간 인기 검색어들은 전반적으로 매우 유사하였으며, “디워,” “진중권,” “100분토론,” “하재근,” “디워 관객 수”와 같이 영화 “디워”와 관련된 질의들을 공통적으로 포함하고 있었다. 즉, 일반적으로 국내 웹 이용자들이 주로 검색하는 주제가 엔터테인먼트나 컴퓨터 관련 질의에 집중된다는 점을 고려할 때, 네 포털들에 입력되는 질의의 주제나 성격은 유사한 것으로 사료된다. 셋째, 위에서도 언급되었듯이 이용자들의 실제 질의에 대한 상세한 정보를 구하기 어려운 현실에서, 결과의 적합도 등을 평가할 때 참고할 수 있는 질의 로그와 클릭 로그가 네이버로부터 확보되어 있기 때문이다.

700개의 질의를 선택한 이유는 하루에 네이버에 입력되는 통합 검색 질의의 수를 고려할 때, 표본 오차 95% 신뢰수준  $\pm 4\%$ 와  $\pm 5\%$ 를 허용할 경우 필요한 표본의 크기가 각각 600개와 384개로 통계학 문헌에서 제시되고 있기 때문이다(Arkin and Colton, 1963). 한편 700개 질의의 주제를 박소연, 이준호, 김지승(2005)의 연구에서 도출된 주제 분류 체계에 따라 분석한 결과는 <표 1>과 같다. 이 결과는 700개 질의를 중복 분류한 것이므로, <표 1>의 총계는 분석된 질의 수인 700개보다 약간 많다.

1) 매트릭스의 2007년도 8월 조사에 따르면, 네이버의 국내 검색 시장 점유율이 74.6%로 나타났으며, 코리안 클릭에 따르면, 네티즌들의 한 달 인터넷 검색 건 수 중 네이버에 입력된 검색 건 수가 70%에 이른다.

〈표 1〉 질의 주제 분석

주제	빈도	%
엔터테인먼트	196	26.0
게임	79	10.5
라이프스타일	75	9.9
쇼핑	71	9.4
컴퓨터/인터넷	69	9.1
교육/학문	61	8.1
지역	34	4.5
금융/경제	32	4.2
기업	26	3.4
문화/예술	25	3.3
뉴스/미디어	20	2.6
건강	18	2.4
사회	11	1.5
기관	7	0.9
성인	6	0.8
주제 불분명	25	3.4
총계	755	100

## 2. 평가 방법

통합 검색 결과의 적합도 평가 시 평가자들은 이용자가 실제로 조회한 문서를 기록한 클릭 로그에 근거하여 평가를 수행하였다. 일반적으로 웹 검색 서비스의 로그는 이용자들이 입력한 질의를 기록한 질의로그와 질의에 대한 검색 결과에서 이용자가 조회한 문서를 기록한 클릭 로그로 구성된다. 웹 검색 분야의 대부분의 선행 연구들은 검색 질의를 살펴본 연구자의 판단에 근거하여 질의의 주제를 분석하였다. 선행 연구들이 질의만을 살펴 본 후 주제를 분석한 주된 이유는 클릭 로그를 확보하지 못하였기 때문인데, 웹 검색 질의의 주제 분야가 방대하고 다양하여서 이용자가 검색 결과에서 실제로 조회한 문서를 모르는 상태에서 연구자의 판단에 근거하여 질의의 주제를 분류하기에는 한계가 있다. 이에 본 연구에서는 질의 로그와 질의에 대한 검색 결과에서 이용자가 조회한 문서를 기록한 클릭 로그에 근거하여 질의의 주제를 분석하였다. 일반적으로 개별 질의에 대한 클릭 로그는 이용자들이 조회한 수십만 건에서 수천만 건의 문서로 구성되며, 대다수 이용자들의 정보 요구가 집대성된 것으로 간주될 수 있다. 클릭 로그가 부재한 일부 질의의 경우, 이용자의 입장에서 적합하다고 판단할 수 있는 결과는 평가자들 간의 논의를 통하여 “적합”으로 평가하였다. 한편 클릭 로그의 분석 결과 복수 개의 주제를 포괄하는 질의의 경우 검색 결과가 이 중 하나의 주제라도 포함하고 있으면 “적합”으로 분류하였다.

본 연구에서는 통합 검색 서비스 평가를 위한 상세한 기준을 작성하였으며, 이러한 기준에 근거하여 문헌정보학과 전공자인 평가자 4명이 2007년 8월 22일부터 2007년 9월 18일까지 약 삼주일 동안 선정된 700개의 질의를 대상으로 네이버, 다음, 야후, 엠파스의 통합 검색 서비스 평가를 수행하였다. 평가자들은 연구자들로부터 수 개월 동안 통합 검색 평가에 관한 교육을 받고 실습을 수행하였으며, 클릭 로그에 근거하여 주제 분석 및 평가 작업을 수행하였기 때문에 평가자들의 주관적 개입할 여지는 매우 적다고 할 수 있다. 연구의 일관성을 위하여 동일한 질의를 네이버, 다음, 야후, 엠파스에 동시에 입력한 후 노출되는 통합 검색 결과에 대해 평가를 수행하였다. 또한 대부분의 웹 이용자들이 결과 화면의 첫 페이지만을 주로 조회하며 다음 결과 페이지를 조회하는 경우가 드물다는 사실을 고려하여(이준호, 박소연, 권혁성, 2003; Jansen, Spink, 2005; Jansen, Spink, 2003; Silverstein et al., 1999)<sup>2)</sup>, 본 연구에서는 질의 입력 후 첫 화면에 노출되는 결과들만을 평가 대상으로 하였다.

한편 국내 검색 포탈들은 통합 검색, 지식 검색, 블로그, 카페, 뉴스, 디렉토리 등 다양한 서비스들을 제공하며, 이러한 포탈들의 컬렉션들은 크게 사이트 제공 컬렉션과 내용 제공 컬렉션으로 구분될 수 있다(박소연, 이준호, 김지승, 2005). 사이트 제공 컬렉션은 “바로가기,” “카테고리,” “사이트,” “비즈사이트”와 같이 웹 사이트들을 검색 결과로 제공하며, 내용 제공 컬렉션은 “지식인,” “뉴스,” “웹문서,” “블로그,” “카페,” “전문지식”과 같이 특정한 주제에 관한 문서나 신문 기사, 게시판 글, 지식인에 올라간 글들을 제공한다.

클릭 로그를 통한 질의의 주제 분석 이후, 본 연구에서는 이 두 범주에 대해 별도의 적합도 평가 기준을 적용하였다. 평가 기준 적용 시 TREC(Text REtrieval Conference)내 Web Track의 검색 과제와 평가 기준을 일부 참고하였다. TREC은 1992년부터 시작된 미국 내 정보 검색 분야의 주요 학회 중의 하나이며, TREC안의 여러 특수 분야 중 Web Track에서는 참가자들에게 검색 과제로서 “topic distillation task,” “homepage finding task”와 “named page finding task”를 요구하고 있다. “topic distillation task”은 특정한 주제에 관한 적합한 문서들을 찾는 과제이며, “homepage finding task”는 질의와 일치하는 특정한 사이트를 찾는 과제이다. “homepage finding task”의 경우 결과 평가 시 특정한 사이트가 정확하게 검색되었는지의 여부만을 평가하며, “topic distillation task”의 경우 보다 복잡한 적합도 평가 단계를 거치게 된다(Crasswell and Hawking, 2005).

2) 이준호, 박소연, 권혁성의 2003년 논문에서 2003년 1월 5일부터 11일까지 네이버에서 생성된 질의 4천여 만개를 분석하였으며, 이용자들이 평균적으로 출력하는 결과 화면이 1.11개 페이지로 나타났다. Silverstein et al.의 1999년 논문에서 분석된 Altavista 의 9억 9천만개 이상 질의 중 결과의 첫 화면만을 조회하는 비율이 85.2%로 나타났다. 또한 Jansen과 Spink에 따르면(2005), 2002년 수집된 Altavista 369,350개 세션 중 첫 페이지만을 조회하는 비율이 73%로, 2002년 수집된 European AlltheWeb 345,093개 세션 중 첫 페이지만을 조회하는 비율이 83%로 나타났다.

사이트 제공 컬렉션의 결과 적합도 평가 시 질의와 결과 사이트의 일치 여부가 가장 중요한 평가 기준이므로, 본 연구에서는 <표 2>와 같이 “적합,” “부적합”의 2점 척도를 사용하였다. 내용 제공 컬렉션의 경우 결과 문서의 내용에 대한 보다 상세한 분석이 필요하므로 이분법적인 방식보다 <표 3>과 같이 “적합,” “보통,” “부적합”의 3점 척도를 사용하기로 결정하였다. 내용 제공 컬렉션들에 대해 3점 척도를 사용한 이유는 1950년대 말 이후부터 정보 검색 분야에서 적합도에 등급 또는 정도가 존재하므로 이분법적 평가 방법보다는 복수의 등급(degree)을 허용하는 것이 필요하다는 논의가 활발히 전개되어 왔기 때문이다(National Academy of Sciences 1959; Goffman 1964; Hillman 1964; Saracevic 1975; Saracevic 2007). 적합도 평가 시 복수의 등급을 허용할 경우 적합 정도를 보다 섬세하게 표현할 수 있으며, 평가자 간에 발생할 수 있는 관점의 차이나 생각의 차이가 양극화되는 위험을 방지할 수 있다는 장점이 있어(맹성현 외 1999), 많은 선행 연구들이 3점 척도나 5점 척도를 통한 적합도 평가를 수행하여 왔다(노정순 2004; 맹성현 외 1999).

<표 2> 사이트 제공 컬렉션의 적합도 평가 기준 및 예

평 가	내 용
적합	- 질의와 일치하는 사이트에 대한 결과가 제공될 때 예) 질의: 다음 결과: www.daum.net
부적합	- 질의와 일치하지 않는 사이트에 대한 결과가 제공될 때 - 질의의 일부가 광고 사이트의 키워드에 등장하지만, 질의와 무관한 사이트에 대한 결과가 제공될 때 예) 질의: 케이블 가입자 수 결과: 디지털 랜드 - 컴퓨터 주변기기 쇼핑몰, 공유기, CCTV, USB, 네트워크 장비, 허브, 케이블 판매. www.idigitalland.co.kr 예) 질의: 옥션 결과: 간지나라 - 지금은 이벤트중, 옥션, 일본 및 유럽 스타일 구제 일본풍, 거리 패션, 예쁜 옷. 이벤트/경품 반스 2005년판 밀리터리 봄버판매, 리뉴얼 및 사은품 이벤트

<표 3> 내용 제공 컬렉션의 적합도 평가 기준 및 예

평 가	내 용
적합	- 결과의 전체적인 내용이 질의와 일치하거나 밀접한 관련이 있을 때 예) 질의: 서울대학병원 결과: 서울대학 병원 위치, 진료시간, 의료진 등의 내용 포함
보통	- 결과의 내용이 질의와 일부 관련이 있을 때 예) 질의: 제모 비용 변화 결과: 제모 비용에 관한 결과만 제공될 때(변화에 관한 정보는 제공 안 됨) - 상품명에 질의인 경우 상품명과 일부만 일치하는 결과가 제공될 때 예) 질의: 슬래드 1300 결과: “슬래드”에 관한 일반 정보만 제공될 때(슬래드 1300 버전에 관한 구체적인 정보가 제공 안 됨)
부적합	- 결과의 내용이 질의와 전혀 일치하지 않거나 관련이 없을 때 예) 질의: 주온 결과: 주전층

## IV. 연구 결과

### 1. 컬렉션 매칭

통합 검색 서비스의 객관적인 평가를 위하여 포털별로 통합 검색 서비스에서 제공되는 컬렉션들을 파악한 후, 유사한 성격을 지닌 컬렉션을 파악하는 작업이 요구된다. 2006년 8월과 2007년 8월에 네이버, 다음, 야후, 엠파스의 통합 검색 결과에 노출되는 컬렉션 수를 각각 비교한 결과는 <표 4>와 같다. 본 연구의 수행 시점인 2007년 8월 기준으로, 엠파스에서 노출되는 컬렉션들이 64개로 가장 많았으며, 야후에서 노출되는 컬렉션 수가 23개로 가장 적었다. 2006년 8월에도 역시 엠파스에서 노출되는 컬렉션들이 가장 많았으며, 야후에서 노출되는 컬렉션 수가 가장 적었다. 엠파스와 다음이 제공하는 컬렉션들 중 가장 큰 비중을 차지하는 것은 “자체 제작 콘텐츠”이며, “자체 제작 콘텐츠”란 “공연 정보,” “영화 정보,” “방송 정보,” “기업정보” “게임 정보” 등과 같이 미리 지정해 놓은 키워드에 대해 포털들이 직접 제작하는 콘텐츠들을 통칭하여 의미한다.

1년 전에 비해 컬렉션의 변화가 가장 컸던 포털은 엠파스였으며, 이어서 다음, 네이버, 야후의 순으로 조사되었다. 즉, 엠파스의 통합 검색 서비스에서 제공되는 컬렉션 수가 가장 많았으며, 컬렉션의 변화도 가장 컸다. 그러나 엠파스의 경우처럼 컬렉션의 수가 지나치게 많거나 컬렉션의 변화가 빈번한 것은 이용자에게 혼돈을 초래할 수 있어서 바람직하지 않은 것으로 사료된다. 야후의 경우 제공되는 컬렉션의 수가 일정하고, 컬렉션의 변화도 가장 적은 것으로 나타났다. 1년 동안 모든 포털들의 컬렉션 수가 증가하였는데, 이는 앞에서 언급된 “자체 제작 콘텐츠” 항목의 증가에 기인하는 것으로 보인다.

한편, <표 5>는 2007년 8월 기준으로 4개의 포털들이 제공하는 컬렉션들 중 기능이나 제공 콘텐츠가 유사한 컬렉션들을 범주화한 결과를 보여준다.

<표 4> 통합 검색 컬렉션 수

(괄호는 자체 제작 콘텐츠 통합 시의 컬렉션 수)

	네이버	다음	야후	엠파스
2006년 8월	29(27)	32(29)	22(22)	35(30)
2007년 8월	43(24)	54(34)	23(21)	64(32)



〈표 5〉 통합 검색 컬렉션 매칭<sup>3)</sup>

네이버	다음	야후	엠파스
바로가기	바로가기	바로가기	바로가기
자체 제작 콘텐츠	자체 제작 콘텐츠	자체 제작 콘텐츠	자체 제작 콘텐츠
스폰서링크	스폰서링크·와이드링크	스폰서링크	스폰서링크
파워링크	프리미엄링크	프리미엄사이트	스피드업
다이렉트검색	다이렉트검색	다이렉트검색	다이렉트검색
플러스프로	스페셜링크·스폰서박스	플러스링크	스페셜링크·프리미엄매칭
	자료실		자료실
이미지	이미지	이미지·사진	이미지
			랭킹평가
사전	사전	사전	사전
지역정보	지역정보 위치정보	거기 거기플러스	지역
음악·좋은 노래 뮤직컬렉션	음악	뮤직	음악·배경음악
동영상	동영상	동영상	동영상
책	책		책
웹페이지	웹문서		웹문서
	금융상품		
			리포트
취업	취업정보		열린취업
		이슈n	
지식인	신지식	지식	열린지식
블로그	블로그	블로그	열린블로그
카테고리	카테고리	카테고리	카테고리
	플래닛		미니홈피·타운
지식쇼핑	쇼핑검색·쇼핑하우	쇼핑	쇼핑
비즈사이트	비즈사이트	비즈링크	비즈링크
사이트	사이트	웹	사이트
카페	카페이름·카페글·카페초이스		열린커뮤니티
전문자료	지식파일		전문지식·전문자료
최신뉴스	최신뉴스	뉴스	최신뉴스
	게시판		열린게시판
	오늘의카페		

## 2. 전체 적합도 평가 결과

국내 주요 검색 포털들의 통합 검색 서비스를 평가하기 위하여, 본 연구에서는 2007년 8월 10일 의 질의 로그로부터 선정된 700개의 질의를 활용하였다. 연구의 일관성을 위하여 동일한 질의를 네이버, 다음, 야후, 엠파스에 동시에 입력한 후, 첫 화면에 노출되는 결과들에 대해 평가를 수행하였다. 첫 화면에 노출되는 컬렉션들의 결과의 적합도를 위에 제시된 적합도 평가기준에 따라 개별

3) 자체 제작 콘텐츠의 세부 항목들은 포함되지 않음.

적으로 평가하였다. 본 연구의 수행 시점에 네이버는 43개, 다음은 54개, 엠파스는 64개, 야후는 23개의 컬렉션들을 제공하고 있었다. 따라서 질의별로 184개의 컬렉션들에 대해 첫째, 결과가 노출되는지 여부와 둘째, 결과가 노출되는 경우 적합도의 정도를 평가하였다. 700개의 질의를 분석 대상으로 하였으므로, 총 128,800회의 평가 작업이 수행되었다.

또한 개별 컬렉션들의 결과를 통합한 전반적인 적합도도 평가하였으며, 그 결과는 <표 6>과 같다. 이 표에서, “결과 없음”이란 특정한 질의에 대해 어떠한 컬렉션에서도 결과가 노출되지 않은 경우이다. “적합”이란 특정한 질의에 대해 통합 검색 서비스에서 제공하고 있는 다양한 컬렉션 들 중 적어도 하나의 컬렉션에서 적합한 결과가 나온 경우이며, “보통”이란 평가 결과의 최대값이 “보통”인 경우이며, “부적합”이란 평가 결과의 최대값이 “부적합”인 경우이다. <표 6>에 따르면 네이버 통합 검색 서비스의 전반적인 적합도가 가장 높은 것으로 나타났으며, 근소한 차이로 야후가 2위를 차지하였고, 3, 4위는 다음, 엠파스 순으로 나타났다. 한편 “결과 없음”의 비율은 엠파스, 네이버가 거의 비슷하게 낮았으며, 야후가 가장 높게 나타났다.

<표 6> 적합도 평가 결과(중복 질의 포함 시)

총질의수: 700개

	없음	적합	보통	부적합
네이버	16(2.3%)	628(89.7%)	21(3.0%)	35(5.0%)
다음	26(3.7%)	615(87.9%)	19(2.7%)	40(5.7%)
야후	56(8.0%)	625(89.3%)	13(1.9%)	56(8.0%)
엠파스	15(2.1%)	605(86.4%)	23(3.3%)	57(8.1%)

전체 700개의 질의 중 중복되는 질의를 제거한 후 적합도 결과를 평가한 결과는 <표 7>과 같다. <표 7>의 순위는 중복 질의를 포함한 <표 6>의 순위와 동일하다. 즉 네이버의 적합도가 가장 높으며, 다음은 야후, 다음, 엠파스 순이다.

<표 7> 적합도 평가 결과(중복 질의 제거 시)

총질의수: 662개

	없음	적합	보통	부적합
네이버	16(2.4%)	590(89.1%)	21(3.2%)	35(5.3%)
다음	26(3.9%)	577(87.2%)	19(2.9%)	40(6.0%)
야후	6(0.9%)	587(88.7%)	13(2.0%)	56(8.5%)
엠파스	14(2.1%)	568(85.8%)	23(3.5%)	57(8.6%)

### 3. 컬렉션별 적합도 평가 결과

통합 검색 서비스의 적합도를 개별 컬렉션 별로 세부적으로 분석한 결과는 <표 8>과 같다. 즉 <표 8>에서는 포털들이 제공하는 다양한 컬렉션들 중 성격이 유사한 컬렉션들의 적합도를 개별적으로 비교, 분석하고 있다. 이 때 적어도 세 개 이상의 포털들이 공통으로 제공하는 컬렉션들을 분석 대상으로 하였다. 예를 들어 모든 포털들이 “바로가기” 서비스를 제공하고 있으므로, “바로가기” 서비스 결과의 적합도를 <표 1>의 기준에 따라 평가하였다. 그 결과, 네 포털들의 “바로가기” 서비스 중 야후의 “바로가기” 서비스의 적합도가 가장 높은 것으로 나타났다.

<표 8>에 따르면, 네이버가 다른 포털들에 비해 강세를 보이고 있는 컬렉션들은 지식인, 웹페이지, 디렉트검색, 사전, 카페, 책, 이미지 등이다. 다음이 우세한 컬렉션들은 스페셜 링크, 비즈사이트, 지역정보, 게임정보, 기업정보, 영화정보 등으로 나타났다. 야후는 바로가기, 스폰서링크, 웹, 동영상, 쇼핑 등에서 다른 포털들보다 강세를 보였다. 엠파스가 우세한 컬렉션들은 최신 뉴스, 열린취업, 열린 블로그, 카테고리, 전문자료, 공연, 쇼핑 등으로 나타났다. “쇼핑”서비스의 경우 야후와 엠파스가 공동 1위를, “파워링크/스피드업”서비스는 네이버, 엠파스가 공동 1위를 차지하였다. 엠파스의 경우 “열린” 서비스들이 타 포털들에서 제공하는 검색 결과를 통합하여 제공하기 때문에 이들 서비스들의 적합도가 상대적으로 높게 나타난 것으로 보인다. 또한 엠파스가 우세한 컬렉션들이 다소 많은 것은 엠파스에서 제공하는 컬렉션들이 가장 많다는 사실에도 부분적으로 기인한다. 개별 컬렉션의 적합도를 비교한 이 결과는 각 포털들이 타 포털들보다 우세한 컬렉션들을 더욱 강화하고 상대적으로 취약하게 나타난 컬렉션들의 적합도를 제고하는 데에 활용될 수 있을 것으로 보인다.

<표 8> 개별 컬렉션 적합도 비교 분석(\*은 공동 1위, \*\*은 자체 제작 콘텐츠들임)

네이버의 적합도가 1위인 컬렉션들	다음의 적합도가 1위인 컬렉션들	야후의 적합도가 1위인 컬렉션들	엠파스의 적합도가 1위인 컬렉션들
디렉트검색 파워링크* 책 사전 지식인 카페 웹페이지 이미지 방송정보**	스페셜링크 지역정보 비즈사이트 게임정보** 기업정보** 영화정보** 요리법** 프로필**	바로가기 동영상 스폰서링크 웹 쇼핑* 테마검색**	스피드업* 음악 열린취업 열린블로그 카테고리 최신뉴스 전문자료 쇼핑* 공연**

## V. 결 론

본 연구에서는 통합 검색 서비스의 평가 방법을 제시하고, 이 방법에 따라 국내 주요 검색 포털 들인 네이버, 다음, 야후, 엠파스의 통합 검색 서비스를 비교, 평가하였다. 본 연구는 이러한 평가 시 실제 이용자들이 직접 입력한 실제 질의들을 활용하였다. 연구 결과, 전반적인 통합 검색 적합도의 경우, 네이버가 가장 높은 것으로 나타났으며, 그 다음은 야후, 다음, 엠파스 순으로 나타났다. 시장 점유율에서 2위인 다음보다 야후가 더 순위가 높으며, 근소한 차이로 2위를 차지한 결과가 주목할 만하다. 또한 야후의 경우 시간이 경과해도 제공되는 컬렉션의 수나 종류가 가장 일관성이 높은 것으로 나타났다.

또한 본 연구에서는 포털들이 제공하는 다양한 컬렉션들 중 성격이 유사한 컬렉션들의 적합도를 개별적으로 비교, 분석하였다. 네이버가 다른 포털들에 비해 강세를 보이고 있는 컬렉션들은 지식 인, 웹페이지, 디렉트검색, 사진, 카페, 책, 이미지 등이다. 다음이 우세한 컬렉션들은 스페셜 링 크, 비즈사이트, 지역정보, 게임정보, 기업정보, 영화정보 등으로 나타났다. 야후는 바로가기, 스폰 서링크, 웹, 동영상, 쇼핑 등에서 다른 포털들보다 강세를 보였다. 엠파스가 우세한 컬렉션들은 최 신 뉴스, 열린 취업, 열린 블로그, 카테고리, 전문자료, 공연, 쇼핑 등으로 나타났다.

본 연구의 결과는 향후 포털들의 통합 검색 서비스의 개선에 활용될 수 있을 것으로 기대된다. 즉, 개별 포털들의 보다 효과적인 콘텐츠 구축 및 효율적인 통합 검색 알고리즘 개발에 기여할 것 이다. 또한, 본 연구의 결과는 이용자가 우수한 통합 검색 서비스 선택 시 참고자료로 활용될 수 있을 것으로 기대된다. 한편, 통합 검색 서비스는 국내 포털들의 주력 서비스이며, 통합 검색 항목 이나 콘텐츠, 결과 노출 방법 및 순서 등이 지속적으로 변화되고 있으므로, 향후 통합 검색 서비스 에 대한 주기적인 평가가 요청된다.

본 연구의 수행 결과 향후 연구가 요구되는 사항들은 다음과 같다. 첫째, 본 연구에서 사용된 평 가 방법에 대한 보완 및 세분화가 가능하다. 예를 들어, 본 연구에서는 통합 검색에서 노출되는 첫 화면만을 조사 대상으로 하였으나, 보다 심층적인 평가를 위해서는 둘째 화면에서 노출되는 결과에 대한 평가도 병행하는 것이 필요하다. 둘째, 본 연구에서는 국내 검색 포털들의 통합 검색 서비스를 결과의 적합도 측면에서 분석하였다. 향후 연구에서는 본 연구에서 포함되지 않았던 기준들에 따라 통합 검색 서비스를 분석, 평가하는 것이 필요할 것으로 보인다. 예를 들어 통합 검색 서비스의 인터 페이스, 이용자 지원, 부가 서비스 제공 등의 측면에서 평가할 수 있을 것이다. 셋째, 이용자들의 통합 검색 서비스에 대한 만족도 조사를 위해서 인터뷰, 관찰 등을 통한 별도의 보완작업이 요구된 다. 마지막으로, 본 연구에서는 하루 동안 입력된 질의를 대상으로 통합 검색 서비스의 결과를 평가 하였다. 향후 연구에서는 보다 장기간에 걸쳐 수집된 질의를 통한 평가 작업이 요구된다.

## 참 고 문 헌

- 김재원, 최승기, 송명원, 윤병남. “국가지식정보 통합검색시스템 설계 및 구현.” 한국정보과학회 학술 발표논문집, 제29권, 제1호(2002), pp.526-528.
- 노정순. “Invisible Web 탐색도구의 성능 비교 및 분석.” 정보관리학회지, 제21권, 제3호(2004. 9), pp.203-225.
- 맹성현, 이석훈, 이준호, 이응봉, 송사광. “정보 검색 시스템 평가를 위한 균형 테스트 컬렉션 구축.” 정보관리학회지, 제16권, 제2호(1999, 6), pp.135-148.
- 박소연, 이준호, 김지승. “클릭 로그에 근거한 네이버 검색 질의의 형태 및 주제 분석.” 한국문헌정보학회지, 제39권, 제1호(2005, 3), pp.266-278.
- 박소연, 이준호. “국내 주요 검색 포털들의 백과사전 서비스 비교 평가.” 한국도서관정보학회지, 제37권, 제2호(2006, 6), pp.217-230.
- 박소연, 이준호, 전지운. “지식 검색 서비스 개선을 위한 문서의 적합도 및 신뢰도 분석.” 한국문헌정보학회지, 제40권, 제2호(2006, 6), pp.299-314.
- 박주범, 정동열. “지식검색서비스 이용에 관한 실증적 연구.” 정보관리학회지, 제21권, 제3호(2004, 9), pp.83-98.
- 안영수, 박순영, 정원일, 배해영. “GML 문서의 통합 지리 정보 검색을 위한 XQuery의 확장.” 한국정보과학회 학술발표논문집, 제30권, 제1호(2003), pp.617-619.
- 이수상. “디지털도서관의 통합검색 방식에 관한 연구.” 한국도서관정보학회지, 제37권, 제2호(2006, 6), pp.127-144.
- 이준호, 박소연, 권혁성. “질의 로그 분석을 통한 네이버 이용자의 검색 행태 연구.” 한국정보관리학회지, 제48권, 제2호(2003, 6), pp.27-40.
- 임원기. “구글 야후 등 통합 검색 지식iN ‘네이버 따라하기.’ 한국경제신문, (2007. 5. 23).  
<<http://www.hankyung.com/news/app/newsview.php?aid=2007052347821>> [인용 2008. 3. 14].
- 장혜란, 이은태. “인터넷 지식검색의 질문응답서비스 성능평가.” 정보관리연구, 제37권, 제2호(2006, 6), pp.137-156.
- 최상희. “문서구조를 이용한 질의응답문서 클러스터링에 관한 연구.” 한국문헌정보학회지, 제39권, 제4호(2005, 12), pp.105-118.
- Arkin, H., and Colton, R. *Tables for Statisticians*. New York: Barnes & Noble Inc., 1963.
- Crasswell, N., and Hawking, D. *Overview of the TREC 2004 Web Track*.  
<<http://trec.nist.gov/pubs/trec13/papers/WEB.OVERVIEW.pdf>>. 2005. [Cited

2008. 2. 20].
- Goffman, W. "On relevance as a measure." *Information Storage and Retrieval*, Vol.2, No.3(1964), pp.201-203.
- Hillman, D. J. "The Notion of Relevance(1)." *American Documentation*, Vol.15, No.1(1964), pp.26-34.
- Jansen, B. J., and Spink, A. "How are we searching the World Wide Web? A comparison of nine search engine transaction logs." *Information Processing and Management*, Vol.42, No.1(2005), pp.248-263.
- Jansen, B. J., and Spink, A. "An analysis of web documents retrieved and viewed." *The 4th International Conference on Internet Computing*(2003), Las Vegas, Nevada, pp.65-69.
- Moon, I., and Woyke, E. "NHN: The little search engine that could." *Businessweek* (2006, January, 30).
- National Academy of Sciences. *Proceedings of the International Conference on Science Information*, 2 Vols.(1959). National Academy of Sciences, Washington, D. C.
- Saracevic, T. 1975. "Relevance: A review of and a framework for the thinking on the notion in information science." *Journal of the American Society for Information Science*, Vol.26, No.6(1975), pp.321-343.
- Saracevic, T. 2007. "Relevance: A review of the literature and a framework for thinking on the notion in information science. Part II: nature and manifestations of relevance." *Journal of the American Society for Information Science and Technology*, Vol.58, No. 3(2007), pp.1915-1933.
- Silverstein, C., Henzinger, M., Marais, H., and Moricz, M. "Analysis of a very large web search engine query log." *SIGIR Forum*, Vol.33, No.1(1999), pp.6-12.