

工學碩士 學位論文

인터넷 검색포탈의 시장구조분석에
관한 연구

Hierarchical Market Structures on Internet Search Portal

指導教授 辛 昌 勳

2003年 2月

韓國海洋大學校 大學院

物流시스템工學科

李 池 勳

本 論文을 李池勳의 工學碩士 學位論文으로 認准함

委員長 工學博士 申宰榮 ⑩

委 員 工學博士 李哲榮 ⑩

委 員 工學博士 辛昌勳 ⑩

2002年 12月

韓國海洋大學校 大學院

物流시스템工學科

李池勳

목 차

제 1 장 서 론.....	1
제 2 장 검색포탈의 정의.....	4
제 3 장 이론적 고찰.....	6
3.1. 인터넷 검색포탈에 관한 선행연구.....	6
3.2. 시장구조분석에 관한 선행연구.....	8
제 4 장 연구모형.....	10
4.1. 검색포탈의 시장구조분석모형.....	10
4.2. 예제 모형.....	14
4.3. 모형의 확장.....	18
4.4. 분석절차.....	20
제 5 장 실증분석.....	21
5.1. 키워드(Key-Word)와 검색포탈의 선정.....	22
5.2. 자료 수집 및 데이터베이스구축.....	23
제 6 장 결론 및 미래 연구 방향.....	31
참 고 문 헌.....	35

표 목 차

<표 1> 국내 포털사이트에서의 검색포털의 순위와 이용자수	6
<표 2> 검색포털서비스의 URL 소속의 예	1 4
<표 3> 검색포털의 개별 URL의 존재확률의 예	1 6
<표 4> 검색포털서비스 간 전환율의 예($\hat{P}_i(j)$)	1 8
<표 5> 데이터베이스 구조의 예(Yahoo)	2 3
<표 6> 수집된 전체 데이터의 수	2 4
<표 7> 전체 검색포털의 전환율	2 5
<표 8> 검색포털의 키워드광고의 판매가격	3 0

그림 목 차

<그림 1> 검색포탈의 비즈니스모델변화	5
<그림 2> 검색포탈서비스간 중복관계의 예	1 5
<그림 3> 분석절차	2 1
<그림 4> 검색포탈서비스의 시장구조모형 (키워드, 전체, $\lambda = 0$)	2 7
<그림 5> 검색포탈서비스의 시장구조모형 (키워드, 전체, $\lambda = 0.05$)	2 8

Hierarchical Market Structures on Internet Search Portal

Lee Ji-Hoon

Department of Logistics Engineering
Graduate School of Korea Maritime University

Abstract

Internet search services hold the key to helping consumers to access and sort the wealth of information. As new search Providers enter successively, they are trying to take a competitive advantage in search service market. We estimated the structure of search service market to regard the degree of overlap among search engines' results as a similarity measure.

This study identified a market structure in the level of the whole URLs and different structures in the level of search keywords. In conclusion, these models could be a useful tool to aid in developing tool to in developing search service strategy formulation.

제 1 장 서 론

한국인터넷정보센터(KRNIC)에서 7월에 발표한 국내 인터넷 이용자수는 2,565만 명으로 전체 인구의 58%에 달하며 미국의 온라인설문조사기관인 Nielsen/NetRating의 전세계 28개국을 대상으로 인터넷 이용 현황을 조사한 보고서에 따르면, 한국 가정의 인터넷 이용자수는 세계 4위에 랭크 되었다. 이러한 보고서들을 종합해 보면 국내 인터넷의 사용은 이미 성숙기에 접어들고 있으며 사용자의 성별, 연령층이 라디오, TV에 이어 새로운 미디어(Media)의 역할을 하고 있다.

현재 인터넷을 통해 다양한 서비스가 제공되고 있으며 이용자들은 이를 통해 가치를 창조하고 향유하기 시작했고, 그런 서비스를 이용하는데 있어 가장 중요한 것이 있다면 바로 검색포탈을 꼽을 수 있다. 조지아 연구센터의 GVU(Graphic, Visualization, and Usability) 보고서에 따르면 인터넷에서 특정 사이트를 발견하는 방법 중 검색포탈을 통해 찾아가게 되는 비율이 가장 높게 나타났다(GVU 1999). 이는 검색포탈의 중요성을 새삼 일깨워주고 있다. 인터넷에서 정보의 생산량이 폭발적으로 증가하고 있음에 따라 유용한 정보를 찾는 일이 더욱 힘들어 질 것이므로 검색포탈에 대한 의존도는 갈수록 높아질 수밖에 없을 것이다.

인터넷 포탈서비스 중에서 가장 많은 이용자와 사용빈도를 보이는 검색포탈 서비스는 1994년 야후(Yahoo)¹⁾가 처음으로 등장한 이래 지금까지 국내외적으로 급격히 증가하고 있다. 국내의 경우 96년 초부터 검색포탈을 제공한 심마니

¹⁾ 야후(<http://www.kr.yahoo.com>)

(Simmani)²⁾ 와 네이버(Naver)³⁾, 한미르(Hanmir)⁴⁾ 가 검색포털을 제공하고 있으며 외국계 검색포털인 알타비스타(Altavista)⁵⁾와 라이코스(Lycos)⁶⁾ 그리고 보다 진보된 검색기법인 자연어검색기능을 제공하는 엠파스(Empas)⁷⁾ 등 수많은 검색포털이 등장하고 있어서 향후 검색포털의 경쟁이 치열해 질 것으로 예상된다. 또한 각각의 검색포털들은 이용자들을 어떤 특정한 목적지에 도달하게 하는 포털사이트의 역할에서 모든 업무를 처리할 수 있는 허브사이트로의 성장을 위해 보다 나은 검색정보와 보다 다양한 콘텐츠 제공을 위해 노력하고 있다.

이러한 노력이 학계에서도 진행되고 있는데, 대부분의 선행연구들은 검색포털의 성능평가에 치우쳐 있다. Lake(1997)는 일반적인 검색질의어를 통해 얻어진 웹사이트의 주소인 URL(Uniform Resource Locator, 이하 URL로 칭함)을 이용해 적합성이나 정확성 등의 기준으로 검색포털을 평가하였다. 평가에 대한 개선을 위해 자연어 처리기술, 각종 알고리즘을 통한 해법 등을 제시하고 있다. 이를 통해 구글(google)⁸⁾의 페이지랭크(PageRank)와 같은 새로운 알고리즘들이 현재 이용되고 있다. 최근 Bradlow and Schmittlein(2000)의 연구에서는 특성별로 나누어진 검색어의 질의를 통해 얻어진 URL을 검색포털 간의 거리개념으로 변환하여 검색포털의 수평적인 시장구조분석을 하였다. 이는 기존의 연구와는 달리 검색포털의 내부 알고리즘의 개발이 아닌 이용자의 측면을 다루고 있으며 검색포털에 대한 새로운 연구방안을 제시하고 있다.

본 연구에서는 검색포털의 가장 기본이 되어왔고 현재에도 가장 핵심이 되는 검색기능을 연구대상으로 질의어에 대한 검색결과인 URL의 중복 정도를 활용하여 검색포털시장의 구조를 파악하였다. 기존의 연구에서는 검색포털시장 전

2) 심마니(<http://www.simmani.com>)
3) 네이버(<http://www.naver.com>)
4) 한미르(<http://www.hanmir.com>)
5) 알타비스타(<http://www.altavista.co.kr>)
6) 라이코스(<http://www.lycos.co.kr>)
7) 엠파스(<http://www.empas.co.kr>)
8) 구글(<http://www.google.com>)

체를 다루기 보다는 검색포탈 각각을 다루고 있고, 검색포탈의 디자인이나 자체적으로 제공되고 있는 컨텐츠의 비교만이 진행되어왔다. 이러한 기존 연구와는 달리 본 연구에서는 검색포탈시장이 구조화되지 않았다는 가설과 본 연구에서 상정된 구조화된 가설과의 비교검정을 통해서 검색포탈의 시장구조를 파악하였다. 시장구조를 파악함으로써 검색포탈시장을 이루고 있는 서비스간의 관계를 알 수 있었고, 세부적으로 검색에 사용된 키워드별로의 시장구조 또한 알게 되었다. 이러한 시장구조분석을 통해 검색포탈간의 유사성과약을 통해서 제품속성, 즉 향후 검색포탈에서 제공하게 될 각각의 컨텐츠(Contents) 차원에서의 경쟁상황을 유추할 수 있었다. 본 연구의 모형 및 결과는 새로운 컨텐츠의 추가 및 삭제 등의 차별화 전략 수립에 도움을 줄 수 있으며, 이용자관점에서의 검색포탈간 대체성 분석과 상호 보완적인 의미를 제시하고 있다. 또한 키워드광고와 같이 검색포탈에 상위에 랭크(Rank) 되기 위해 자신의 사이트 혹은 쇼핑몰을 유료로 등록하는 광고주들에게도 보다 전략적인 광고방법을 제시할 수 있다.

본 논문의 내용을 요약하면 다음과 같다. 제 2 장에서는 인터넷 포탈의 정의와 검색포탈에 관한 현황에 대해서 살펴보았다. 이를 통해 연구의 대상이 되는 검색포탈을 선택하였다. 제 3 장에서는 검색포탈을 대상으로 한 연구와 시장구조 분석에 관한 이론적 고찰을 다루고 있다. 문헌고찰을 통해서 자료의 수집과 분석 방법을 수립하게 된다. 제 4 장은 기존의 강제상표전환자료를 이용한 시장구조분석을 토대로 본 연구에 이용될 모델과 가설설정 및 검정에 관하여 다루었다. 제 5장은 본 논문에서 제시한 검색포탈의 시장구조모형을 실증적으로 분석하기 위해 가장 대표적인 검색포탈 8개를 대상으로 15가지 질의어를 이용하였으며, 결과물인 URL을 이용해서 검색포탈의 시장구조를 도출하였다. 제 6 장에서는 본 연구결과를 요약하고 연구결과의 적용방안과 한계점을 알아보고, 향후 연구과제를 제시하였다.

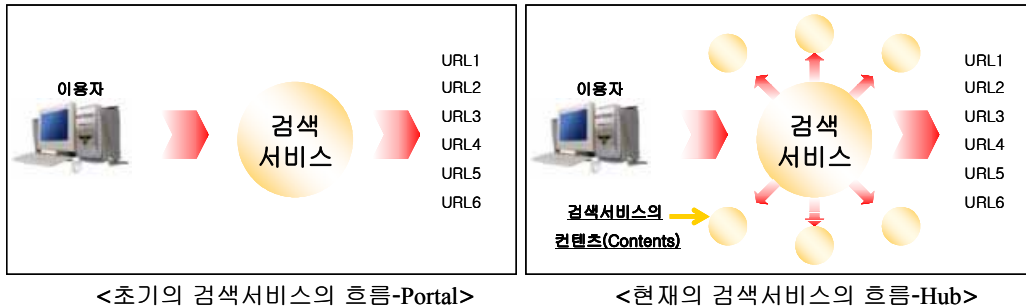
제 2 장 검색포탈의 정의

포탈(Portal)이란 인터넷이용자들이 인터넷을 이용할 때 관문역할을 하는 사이트를 일컫는 말로 사용되어왔다. 이러한 개념에서 출발하였지만 최근 포탈의 모습을 보면 인터넷과 관련된 검색기능, 전자메일 기능, 뉴스 제공 기능, 쇼핑 기능 등 매우 다양한 기능을 제공하는 초대형 사이트로 정의되고 있는 추세이다.

포탈이 기본적으로 관문의 역할을 하기 위해서는 일단 그 관문을 통과하는 사람이 많아야 포탈의 미디어적인 성격을 이용한 비즈니스모델의 실현이 가능하다. 많은 사이트 이용자를 모으기 위한 방법으로는 초창기처럼 검색포탈이나 접속서비스를 제공한 유형이 있는가 하면 기타 이메일이나 다른 서비스로 방문자를 모은 경우도 있을 수 있다. Yahoo는 디렉토리 서비스에서, AOL이나 국내의 천리안은 접속서비스로부터, Daum은 한메일을 이용하여 일단 많은 이용자를 모으는데 성공하여 현재 포탈로 불리고 있다.

좀더 거슬러 올라가서 초기 검색포탈의 등장에 대해서 알아보면 다음과 같다. 인터넷이 보편화되고 네트워크를 통해 흩어져있는 수많은 정보에 접근하고 공유하는 것이 가능하게 되자, 자연스럽게 필요에 맞는 정보를 효율적으로 검색하는 것이 주요 이슈가 되었다. 웹 브라우저가 대중화되고, 하루가 다르게 수많은 웹사이트들이 생겨나자, 이들을 소개하고 찾기 쉽게 정리한 서비스가 폭발적인 인기를 얻게 된 것이 바로 검색업체가 주요 포탈로 등장하게 된 배경이다. 이 중 Yahoo는 디렉토리 검색서비스로 포탈 시장을 선점하였고 잇따라 Altavista, Lycos 등의 갖가지 다양한 검색서비스가 뒤를 이었다.

<그림 1>의 초기 검색포탈의 흐름은 단순히 이용자들에게 검색질의 결과를 보여주는 서비스에서 현재 검색포탈이 추구하는 Hub-포탈 또는 Mega-포탈로 불리는 자체 콘텐츠를 이용자에게 제공하는 구조로 변화됨을 보여준다.



<그림 1> 검색포탈의 비즈니스모델변화

본 연구에서는 다양한 포탈의 분류 중에서 검색서비스를 통한 많은 이용자 트래픽(Traffic)을 올리고 있는 검색포탈을 그 대상으로 하고 있다. <표 1>은 국내 인터넷관련 조사기관인 코리안클릭(Korean Click)의 2002년 상반기 Top 50개의 포탈 중에서 검색포탈의 순위를 보여주며 Daum과 Yahoo가 1위와 2위를 차지하고 있다. Daum은 E-mail서비스를 기반으로 포탈시장에서 가장 선두를 달리고 있으며, Yahoo는 디렉토리 기반 검색포탈로 검색포탈 중 가장 많은 이용자를 보이고 있다.

Daum을 검색포탈이라고 분류하기는 어렵지만 전문 검색엔진회사인 잉크토미(Inktomi)⁹⁾와의 제휴로 검색포탈 역시 다른 검색포탈과 비교해 볼 때 손색이 없는 것으로 평가 된다. 국내 대형 포탈들 역시 검색기능을 제공하고 있지만 여기서는 그 포탈의 제공서비스 중 검색포탈의 성격이 약하기 때문에 제외시켰다.

⁹⁾ 잉크토미(<http://www.inktomi.com>)

<표 1> 국내 포털사이트에서의 검색포털의 순위와 이용자수

상반기 순위	사이트	카테고리	상반기 월평균 방문자수(천명)
1위	Daum	커뮤니티/검색 포털	22,562
2위	Yahoo	검색 포털	20,514
3위	Naver	검색 포털	19,542
5위	Hanmir	검색 포털	14,526
6위	Lycos	검색 포털	14,169
11위	Empas	검색 포털	11,003
25위	Simmani	검색 포털	6,415

* 출처 : 코리아 클릭(2002)

여기서 검색엔진부문에서 최고의 결과와 많은 이용자를 보이는 Google은 검색포털이라고 분류하기 보다는 검색전문사이트로 여기는 것이 적절하겠다. 왜냐하면 검색포털을 제공하지만 검색포털의 비즈니스모델인 이용자들에게 충성도 확보를 위해 부가서비스 등을 제공하지 않고 있기 때문이다.

제 3 장 이론적 고찰

3.1. 인터넷 검색포털에 관한 선행연구

인터넷 검색포털에 대한 연구는 1996년 이후 꾸준히 계속되어 오고 있으며 각기 다양한 키워드를 사용하여 연구가 진행되어 왔다. Westera(1996)는 ‘Wine’에 관한 연구를 하였고, Feldman(1997)은 ‘Car, Tennis, Elbow’ 등의 다양한 질의를 통해서 연구를 하였다. 대부분의 검색포털 관련 연구들은 검색어를 통해서 수집된 자료를 바탕으로 진행되어 왔다. 그리고 연구에 사용된 검색어의 선택은 관련 서

적, 전문가의 의견 및 통계자료 등을 바탕으로 선택하였다. 이와 같은 한정된 키워드로 검색포탈을 전반적으로 평가한다는 데에는 한계가 있으나, 검색어를 통해서 검색포탈의 대략적인 특성을 파악하는 데는 효과적이라 할 수 있다. 이렇게 다양한 검색포탈에 동일한 키워드를 입력하여 나온 데이터를 분석하는 연구는 꾸준히 진행되어 왔다.

Lake(1997)는 일상생활에서 흔히 볼 수 있는 실용적인 검색어를 통해서 검색된 결과물에 대한 적합성 및 정확성을 평가하였다. 이를 통해 보다 신뢰성 있는 검색포탈에 관한 연구가 진행되었다. 그러나 각 검색포탈에 대한 정의가 불분명하며, 키워드 및 디렉토리의 명확한 분류와 수집된 자료에 대한 정확한 통계적 검정에 문제점을 가지고 있었다. Gordon and Pathak(1999)은 8개의 검색포탈과 33개의 검색어를 통하여 이러한 선행연구의 단점을 보완하였다.

지금까지의 검색포탈에 관한 연구는 주로 검색엔진의 기술적인 측면에서 다루지고 있었다. 그러나 최근에는 마케팅적 접근이 시도되고 있다. Bradlow and Schmittlein(2000)는 20가지 마케팅용어와 6가지 검색포탈의 중복 URL자료를 바탕으로 수평적으로 검색포탈시장의 구조를 분석하였다. 수평적 시장구조분석을 통해서 검색포탈을 이용하는 고객이 검색결과 찾고자 하는 정보를 찾지 못했을 경우 다른 검색포탈을 효율적으로 선택하게 하는 검색포탈 이용순위를 제시하였다. 이를 통해서 이용자의 효율적인 검색뿐만 아니라 서비스제공자는 자사와 유사한 검색포탈이 어느 것인지를 알게 될 것이며, 차별화 전략에 도움을 줄 수 있게 된다. 검색포탈시장 전체 차원에서의 연구인 Gandal(2001)의 연구에서는 검색포탈시장이 이용자의 전환비용이 거의 들지 않기 때문에 진입장벽이 낮으며, 이러한 진입장벽이 낮음에도 불구하고 Yahoo가 시장에서의 우위를 점하고 있는 것은 지속적인 혁신을 통한 우월한 서비스의 제공에 기인한다고 하였다. 또한 낮은 진입장벽은 검색포탈시장을 복잡한 경쟁상황으로 몰아넣게 될 것이며, 이를 위해

서 서비스제공자는 그들이 가진 서비스의 차별화를 시도해야 한다고 주장했다.

3.2. 시장구조분석에 관한 선행연구

연구모형의 설명에 앞서 기존연구들의 시장에 대한 정의를 본다면, 시장은 경우에 따라 다양한 해석이 가능하겠지만 고객과 상품제공자의 욕구충족을 위한 장이라고 할 수 있다. 1970년대 후반 이후 활발하게 연구되어온 시장구조분석이라고 할 때의 시장은 고객이 추구하는 어떤 혜택을 충족할 수 있는 서로 대체성이 높은 상품집합을 지칭하는 상품시장의 개념이다. 이처럼 시장구조분석은 시장의 개념에 차이가 있고 상품시장에 관한 연구와 고객시장관점의 연구들이 각각 독립된 관점으로 연구가 되었다.

시장구조분석은 상호관계가 명백한 일련의 제품을 묶기 위한 과정이라 할 수 있다(Grover and Dillon 1985). 시장구조를 분석하는 것은 기업의 마케팅 목적을 위해 소비자 행동의 여러 가지 상이한 측면을 조직화하기 위한 것으로서 이러한 조직화는 소비자행동의 다양한 측면으로부터 일종의 체제나 질서를 제공하게 된다.

경쟁적 시장구조에 대한 정확한 이해는 기업의 마케팅 관리자들에게 제품전략을 평가할 때 중요하다. 즉 마케팅 관리자는 신제품 출하, 제품 재구성, 포지셔닝 결정에 있어 경쟁자가 무엇에 의해 가장 많은 영향을 받고, 또한 무엇에 대해 가장 민감하게 반응할 것인지를 알려는 욕구가 있다. 이러한 욕구는 경쟁적 시장구조를 분석함으로써 충족될 수 있다. 마케팅 관리자의 마케팅전략에 의해 어떤 제품이 가장 민감하게 반응하는지를 경쟁적 시장구조를 통해 알 수 있다면 경쟁적 시장구조가 유용한 개념이 될 수 있는 것이다(Urban, Johnson and Hauser 1984). 시장구조분석을 통해 상표들을 동질적으로 집단화시키면 집단 내에서의 경쟁 정도와 집단 사이의 경쟁의 정도를 파악할 수 있다. 즉, 상표들 사이의 구조관계를

이해함으로써 시장세분화 결과 나타난 목표시장에서 경쟁군을 이해하면 기존상표를 재정위화 시키고 새로운 상표들에 대한 소비자반응을 평가하여 신제품개발과 상표확장 기회를 찾을 수 있게 된다.

수직적 시장구조분석의 대표적인 연구 중의 하나인 Urban, Johnson and Hauser(1984)의 시장구조분석 모형에서 적용한 총체적 수준의 상표전환 자료를 이용한 연구들의 기본전제는 상표간의 전환이 발생하는 정도를 상표간의 유사성 정도로 간주하여서 제품시장을 유사한 상표들의 묶음인 하부시장들로 계층적으로 분할하여 상표간 대체성 내지 경쟁 정도를 나타내고자 하는 것이다. 상표전환은 조건부확률 즉, 지난번 구매에서 상표 A를 구매한 경우 이번 구매에 상표 B를 구매할 확률로 측정되는 것이 일반적이다. 대개 이러한 조건부확률은 설문 또는 패널자료로 계산되는데 상표전환연구에서는 연속적인 구매를 상정하고 있으며 일정기간 동안에는 구매행위나 사용행위가 유사하다고 가정하고 있다. 또 소비자들은 대체성이 낮은 제품들보다는 대체성이 높은 제품들간에 상표전환행동을 하기가 쉬우며, 상표전환비율은 대체성 측정의 척도로 제공된다고 가정하고 있다. 이 가정은 상표전환을 이용한 대체성 연구의 기본 논리를 제공한다고 하겠다(Rao and Sabavala 1981).

본 연구에서는 과거의 시장구조분석의 커피, 자동차, 음료시장의 분석과는 달리 인터넷검색포탈이라는 새로운 시장을 다루고 있다. 또한 상표전환자료가 아닌 검색포탈간의 중복(Overlap)자료를 이용하고 있으므로 연구모형의 적용이 다소 생소하게 느껴질 수 있다. 그러나 이용자들의 상표전환 행동자료 대신 앞서 서술한 기존의 검색포탈에 관련된 연구를 통해 검색포탈간의 중복되는 URL에서 중복이 일어나고 있다는 것을 알 수 있었고, 이러한 중복되는 자료를 바탕으로 서비스간의 유사성이 또는 대체성이라는 개념을 이끌어 낼 수 있다. 만약 2가지 이상의 검색포탈을 이용하고 있는 사용자가 첫 번째 선택한 검색포탈에서 자신

이 원하는 정보(URL)를 찾지 못하거나 이용 중이던 검색포탈이 운영이 중단된다면 다른 검색포탈을 이용하게 될 것이라는 점이 기존의 상품 시장구조분석 연구의 상표전환과 유사한 의미를 가질 수 있기 때문이다. 정리하자면 본 연구는 검색포탈 간의 전환을 검색포탈의 검색결과물의 중복이라는 개념에서 출발했으며, Urban, Johnson and Hauser(1984)의 강제상표전환을 이용한 시장구조분석에서 사용된 확률적인 모형을 기반으로 인터넷 검색포탈의 시장구조를 확인하고자 한다.

제 4 장 연구모형

4.1. 검색포탈의 시장구조분석모형

각각의 검색포탈은 인터넷이라는 하나의 공간에서 정보를 수집하게 됨으로써, 검색포탈 간의 중복되는 자료를 가질 수밖에 없다. 중복 URL을 이용해서 특정 URL이 각 검색포탈에 존재하는지의 여부를 통해 상품의 대체성과 같은 개념인 URL의 전환율을 얻을 수 있었다. 이렇듯 각 검색포탈에서 제공하는 검색결과들은 다양한 중복관계가 존재하고 있으며, 이들을 총합적인 수준에서 본다면 결국 검색포탈간의 유사성을 알 수 있게 될 것이다.

본 연구에서는 URL 전체 차원(aggregate level)의 관점에서 볼 때, 구조화가 전혀 안된 시장이라면 ACRM(aggregate constant ratio model, 이하 ACRM으로 칭함)이 적용되어야 한다는 Urban, Johnson and Hauser(1984)의 시장구조에 관한 연구에서처럼 검색포탈시장 역시 구조화가 되어 있지 않다면 ACRM이 적용된다고 전제하고 이로부터 벗어난 정도를 단측 Z검정(one tail Z-test)을 실시했다. 검색포탈 시장에서 ACRM이 적용된다는 것은 어떤 검색포탈이 없어지거나 이용 불가능하다고 할 때, 나머지 각 검색포탈에서 발견되거나 이들을 이용하는 정도가 그 검색포탈의 기존 점유율에 비례함을 의미한다.

본 연구에서 점유율은 두 개 이상의 검색포탈에서 존재하는(검색포탈간 중복되는) URL의 수를 통해서 계산된다. 기존 시장구조분석에서는 상품전환이 일어나는 고객의 수를 통해서 얻어진다는 점에서 본 연구와의 차이점이 존재한다. 그러나 본 연구에서는 URL의 집합인 검색결과가 각 검색포탈마다 다름으로 인해 검색포탈의 전환을 발생시킬 수 있는 가능성이 있다는 점에서는 기존의 시장구조분석과 동일한 관점이나, 소비자 행동의 관점에서의 전환이 아니라 발견되는 URL이라는 제품속성차원에서의 유사성을 이용한다는 점이 다르다고 할 수 있다.

ACRM의 조건하에서는 아래의 식 (1) 이 성립한다. 즉, 하부시장 A 에 j 가 존재할 확률은 하부시장전체의 점유율과 j 의 점유율을 통해서 얻을 수 있다.

$$(1) \quad P(j | A) = \frac{m_j}{\sum_{k \in A} m_k}$$

만약 하부시장 A 가 전체시장 T 일 경우, $\sum_{k \in T} m_k = 1$ 이 된다.

수식 (1)을 통해서 식 (2)와 (3)이 성립된다.

$$(2) \quad P_i(j) = \frac{m_j}{\sum_{k \neq i} m_k} = \frac{m_j}{1 - m_i}$$

$$(3) \quad P_i(s) = \sum_{j \in s, j \neq i} P_i(j) = \frac{\sum_{j \in s, j \neq i} m_j}{1 - m_i}$$

$P_i(j)$ = 검색포탈 i 가 없을 경우 j 의 점유율, $i \neq j$, i 의 URL이 j 에 존재함.

s = 검색포탈의 집합, 하부시장 s .

$P_i(s)$ = 검색포탈 i 가 없을 경우 전체 시장에서 하부시장 s 의 점유율.

수식 (2)와 (3)은 검색포탈의 시장구조가 없을 경우를 전제하여 성립된다. 여기서 점유율 m_i 는 뒤에 나올 식 (15)에서 구할 수 있다.

다음은 앞서 설명한 ACRM과 비교하기 위한 URL의 중복여부와 관련된 부분을 정의한다.¹⁰⁾

n_i = 검색포탈 i 의 URL중에서 적어도 1개 이상의 다른 검색포탈에 중복해서 존재하는 URL수.

$n_i(j)$ = n_i 중 i 가 없어질 경우 검색포탈 j 에서 발견되는 URL 수.

$n_i(s)$ = n_i 중 i 가 없어질 경우 하부시장 s 의 URL 수.

만약 검색포탈시장이 ACRM이 적용된다면 검색 결과물에 의한 예측치는 다음과 같이 될 것이다.

$$(4) \quad n_i(j)/n_i \approx P_i(j)$$

$$(5) \quad n_i(s)/n_i \approx P_i(s)$$

다음은 앞서 설명한 ACRM의 예측치와 비교하기 위한 검색포탈의 검색결과물을 통한 방법을 알아본다. 중복이 되는 URL중 검색포탈 i 가 없어질 경우 하부시장 s 에서의 존재할 확률은 식 (6)으로 구해진다.

$$(6) \quad \hat{P}_i(s) = n_i(s)/n_i$$

¹⁰⁾ 다음 $n_i, n_i(j), n_i(s)$ 의 URL의 수는 뒤에서 설명하게 될 식 (10)을 통해 얻어지게 된다.

이를 통해 우리가 정의한 시장구조모델이 구조가 없는 경우와 비교했을 때 보다 어떤 차이가 나타난다면 각각의 하부시장 s 에서 다음과 같이 예상할 수 있다.

$$(7) \quad \hat{P}_i(s) \geq P_i(s) \text{ (검색포탈 } i \text{가 하부시장 } s \text{에 속하는 경우)}$$

$$(8) \quad \hat{P}_i(s) \leq P_i(s) \text{ (검색포탈 } i \text{가 하부시장 } s \text{에 속하지 않는 경우)}$$

여기서 수식 (7) 과 (8)의 결정은 본 연구에서 설정하게 될 검색포탈 시장구조모형의 하부시장이 어떻게 이루어져 있는지에 의해 각기 다른 결과가 나타날 것이다. ACRM과 상정된 시장구조와 비교검정을 위해 다음과 같이 가설을 설정한다.

$$\left\{ \begin{array}{l} H_0 : \hat{P}_i(s) \leq P_i(s) : \text{시장구조 없음. 검색포탈 } i \text{가 하부시장 } s \text{에 속하지 않음.} \\ H_1 : \hat{P}_i(s) \geq P_i(s) : \text{시장구조가 있음. 검색포탈 } i \text{가 하부시장 } s \text{에 속함.} \end{array} \right.$$

만약 검색포탈 i 에 속한 URL중 무작위 추출한 경우, 하부시장 s 에서 발견될 확률 $P_i(s)$ 가 독립적이라면 이항분포로서 평균 $n_i P_i(s)$, 분산 $n_i P_i(s)(1 - P_i(s))$ 로 만약 n_i 가 상당히 클 경우, 중심극한정리에 의해서 $n_i(s)$ 는 (9)와 같이 표현 가능하다. 즉, ACRM이 적용되는 시장구조가 없는 H_0 의 가정 하에 수식 (3)의 $P_i(s)$ 와 수식 (6)의 $\hat{P}_i(s)$ 를 단측 Z-test로 검정한다.

$$(9) \quad n_i(s) \sim \text{Normal}[n_i P_i(s), n_i P_i(s)(1 - P_i(s))]$$

4.2. 예제 모형

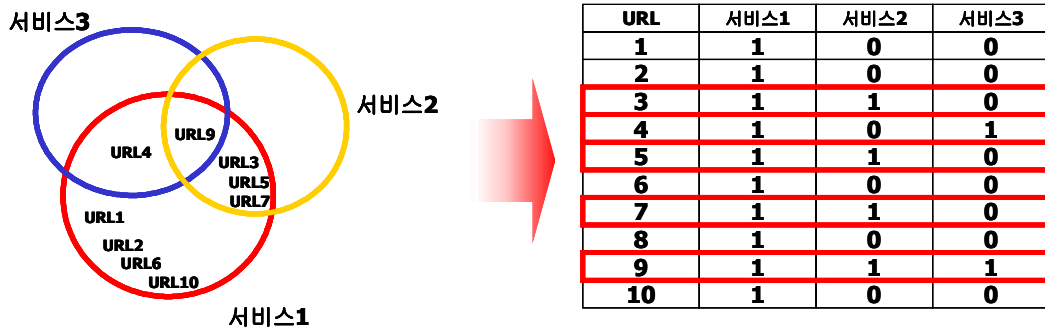
예를 통해서 본 연구의 분석방법을 자세히 설명하면 다음과 같다. 질의어에 대한 URL이 10개가 있고 검색포탈이 보유하고 있는 URL이 다음 <표 2>와 같이 구성되어 있을 때, 우선 두 가지 이상의 검색포탈에서 중복이 있을 경우만을 선택한다. 중복되는 URL이 있는 경우에만 서비스간의 전환이라는 개념이 적용될 수 있기 때문이다.

<표 2> 검색포탈서비스의 URL 소속의 예

URL	포탈 서비스1	포탈 서비스2	포탈 서비스3	URL	포탈 서비스1	포탈 서비스2	포탈 서비스3
1	1	0	0	6	1	0	0
2	1	0	0	7	1	1	0
3	1	1	0	8	1	0	0
4	1	0	1	9	1	1	1
5	1	1	0	10	1	0	0

주: 0=해당 검색포탈에서 발견되지 않음. 1=해당 검색포탈에서 발견됨.

<표 2>는 3가지 검색포탈에서 10개의 URL을 예를 들어 나타냈다. URL 1번과 2번은 서비스 1에만 존재하고, URL 9번과 같은 경우는 3개의 검색포탈에 모두 있다는 것을 알 수 있다.



<그림 2> 검색포탈서비스간 중복관계의 예

<그림 2>는 검색포탈 간의 URL중복을 표로 변환하는 과정을 보다 쉽게 보여준다. 그림에서 URL 3,4,5,7,9가 2개 이상의 검색포탈에 존재하는 URL인 것을 알 수 있다. 서비스간의 전환율을 구하기 위해서 이 중에서 5개의 URL을 선택하여 이용하게 된다. 그리고 이렇게 수집된 URL각각이 해당 검색포탈에 존재할 확률은 식 (10)과 같다. 기존의 시장구조분석에서 이용되는 자료는 상표선택을 위한 소비자들의 행동 및 인식과 관련된 전환 자료인 반면에 본 연구에서는 URL 관점에서 이들이 검색포탈에서 중복되어 발견되는 비율을 이용한 검색포탈간의 유사성을 검정한다.

$$(10) \quad P'_u(i) = \frac{u_{ij}}{\sum_j u_{ij}}, (i \neq j)$$

$P'_u(i)$ = 특정 URL u 가 검색포탈 i 에 존재할 확률.

u_{ij} = i 의 URL u 가 j 검색포탈에서의 존재여부(1=존재, 0=존재하지 않음)

<표 3> 검색포탈의 개별 URL의 존재확률의 예

URL	포탈 서비스1	포탈 서비스2	포탈 3 서비스
3	0.5	0.5	0
4	0.5	0	0.5
5	0.5	0.5	0
7	0.5	0.5	0
9	0.33	0.33	0.33

만약 URL u 가 검색포탈 i 에 존재하지 않을 경우 $P'_u(i)$ 는 0이 된다. URL 3의 경우 $P'_3(1) = 0.5$, $P'_3(2) = 0.5$, $P'_3(3) = 0$ 이 된다. 이 과정은 특정 검색포탈이 강제적으로 제거될 경우 URL별로 존재확률의 변화를 위해 필요하다.

그 다음으로 특정 URL u 가 검색포탈 i 가 제거되었을 경우 검색포탈 j 에 있을 확률은

$$(11) \quad P_{ui}(j) = \frac{P'_u(j)}{1 - P'_u(i)}$$

로 구해지며,

$P_{31}(2) = 1.0(1.0 = 0.5/(1-0.5))$, $P_{41}(3) = 1.0$, $P_{51}(2) = 1.0$, $P_{71}(2) = 1.0$, $P_{91}(2) = 0.5$, $P_{91}(3) = 0.5$ 와 같이 계산할 수 있다.

이를 통해 n_i 중 서비스 i 가 없어졌을 경우 j 검색포탈에 속하는 URL수 $n_i(j)$ 를 다음과 같이 구할 수 있게 된다.

$$(12) \quad n_i(j) = \sum_u P_{ui}(j)P'_u(i)$$

$$(13) \quad n_i = \sum_u P'_u(i)$$

$$(14) \quad \hat{P}_i(j) = n_i(j)/n_i$$

식 (13)의 n_i 는 특정 URL u 가 검색포탈 i 에 존재할 확률의 합으로 구할 수 있으며 이를 통해 식 (14)를 구할 수 있게 된다. 예제의 경우를 계산해 보면,

$$n_1(2) = P_{31}(2)P'_3(1) + P_{41}(2)P'_4(1) + \dots + P_{91}(2)P'_9(1) = 0.5 + 0.5 + 0.5 + 0.165 = 1.665$$

,

$$n_1 = P'_3(1) + P'_4(1) + P'_5(1) + P'_7(1) + P'_9(1) = 0.5 + 0.5 + 0.5 + 0.33 = 2.33$$

와 같이 구할 수 있다. 수식 (6)을 통해 검색포탈 1이 없을 경우 검색포탈 2에서 존재할 확률은

$$\hat{P}_1(2) = \frac{n_1(2)}{n_1} = 0.714 \text{ 임을 알 수 있다.}$$

또한, ACRM모형을 위한 검색포탈 i 의 점유율은,

$$(15) \quad m_j = \frac{n_i}{\sum_i n_i}$$

로 구할 수 있다.

<표 4> 검색포탈서비스 간 전환율의 예($\hat{P}_i(j)$)

$i \backslash j$	n_i	포탈서비스 1	포탈서비스 2	포탈서비스 3
포탈서비스 1	2.33	-	0.714	0.286
포탈서비스 2	1.83	0.910	-	0.090
포탈서비스 3	0.83	0.801	0.199	-

<표 4>는 예제 자료를 이용한 서비스간 전환율을 구한 표이다. 이를 보면 검색포탈 1이 없어졌을 경우 검색포탈 2로의 전환율이 0.714로 나타나는 것을 알 수 있다. 이는 서비스 1이 제거되었을 때 서비스 3보다 서비스 2로 더 많이 전환됨을 나타낸다. 이렇게 얻어진 표를 다음에 설명할 절차대로 ACRM기반의 구조가 없는 경우와의 비교검정을 실시한다.

4.3. 모형의 확장

본 연구에서는 앞서 설명한 예와 같이 분석을 실시하였으나 현실적인 문제를 고려하기 위해 약간의 모형 추가가 불가피 했다. 그 이유는 다음과 같다.

국내외 인터넷 조사기관과 검색포탈관계자의 보고에 의하면 검색포탈 이용자의 80% 이상이 검색결과순위 3페이지 이상은 보지 않는다고 한다. 이는 30개의 결과까지의 URL이 그 이후의 순위에 있는 결과물 보다 중요도가 높다는 점을 말해 준다.

앞서 설명한 예에서는 첫 번째 URL과 천 번째 또는 그보다 더 뒤에 나오는 URL이 동일하게 다뤄지고 있으므로 현실적인 문제를 간과하고 있다. 이러한 문제는 다음에 설명하게 될 발견확률의 가중치의 조정을 통해서 해결할 수 있다.

앞서 언급한 가중치의 조정을 위해서 본 연구에서는 위험함수(hazard function)를 이용하였다. 위험함수는 확률 분포함수보다 제약이 덜하며, 이용이 편리하다는 장점이 있다. 적용 예로는 지체시간(duration) 또는 고장시간(failure time)이나 생존(survivor)자료의 분석에 이용되는 기법으로 많은 이론 및 실증연구가 있다 (Heckman and Singer, 1984; Jain and Vilcassim, 1991).

본 연구에서는 앞서 언급했던 검색포탈을 이용하는 행태를 음의 지수함수(negative exponential function)형태로 정의하고 이를 다음과 같이 전개하도록 한다.

검색포탈간 중복되는 URL의 랭크(r)에 대한 가중치는 식 (16)과 같은 위험함수로 나타낸다.

$$(16) \quad F(r) = e^{-\lambda r}, (1 \leq r \leq 200)$$

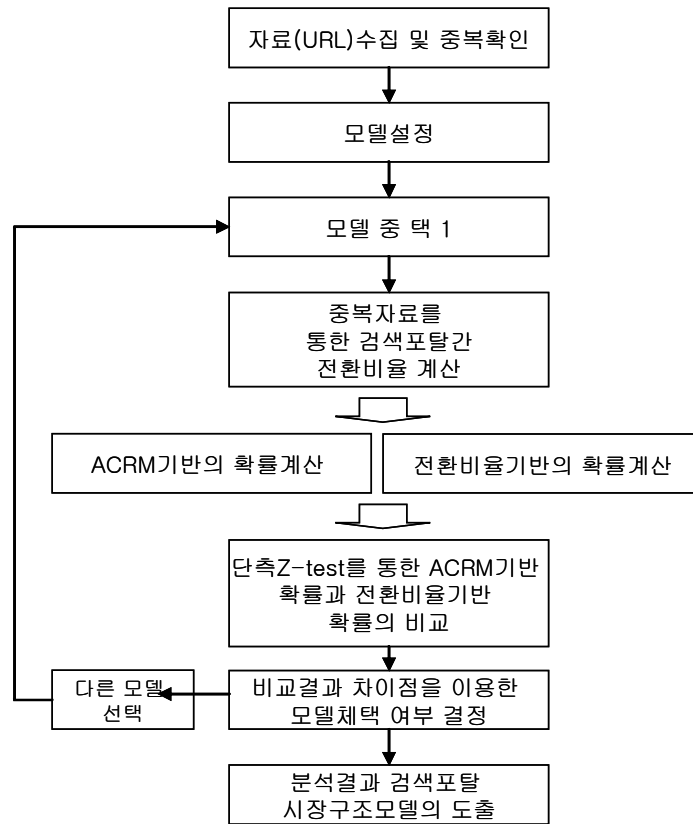
여기서 λ 는 앞서 검색포탈을 이용하는 이용자들이 30페이지까지 이용하는 비율이 30%내외이므로 본 연구에서는 그 적정값으로 0.05를 선택하였다 ($(F(31) \approx 0.2)$). 이를 앞의 식 (10)의 u_{ij} 에 적용하면 다음과 같다.

$$(17) \quad u'_{ij} = u_{ij}F(r)$$

식 (17)의 u'_{ij} 는 기존의 모형에서 단순히 0과 1의 값만 가지는 u_{ij} 와는 달리 위험함수를 이용해서 랭킹에 따른 값을 가지게 된다. 이렇게 얻어진 식 (17)을 앞서의 식 (10)에 대입을 통해서 확장모형에 관한 연구를 진행해 본다. 이때 주의할 것은 전환이 일어나는 검색포탈의 개수가 2개 이상일 경우에만 적용 가능하다는 것이다. 만약 어떤 URL이 단지 두 개의 검색포탈에만 존재할 경우에는 그 발견확률은 기존모델의 u_{ij} 에서와 같이 1의 값을 가지게 된다. 이러한 검색된 URL별 가중치 적용을 통해서 보다 현실성 있는 모형이 구축될 것으로 예측하며, 이후 두 모형을 비교해보도록 한다.

4.4. 분석절차

<그림 3>은 본 연구의 분석절차를 흐름도로 요약한 것으로 우선 가설을 설정한 후 검정하는 탐색적인 방법을 취하고 있다. 첫 번째 단계로 검색포탈시장의 구조를 한 단계의 하부시장의 수준으로 구성된 다양한 모델로 상정한다. 그런 다음 각각의 선택된 검색포탈과 키워드를 이용해서 URL등의 결과물 자료를 수집한다. 두 번째 단계로 하나의 모형을 선택 (H_a) 하고, 이때의 ACRM기반의 확률값 ($P_i(s), P(s)$) 을 구한다. 그와 동시에 수집된 자료에 기반해서 검색포탈의 중복관계를 고려한 서비스 전환율을 바탕으로 H_a 의 확률값 ($\hat{P}_i(s), \hat{P}(s)$) 을 구한다. 세 번째 단계로 앞서 얻은 ACRM기반의 경우와 검색포탈 전환율을 이용한 경우를 단측 Z-test로 가설검정을 실시한다. 이때 유의수준($\alpha = 0.10, Z \geq 1.28$)을 만족시키지 못한다면 다시 모델선택과정으로 돌아가 다른 모형을 선택해서 분석절차를 따른다. 마지막 단계로 유의수준을 만족하는 시장구조 중 총합적으로 가장 적합한 모형을 선택하게 된다.



<그림 3> 분석절차

제 5 장 실증분석

연구목적 달성을 위하여 기존의 검색포탈관련 연구를 바탕으로 다음과 같은 단계로 분석하고자 한다. 우선, 분석을 위한 데이터베이스의 구축을 위한 첫 번째 단계로 검색포탈에 입력할 키워드를 선정한다. 본 연구에서는 2002년 8월에 각 검색사이트에서 발표한 네티즌 인기 검색어 중에서 비교적 유행에 민감하지 않은 15개의 키워드를 선택하였다. 두 번째 단계는 실험 대상인 검색 사이트의

선정이다. 국내 및 인터넷조사기관의 이용자순위자료 등을 통하여 가장 선호하는 8종류의 검색포탈인 Daum, Yahoo, Empas, Altavista, Hanmir, Lycos, Naver, Simmani를 선택하였다. 세 번째 단계로는 앞서 선정된 검색 사이트에 키워드를 검색해서 나온 결과 URL을 기준으로 데이터베이스를 구축하였다. 끝으로 데이터 분석을 통하여 검색포탈의 시장구조분석을 실시하였다. 자료수집은 2002년 8월 중순에 보름간 이루어졌으며 총 310,991개의 결과물 중 23,990개를 샘플로 선정하였다.

5.1. 키워드(Key-Word)와 검색포탈의 선정

검색포탈을 이용하는 이용자들의 키워드를 종합해서 각 검색포탈들은 공시하고 있다. 이렇게 하는 이유는 이용자들이 필요로 하는 정보를 알기 위해서이다. 이런 공시된 인기검색어의 컨텐츠화를 통해서 현재 어려움을 겪고 있는 유료화의 새로운 방향을 모색할 수 있다. 또한 네티즌의 평균 연령이 점차 인구분포에 가까워지고 있기 때문에 인기검색어의 중요성은 보다 높아진다고 할 수 있겠다.

연구에 필요한 URL 데이터베이스 구축을 위해 사용될 키워드를 선정해야 했다. 국내 검색포탈업체 중 Naver와 Lycos에서는 이용자들이 선호하는 검색용어를 정기적으로 공시하고 있으며, 본 연구에서는 2002년 8월에 공시된 네티즌들이 선호하는 키워드를 사용하기로 하였다. 인기검색어 중 다음과 같은 사항의 용어는 제외하기로 했다.

- √ 지나치게 유행에 민감한 검색어
- √ 연예인과 같은 특정인을 찾는 검색어
- √ 특정 사이트나 검색결과 URL수가 극히 작은 경우

이러한 선별 작업을 거쳐서 총 15개의 키워드(검색엔진, 논문, 만화, 부동산,

쇼핑몰, 영화, 운세, 유머, 지도, 채팅, MP3, 게임, 취업, 자료실, 신문)를 사용하였다.

검색포탈의 선택은 2장에서 조사된 국내 8대 검색포탈인 Daum, Yahoo, Empas, Altavista, Hanmir, Lycos, Naver, Simmani를 선택하였다.

5.2. 자료 수집 및 데이터베이스구축

분석을 위해 키워드로 질의 후 검색된 URL결과물을 MS-ACCESS프로그램을 이용해서 데이터베이스화 하였다. 또한 SQL질의를 통해서 검색포탈간의 중복관계를 URL별로 정리하였다. 구축된 데이터베이스의 구조는 <표 5>와 같이 출력순위(Rank), URL, 검색키워드(Key), 검색에 이용된 포탈(Se), 중복체크의 필드로 구성되어 있다.

<표 5> 데이터베이스 구조의 예(Yahoo)

rank	URL	key	se	A	H	L	N	S	Y	E	D
1	http://kr.yahoo.com/	1	6	0	0	0	1	1	0	0	0
2	http://www.chome.co.kr/	1	6	0	1	1	1	0	0	1	1
3	http://www.koreansearchengines.com/	1	6	0	1	0	1	1	0	0	1
4	http://www.wisengine.com/	1	6	0	0	0	1	0	0	1	0

우선 첫 번째 열 Rank는 검색포탈업체에서 자체적인 순위부여기준에 의해 나타난 순서로 보통 출력화면당 10개의 검색결과가 나타난다. 두 번째는 URL데이터이며, 세 번째 열 Key는 수집에 사용된 키워드를 나타내며 1번 ‘검색엔진’부터 15번 ‘신문’을 나타낸다. 네 번째 열은 검색에 사용된 포탈의 번호를 나타내며 1번 Altavista에서 8번 Daum을 나타낸다. 5~12번째의 열은 검색포탈 간의 검색결과를 통합한 이후 중복되는 여부를 ‘0’ 과 ‘1’ 로 나타내었다. 예를 들어 Rank

1번의 경우 Yahoo에서 수집된 검색엔진이라는 키워드를 통해 얻은 자료이며 (0,0,0,1,1,0,0,0)로 표현되며 Naver, Simmani에서 관찰되었다는 것을 알 수 있다. 다시 말해서 8개의 검색포탈 중 Yahoo, Naver, Simmani가 Rank1의 URL을 가지고 있음을 뜻한다.

<표 6> 수집된 전체 데이터의 수

	Altavista	Hanmir	Lycos	Naver	Simmani	Yahoo	Empas	Daum	합계
검색엔진	1959	1696	1056	539	1357	402	414	190	7613
논문	1539	2381	1360	921	1456	1473	716	927	10773
만화	1674	2658	6972	1142	3328	1541	1187	919	19421
부동산	2628	5108	3798	2590	4123	1201	998	768	21214
쇼핑몰	5144	12882	8680	7030	7026	1796	1048	3687	47293
영화	4524	5274	6932	2301	7078	3261	2664	2015	34049
운세	407	834	1086	471	565	290	209	223	4085
유머	1691	2077	3234	418	3232	746	569	441	12408
지도	1348	2772	1330	1020	1375	1036	839	506	10226
채팅	1487	1722	1167	717	1171	490	495	233	7482
mp3	2202	2341	3836	613	4423	536	690	334	14975
게임	5031	7139	14007	2569	10811	2914	2636	2279	47386
취업	1806	2755	3491	1155	1702	759	569	925	13162
자료실	8499	7750	9033	4596	9275	1635	1850	6383	49021
신문	1927	2518	1662	1094	1876	931	947	928	11883
합계	41866	59907	67644	27176	58798	19011	15831	20758	310991

수집된 전체데이터는 <표 6>과 같이 총 310,991개의 결과물이 검색되었으며 Lycos의 결과물 수가 가장 많은 것을 알 수 있다. 이 중에서 보다 효율적인 연구를 위해 각 검색포탈별/키워드별 상위 200개의 결과물을 이용하기로 했다.

이렇게 구축된 <표 5>의 데이터베이스의 자료를 4장의 분석절차를 통해 <표 7>과 같이 서비스 전환 표를 만들었다. 여기서 n_i 는 시장점유율과 같이 사용되는 2가지 이상의 검색포탈에서 중복되는 URL수를 나타내며 각 수치는 전환율을 나타낸다. 표에서 보면 Altavista가 없을 경우 타 검색포탈과의 전환은 그리 차이가 없으나 반대의 경우에는 전환율이 매우 작게 나타난다는 것을 알 수 있다. 이

는 최근 Altavista가 검색포탈 중에서도 가장 작은 이용자를 보유하는 것과 연관 지어 생각해 보면 다른 검색포탈과 너무 동떨어진 결과를 보여주는 것 보다는 다른 포탈과 좀더 유사한 결과를 얻을 수 있도록 노력하는 방안을 생각해 볼 수 있다.

<표 7> 전체 검색포탈의 전환율

$i \backslash j$	n_i	Alta vista	Hanmir	Lycos	Naver	Sim mani	Yahoo	Empas	Daum
Altavista	721	0	0.18	0.20	0.12	0.10	0.17	0.13	0.10
Hanmir	1666	0.06	0	0.14	0.17	0.11	0.16	0.17	0.19
Lycos	1146	0.10	0.21	0	0.15	0.13	0.12	0.15	0.14
Naver	1361	0.05	0.21	0.11	0	0.13	0.15	0.18	0.16
Simmani	919	0.06	0.18	0.16	0.18	0	0.14	0.16	0.13
Yahoo	1371	0.07	0.19	0.1	0.16	0.11	0	0.16	0.22
Empas	1278	0.06	0.21	0.12	0.19	0.11	0.15	0	0.17
Daum	1283	0.04	0.22	0.11	0.15	0.11	0.21	0.16	0

다음으로 검색포탈시장의 여러 가지 다양한 구조를 상정해 보았다. 본 연구에서는 한 단계 레벨의 하부시장을 가지는 구조만을 고려하였지만 그 경우의 수가 너무 많아서 일일이 대입해서 구할 수는 없었다. 그래서 8개의 검색포탈을 2개(28가지), 3개(56가지), 4개(70가지), 5개(56가지), 6개(28가지)검색포탈이 하나의 하부구조로 설정하여 각각의 경우를 가설검정을 통해 유의한 구조만을 우선적인 대상으로 하는 휴리스틱(heuristic)한 방법을 사용했다.

여기서 만약 검색포탈의 이용자가 연구에서 사용된 8가지의 검색포탈을 모두 사용하고 있고 각각의 검색결과를 비교하게 된다면 이용자들은 8가지 검색포탈의 관계를 정확히는 파악할 수 없지만 비슷한 결과를 보여주는 포탈들을 인지

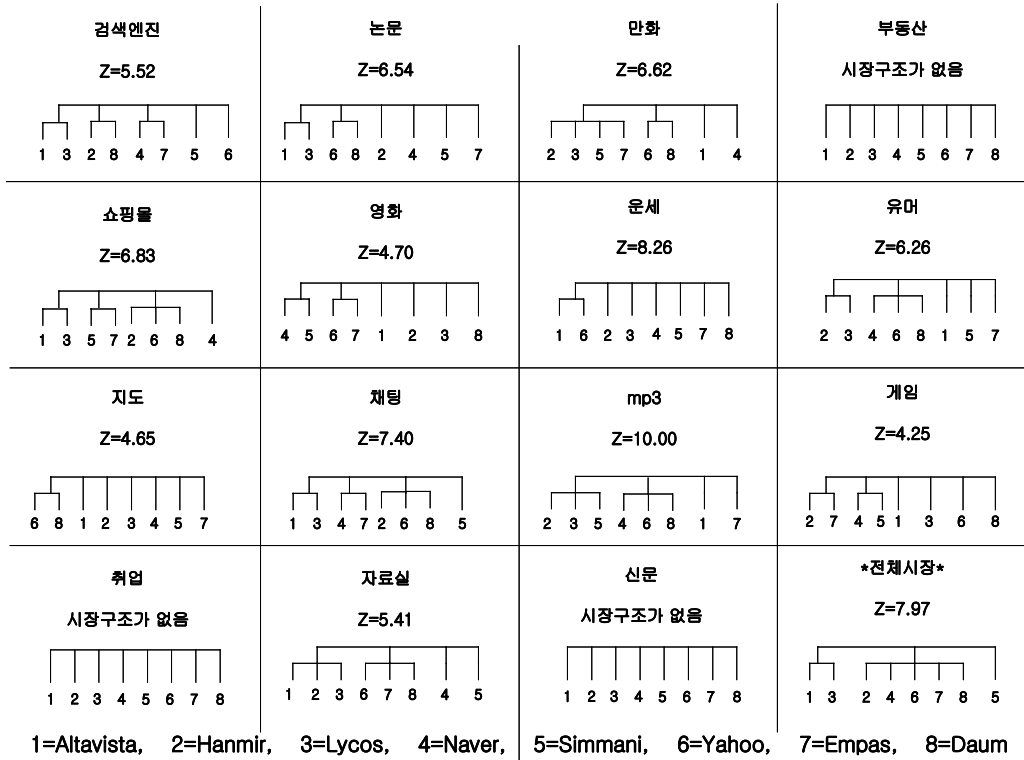
할 수 있게 될 것이다. 그러나 이용자들은 그들의 개인성향과 검색의도 등에 의해 상이한 반응을 나타낼 수 있다. 기존에 이용하던 검색포탈이 사용불가 상태일 때 어떤 이용자들은 비슷한 결과물을 보여주는 검색포탈을 선택할 것이다. 혹은 그와 반대의 현상도 일어날 것이다. 전자의 경우는 하부시장에서의 경쟁이 예상되며, 후자의 경우는 보완적이라고 해석할 수 있다. 이러한 이유로 같은 하부시장으로 묶이는 검색포탈간의 경쟁관계를 정확히 설명할 수는 없다. 그러나 하부시장간의 차이점은 분명히 존재하며, 이들간의 경쟁 혹은 보완관계에 대해서는 앞으로 연구되어야 할 것이다.

분석을 위해 선정된 8개의 검색포탈은 각각 어떠한 경쟁/보완적인 관계를 가지고 있다는 것을 알 수 있었다. 이러한 분석 결과를 통해 검색포탈 관리자들과 그들과 직접적으로 경쟁/보완 관계가 예상되는 서비스업체를 알아낼 수 있으며, 경쟁업체와 보다 차별화 된 서비스의 제공을 위한 마케팅전략의 수립이 가능해지게 된다. 이를 통해 경쟁적 우위도 얻을 수 있을 것이다.

최근의 검색포탈 업체들은 단순히 검색포탈만을 제공하는 것은 아니다. 컨텐츠와 커뮤니티를 구축해서 이용자들을 보다 오랫동안 머물게 하려고 노력 중이며, 이를 통해서 유료 컨텐츠의 이용증가를 위해 노력 중이다. 검색포탈 제공자는 이용자들이 원하는 바를 검색 질의어를 통해 손쉽게 얻을 수 있으며 자사의 컨텐츠로 변형/가공을 통해 수익모델을 만들고 있다. 이용자들의 인기검색어를 통해 서비스제공자는 신규 컨텐츠의 제공을 모색할 수 있으며, 서비스 이용자들은 어렵게 정보를 얻으려고 인터넷을 향해하는 노력을 줄일 수 있게 될 것이다.

그렇다면 각각의 키워드 별로 보았을 때의 시장구조는 어떻게 될 것인가? 각각의 키워드 별로 시장구조를 알게 된다면 새로운 컨텐츠의 도입 시 예상되는 경쟁자를 알 수 있게 되어 서비스 제공자는 타사에 비해 경쟁적 이점을 얻을 수 있을 것이다. 키워드 별 분석방법은 앞서 기술한 분석기법과 동일하게 적용되었

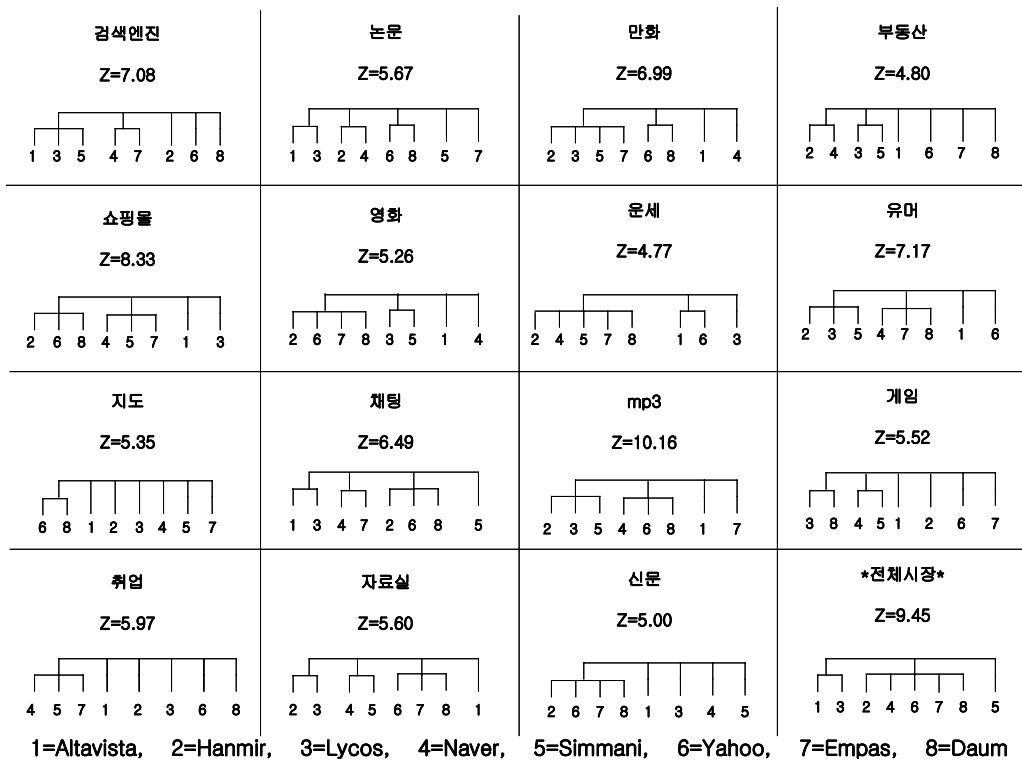
으며, 단지 전체 URL이 아닌 각각의 키워드 별로 알아보았는데 그 차이가 있다.



<그림 4> 검색포탈서비스의 시장구조모형 (키워드, 전체, $\lambda = 0$)

<그림 4>는 4장의 모형확장부분에서 언급한 랭킹을 고려하지 않았을 경우 ($\lambda = 0$)의 시장구조를 보여준다. ‘부동산’, ‘취업’, ‘신문’과 같은 키워드에서는 시장구조가 없는 것으로 나타났으며 나머지 부분에서는 각기 다양한 구조들이 나타났다.

아래 <그림 5>는 확장된 모형인 $\lambda = 0.05$ 일 경우의 시장구조를 보여준다. 앞서 나온 시장구조모형과 비교해 볼 때, 시장구조에 대한 설명력이 높을 뿐만 아니라 앞의 모형에서 찾지 못한 키워드에 대한 시장구조가 나타났다. 이를 통해 확장된 모형의 사용이 보다 더 효율적인 것으로 판단된다.



<그림 5> 검색포탈서비스의 시장구조모형 (키워드, 전체, $\lambda = 0.05$)

<그림 5>와 같이 각각의 키워드 별, 전체시장의 구조를 보면 키워드-‘검색엔진’의 경우 Altavista, Lycos, Simmani가 같은 하부시장으로 묶여 있으며, Naver와 Empas가 또 다른 하부시장으로 묶이는 것을 알 수 있다. 키워드-‘만화’의 경우 Hanmir, Lycos, Simmani, Empas가 하나의 하부시장을 이루고 있으며 Yahoo와 Daum이 같은 하부시장으로 묶인다. 이를 해석해 보면 Yahoo는 같은 하부구조의 Daum과는 이용자에게 보여주는 정보가 다른 검색포탈에 비해서 그 유사성이 매우 높다는 것을 알 수 있다. 만약 Yahoo의 관리자가 만화에 관한 차별화된 콘텐츠를 새롭게 제공하려고 한다면 같은 하부구조의 Daum의 만화 콘텐츠제공업체가 아닌 다른 하부구조의 Hanmir, Lycos, Simmani등과의 제휴 또는 그들의 콘텐츠

제공업체와의 계약 등을 생각해 볼 수 있다. 그러나 이러한 결과는 각 검색포털의 만화에 대한 콘텐츠의 질적인 측면이 고려되어야 할 것이다. 다시 말해 차별화 전략을 위한 검색포털별 이용자들의 평가나 정성적인 자료 등을 토대로 선택해야 할 것이다. 현재 거의 모든 검색포털 업체들은 만화, 운세, 영화 등의 콘텐츠를 유료로 제공하고 있으며 자사에서 콘텐츠를 제작해서 제공하는 경우도 있지만 대부분이 제휴업체의 콘텐츠를 제공 받고 있다. 이렇듯 각각의 검색포털들은 검색엔진의 수집, 색인방법의 변경이나 계약중인 콘텐츠 제공업체를 변경함으로써 시장에서의 차별화 전략을 가능하게 할 수 있을 것이다.

다음으로 검색포털의 검색된 결과물 중 상위에 랭크 시키기 위해 유료로 등록한 업체의 경우를 살펴보자. <표 8>은 각 검색포털이 최근 새롭게 제공하는 유료랭크의 판매현황을 보여준다. 각 검색포털은 각각의 키워드 별로 이용자의 Hit 수를 공시하고 있고 이를 기반으로 가격을 책정하고 있다. 표에서 *표는 현재 판매된 키워드로 모든 검색포털에서 키워드 ‘채팅’이 판매되었음을 알 수 있고, 키워드 ‘쇼핑몰’ 또한 많은 광고주들을 확보하고 있다. 표에서 빠져있는 Yahoo와 Altavista 역시 이러한 프리미엄 키워드를 통해 수익을 얻고 있다.

<표 8> 검색포탈의 키워드광고의 판매가격

(단위 : 만원)

	Hanmir	Lycos	Naver	Simmani	Empas	Daum
검색엔진	40	50	75	30	40	100
논문	15	50	58	19	40	60
만화	100	150	75	74	84	300
부동산	200	150	*150	74	177	*100
쇼핑몰	200	70	150	*41	*108	*100
영화	200	150	*150	74	202	300
운세	91	100	*150	41	70	300
유머	70	70	75	30	79	100
지도	200	150	75	41	216	100
채팅	*183	*150	*225	*41	*109	*400
mp3	182	150	75	74	204	100
게임	200	150	15	107	236	400
취업	200	150	*15	*41	*139	200
자료실	125	150	75	74	77	100
신문	60	30	75	30	66	100

*수집시기 : 2002년 10월 10일 20시

*네이버(50% d/c이벤트중)

*가격앞의 *표시는 현재 판매된 키워드

각 검색포탈에 첫 페이지에 랭크 되기하기를 원하는 광고주들은 단순히 각 검색포탈의 이용자수와 가격을 기반으로 광고대상 검색포탈을 선택한다. 만약 본 연구의 결과인 검색포탈시장의 구조를 파악할 수 있다면 각 하부시장 별로 랭크 시킬 수 있다. 이렇게 함으로써 얻을 수 있는 이점은 서로 다른 하부시장에 랭크 시키게 된다면 자신의 광고 아래로 무수히 많은 대체되는 사이트들이 존재한다는 점에서 매우 중요하다. 더욱이 이용자들은 아무 거리낌없이 자신의 사이트와는 경쟁관계의 사이트들을 단 한번의 클릭으로 옮겨가게 되기 때문이다.

이상의 결과로 각각의 키워드 별 시장구조와 전체 시장구조가 상이한 것을 알 수 있었다. 이러한 결과는 검색포탈 이용자와 검색포탈업체 그리고 광고주들에게 검색포탈의 브랜드만으로 서비스전체를 평가하는 데는 문제가 있다는 것을 말해준다. 우선 검색포탈에서 검색을 하는 이용자들은 키워드 검색 시에 다양한 검색포탈을 이용하게 된다면 보다 폭넓은 자료를 수집할 수 있게 될 것이며, 자

료수집이 미흡할 때 다음으로 선택할 검색포탈을 다른 하부시장에서 선택할 수 있게 될 것이다.

다음으로 검색포탈 제공업체 들은 자사와 하부시장이 다른 검색포탈과의 정보 교환이나 컨텐츠의 교환을 통해 자사의 경쟁력확보에 노력할 수 있게 될 것이다.

끝으로 광고주들은 키워드광고를 실시할 검색포탈을 선택할 때 비용과 관측 효과 면에서 보다 나은 결과를 얻을 수 있을 것이다.

제 6 장 결론 및 미래 연구 방향

지금까지는 검색포탈 제공업체의 관점에서 본 연구의 실제 적용방안을 제시했지만 검색포탈의 이용자의 관점에서 이러한 연구는 다음과 같은 중요한 의미를 가진다.

첫째, 검색포탈업체의 수가 증가함에 따라 이용자들은 점차 다양한 검색포탈을 사용해 보게 될 것이다. 새로운 검색포탈을 이용한 이후 이용자들의 반응은 두 가지로 예상된다. 우선 이용자들이 같은 하부시장으로 분류된 검색포탈을 선택하는 경우와 다른 결과물을 보여주는 검색포탈을 선택하는 경우이다. 두 경우는 기존에 이용하고 있던 검색포탈의 만족여부와 이용자들의 성향에 따라 좌우될 것이다. 본 연구에서는 이 두 경우를 다루기에 앞서 검색포탈시장이 그들이 제공하고 있는 URL을 통해서 세부시장으로 분류가 되고 있음을 보여주고 있다. 이렇게 분류된 시장구조를 보여줌으로써 각 검색포탈 관리자들은 검색포탈시장에서의 위치를 알 수 있게 될 것이다. 따라서 기존의 검색기술개발에 다양한 마케팅전략 등의 수립이 더해질 수 있게 될 것이다.

둘째, 상이한 검색포탈의 결과물을 단지 통합된 형식으로만 보여주는 메타검

색포탈의 경우 위의 결과를 통해 검색포탈들 간의 관계를 고려한 좀 더 효율적인 메타검색사이트를 구축할 수 있게 될 것이다. 동일한 하부시장으로 분류된 검색결과물 보다는 서로 다른 하부시장의 검색포탈들을 선택하게 된다면 보다 다양한 정보의 수집 및 저장되는 데이터의 양이 감소될 것이다.

셋째, 검색포탈을 이용하는 일반 이용자들은 이러한 시장구조를 알게 된다면 다음의 이점을 얻을 수 있겠다. 기존에 이용하던 검색포탈이 이용불가 상태이거나 더 이상 서비스를 제공하지 않게 될 경우 막연히 다른 검색포탈을 선택하게 될 것이다. 이 때 검색포탈의 시장 구조를 알고 있다면 기존에 이용하던 검색포탈과 유사하거나 상이한 검색포탈의 선택에 도움이 될 수 있겠다.

넷째, 현재 Yahoo, Lycos 등은 검색엔진 전문회사인 구글(<http://google.com>)의 검색컨텐츠를 제공 받고 있다. 초기에는 검색포탈을 제공하고 있던 이러한 업체들이 다양한 자체적인 혹은 제휴를 통한 컨텐츠 제공에 주력하고 있다. 이러한 이유는 유료화를 통한 수익창출과 보다 많은 이용자 확보를 위해서이다. 그러나 검색포탈 자신이 제공하는 URL 정보가 자신의 컨텐츠를 제공하는데 걸림돌이 되고 있는 실정이다. 예를 들어 만화라는 유료 컨텐츠를 제공하려고 한다면 이용자들은 컨텐츠의 이용에 앞서 무료만화를 검색포탈을 통해서 찾으려 노력하기 때문이다. 이렇듯 검색포탈 제공자들은 컨텐츠의 경쟁에 앞서 그들 자신이 제공하고 있는 URL에 대한 정확한 정보를 알고 있어야 한다. 이러한 정보를 제공하는데 있어서 본 연구의 또 다른 의미를 찾을 수 있겠다. 또한 광고주들은 인터넷 이용자의 막대한 트래픽을 보다 효율적인 광고전략을 수립할 수 있게 될 것이다.

본 연구는 인터넷의 정보를 체계적으로 수집해서 이용자들에게 정보를 제공하고 있는 검색포탈들의 관계에 대한 이해를 제공함으로써, 검색포탈제공자와 이용자 그리고 광고주들은 많은 이익을 얻게 될 것이다. 서비스 제공자들은 그들의 경쟁업체를 알아냄으로써 새로운 컨텐츠의 도입 시에 예상되는 경쟁자들을 알아

낼 수 있고, 기존의 포탈간의 경쟁관계의 이해를 통해 보다 차별화 된 서비스를 제공할 수 있게 된다. 포탈이용자 역시 좀 더 효율적인 정보의 탐색이 가능하게 된다. 또한 본 연구에서는 인터넷이라는 새로운 연구과제를 기존에 연구되었던 방법을 통해서 풀어나갈 수 있었다. 인터넷을 위한 많은 새로운 연구가 필요하겠지만 기존 연구들의 응용이 가능하다고 말할 수 있겠다.

그러나 본 연구는 다음과 같은 몇 가지 한계점을 지니고 있다. 우선 연구대상 URL을 단지 수량으로 파악했을 뿐 그들이 가지는 정보의 질적인 측면을 고려하지 않았다. 이러한 문제점은 기존 검색포탈관련 연구에서도 거론되어 왔던 내용이다. 이런 문제점은 URL이 연결되어 있는 실제 사이트에 관한 내용의 객관적 평가가 이루어진다면 이러한 문제를 해결 할 수 있게 될 것이다. 그러나 이런 객관적 평가는 인터넷에 존재하는 웹 페이지수가 10억 개 이상이라고 추정만 되고 있는 실정이므로 상당히 어려운 부분이라고 하겠다.

다음으로 검색포탈시장의 구조를 단지 제공되는 정보(URL)의 집합이라는 개념으로 파악하였고, 실제 검색포탈을 이용하는 이용자들의 선택이나 행동들을 고려하지 않았다. 이런 문제점은 이용자들의 로그파일(Log File)을 분석하게 된다면 의외로 쉽게 해결될 것으로 생각되지만 실제 데이터의 수집이 어렵다는 문제가 있다. 대부분의 인터넷 비즈니스 업체들은 자사의 로그파일의 유출을 꺼리고 있으며, 이는 이용자의 정보유출을 막기 위해서와 경쟁사에게로 넘어가는 것을 막기 위해서다.

끝으로 본 연구에서는 2002년 8월의 자료를 바탕으로 하고 있다. 인터넷은 하루가 다르게 정보가 쏟아져 나오고 있다. 분석이 끝날 무렵의 상황과 초기 연구를 진행할 당시와는 검색포탈과 기타 상황들이 변해 있으므로, OLAP(Online Analytical Processing)와 같이 실시간으로 자료의 수집과 분석이 가능하게 된다면 이러한 문제점을 보완할 수 있으리라 생각된다.

앞서 서술한 몇몇 부분들을 조금씩 보완해 나간다면 향후 인터넷비즈니스 중 가장 많은 이용자를 확보할 수 있는 검색포탈시장에 대한 보다 체계적인 연구가 이루어질 것으로 기대한다.

참 고 문 헌

- 코리아 클릭(2002), “2002년 상반기 TOP 50 site,” [http://www.koreanclick.co.kr].
- 한상만, 김소영(1997), “판단적 경쟁구조분석과 행동적 경쟁구조분석의 비교연구: 고려상표군을 중심으로,” *마케팅연구*, 12(2), 67-93.
- Bradlow, Eric T. and David C. Schmittlein(2000), “The Little Engines That Could: Modeling the Performance of World Wide Web Search Engine,” *Marketing Science*, 19(1), 43-62.
- Feldman, Susan(1997), “Just the Answers, Please: Choosing a Web Search Service,” [http://www.infotoday.com/searcher/may/story3.htm].
- Gandal, Neil(2001), “The Dynamic of Competition in the Internet Search Engine Market,” *International Journal of Industrial Organization*, 19, 1103-17.
- Gordon, M. and Praveen Pathak(1999), “Finding Information on the World Wide Web: The Retrieval Effectiveness of Search Engines,” *Information Processing and Management*, 35(2), 141-80.
- Grover, Rajiv and William R. Dillon(1985), “A Probabilistic Model for Testing Hypothesized Hierarchical Market Structures,” *Marketing Science*, 4(4), 312-35.
- GVU(1999), *GVU's User Survey*, Georgia Institute of Technology: Graphic, Visualization, and Usability Center, [http://www.gvu.getech.edu/user_surve

y].

Heckman, J. and B. Singer(1984), "A Model for Minimizing the Impact of Distributional Assumptions in Econometric Model for Duration Data," *Econometrica*, 52, 271-320.

Jain D.C. and N.J. Vilcassim(1991), "Investigating Household Purchase Timing Decisions: A Conditional Hazard Function Approach," *Marketing Science*, 10, 1-23

Kannan, P. K. and Gordon P. Wright(1991), "On Testing Competitive Market Structures," *Marketing Science*, 10(4), 338-47.

Kumar, Ajith and C. M. Sashi(1989), "Confirmatory Analysis of Aggregate Hierarchical Market Structure: Inference From Brand Switching Behavior," *Journal of Marketing Research*, 26(November), 444-53.

Lake, Matthew J.(1997), "2nd Annual Search Engine Shoot-out," [<http://www4.zdnet.com/pccomp/features/exc10997/sear/sear.html>].

Leighton, H. Vernon and Jaideep Srivastava(1997), "Precision among World Wide Web Search Services (Search Engines): AltaVista, Excite, HotBot, Infoseek, Lycos," [<http://www.winona.msus.edu/is-f/library-f/webind2/webind2.htm>].

Rao, Vithala R. and D. Sabavala(1981), "Inferences of Hierarchical Choice Process from Panel Data," *Journal of Consumer Research*, 8(June), 85-96.

Urban, Glen L., Philip L. Johnson, and John R. Hauser(1984), "Testing

Competitive Market Structure,” *Marketing Science*, 3(Spring), 83-112.

Westera, Gillian(1996), “Robot-Driven Search Engine Evaluation Overview,”
[<http://www.curtin.edu.au/curtin/library/staffpages/gwpersonal/senginestudy/index.htm>].

감사의 글

사랑하는 부모님께 이 논문을 바칩니다.