

# 인터넷 이용자의 검색 행동 성향에 관한 연구

## A Study of User's Characteristics of Searching Behavior on the World-Wide Web

오 경 목(Kyung-Mook Oh) \*  
황 상 규(Sang-Kyu Hwang) \*\*  
이 용 현(Yong-Hyun Rhi) \*\*

### 목 차

- |                    |                 |
|--------------------|-----------------|
| 1. 서 론             | 3. 1 평가 시스템     |
| 2. 선행연구            | 3. 2 평가 성향      |
| 2. 1 인터넷 이용자 성향 조사 | 4. 인터넷 이용자 성향분석 |
| 2. 2 로그파일 분석       | 5. 결 론          |
| 3. 인터넷 이용자 평가성향    |                 |

### 초 록

새로운 정보검색 환경인 인터넷에서의 정보검색은 과거의 정보검색행위와는 여러 면에서 다르다. 인터넷 이용자들은 과거의 정보 이용자와는 다른 행동 양식을 보이고 있기 때문에 보다 나은 서비스를 제공하기 위해서 인터넷 검색 시스템은 새로운 이용자들의 특성을 제대로 반영하여야 한다. 정보 수요자인 인터넷 이용자들의 검색 행동 성향을 조사, 분석하는 과정을 통해 이용자들의 요구사항을 파악하고, 이를 통하여 인터넷 기반 정보검색시스템 설계를 위한 가이드라인을 제시하고자 한다.

### ABSTRACTS

Information retrieval on World Wide Web is very different from the one in traditional environment. Since Internet user's searching behavior shows it's own special search characteristics, a study of user's specific characteristics is necessary to offer better services. Through analysis and examination of Internet user behaviors, we suggest guidelines of design for developing Internet-based information retrieval systems.

영문키워드: WWW, Internet Search, searching behavior, information seeking

국문키워드: 웹, 인터넷검색, 검색행위, 정보탐색

\* 숙명여자대학교 정보과학부(Division of Information Science, Sookmyung Women's Univ)

\*\* 홍익대학교 전자계산학과(Department of Computer Science, Hong-Ik Univ)

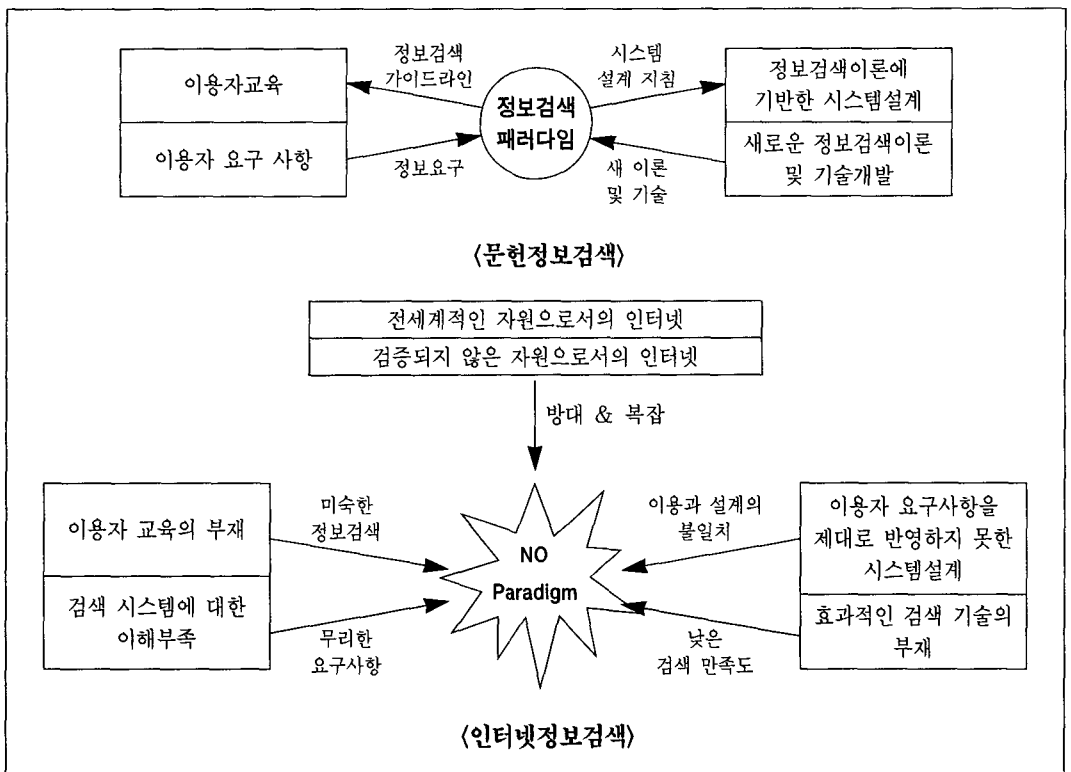
접수일자 1999년 8월 20일

### 1. 서론

소수 전문가들의 소유물이었던 인터넷은 현재 국경을 초월하여 전세계적으로 주요 정보 자원으로 자리잡게 되었다. 요즘 컴퓨터를 사용해본 사람라면 대부분 최소한 한번 이상은 인터넷을 사용한 경험이 있으며, 이와 더불어 인터넷에서 정보를 찾아주는 수많은 인터넷 정보검색시스템들의 존재를 알고 있다. 하지만 대부분의 인터넷 정보검색시스템들은 이용자들의 요구사항을 제대로 반영하지 못하고 있으며, 이용자들 역시 그들이 이용하는 인터넷 정보검색시스템들에 대해 잘 알지 못하고 있는 실정이다(Pollock and Hockley

1997). 서로가 상대방에 대해 잘 알지 못하고 있는 현 상황에서 원활한 정보검색이 이루어 지기는 실로 불가능하며, 이는 이용자나 정보 검색시스템 설계자 모두에게 큰불만으로 작용 하게 된다. 이러한 상황이 발생하게 된 배경에는 여러 가지 이유가 있을 수 있으나 달라진 정보검색 환경 <그림 1>에서 그 근본원인을 찾을 수 있다.

기존의 정보검색에 주류를 이루는 문헌 정보검색에서는 오랜 기간 동안 기존 학계와 업계가 쌓아온 연구를 통해 잘 정돈된 정보검색 패러다임이 존재한다. 이러한 정보검색 패러다임은 이용자에게 효과적인 정보검색방법을 제시해줄 수 있으며, 시스템 설계자에게는 효



<그림 1> 달라진 정보검색 환경

올직한 시스템 설계가 이루어지기 위한 지침을 제시해줄 수 있었다. 무엇보다도 이용자교육을 통해 정보검색의 방법을 가르쳐줄 수 있었으며, 교육의 과정 및 활용과정에서 수집된 이용자들의 요구사항들은 시스템 설계에 반영되어 질 수 있었다. 이에 반하여 새로운 정보검색 환경인 인터넷 정보검색에서는 유감스럽게도 아직까지 잘 정돈된 정보검색 패러다임이 존재하지 않는다. 지금 이 순간에도 끊임없이 변화하고 확대되고있는 인터넷에서는 계속 새로운 기술들이 도입되고 있으며, 검증되지 않은 여러 가지 기술들이 실제 서비스되어지는 과정에서 실험되어 살아남거나 도태되어진다.

새로운 정보검색 환경인 인터넷에서 사용되어진 기술들이 충분한 기간동안 연구를 통해 잘 정돈되어 하나의 정보검색 패러다임을 이룰 충분한 시간적 여유가 없는 현 상황에서 대부분 시스템 설계자들은 과거의 정보검색 패러다임을 기본으로 하여 새로운 인터넷 정보검색 시스템을 설계하게 되는데 이러한 과정에서 실제 이용자들의 요구사항은 충분히 반영되지 못하게 된다. 비록 효과적인 정보검색 시스템이 만들어졌다 하더라도 대부분의 이용자들이 단 한번의 정보검색교육의 기회조차 받지 못한 초보자라는 현 상황에서는 새로 개발된 많은 기능을 갖춘 시스템이 제대로 활용되어질 수 없다. 따라서 시스템 설계자에게는 새로운 기술개발 못지 않게 실제 이용자들의 요구사항을 제대로 파악하여 이를 새로운 인터넷 정보검색 시스템 설계에 반영할 수 있는 노력과 자세가 요구된다.

본 연구에서는 선행된 연구들의 연구결과와

문헌조사를 통해 새로운 정보검색환경인 인터넷의 특징에 대해 조사하였으며, 새로운 정보검색환경인 인터넷에서 이용자들의 특성을 조사, 분석하는 과정을 통해 이용자들의 성향을 연구하였다. 지능형 정보에이전트와 검색엔진 등을 통해 웹정보 검색 이용자들의 평가성향을 비교, 분석하였는데 동물과 관련된 질의어 21개에 대해 4개의 선정된 웹정보 검색시스템 으로부터 최대 100건의 문서를 평가하였다. 또한 연구 과정에서 밝혀진 사실들을 토대로 새로운 인터넷 정보검색 시스템 설계를 위한 가이드라인을 제시하고 인터넷 정보검색 패러다임 확립에 도움이 될 수 있는 내용들을 연구, 분석하였다.

## 2. 선행연구

### 2.1 인터넷 이용자 성향 조사

인터넷이 널리 쓰이기 시작한 이후 선행 연구자들은 새로운 네트워크 환경인 인터넷과 인터넷을 이용하는 이용자들이 가지는 새로운 특성에 주목하였으며, 이들의 특징을 파악하기 위한 다양한 연구 및 조사를 시행하였다. 주요 인터넷 관련 기업에서도 활발한 연구가 진행되어져 왔으며, 문헌정보학(Beavers 1998; 이응봉 1999), 전산학(Henzinger et al. 1999; Josef, Alfred and Andreas 1997), 심리학, 인지과학(박창호, 염성숙, 이정모 1999)등 각기 다른 분야를 전공한 여러 연구자들이 인터넷 정보검색에 관한 연구를 진행해오고 있다.

인터넷은 그 자체의 끊임없는 변화와 복잡함에 의하여 인터넷 고유의 특성을 파악해내기가 수월치 않으며 인터넷의 특성을 파악한 효과적인 검색 방법을 찾아내기 힘들다. 하지만 인터넷을 이용하는 이용자들은 구성원의 다양성에도 불구하고 상당부분 공통된 특성을 파악해 낼 수 있었다. 인터넷 이용자들은 그들이 가진 지식의 정도, 지역, 성별, 학력 등 다양한 요인에도 불구하고 공통된 성향이 존재하며, 심지어는 인터넷을 이용하는 초보자와 고급이용자간에도 공통된 성향이 존재하였다. 인터넷 이용자들의 가장 큰 특징으로는 검색 방식의 단순성을 이야기할 수 있다. 인터넷 이용자들은 대부분이 자신이 가진 사전지식을 잘 활용하지도 않으며 검색연산자도 사용하지 않고 적은 수의 검색어로 구성된 단순한 질의문을 통해 정보검색을 수행한다(Abdulla, Liu and Fox 1998). 박창호와 박민규의 연구결과(1998) 역시 비슷한 결론에 도달하였는데 “정보검색 초보자의 전략은 비교적 단순하며 이는 정보검색 숙련자 역시 별반 큰 차이를 보이지 않는다”고 한다. 정보사냥 대회에 참가한 숙련자의 검색행동의 분석 결과, 검색어의 수, 연산자의 종류와 사용빈도, 검색어 유형 등에서 초보자의 결과와 크게 다르지 않았으며, 질의문 작성시 초보자와 숙련자 모두(최초의) 한두 검색어를 중심으로 하여 다른 검색어들을 교체하곤 하였다. 이러한 특징 때문에 정보검색행동 관찰을 통해 상당부분 이용자의 성향 및 검색 의도를 파악하는 방법이 사용되어질 수 있다. 이러한 방식의 연구로는 Microsoft사에서 주도되어진 인터넷 이용자 성향 파악 및 이용자 친화적인 시스템 설계에

관한 여러 연구(Microsoft 1996)들을 찾아볼 수 있으며, 그밖에 여러 다양한 연구들이 진행되어오고 있다.

## 2. 2 로그파일 분석

일반 이용자들의 성향을 파악하기 위해 인문사회과학 연구자들이 주로 사용하는 방법에는 다음 세 가지 방식(Bains 1997)이 있다. 먼저 인터뷰나 설문지를 통한 직접적인 조사가 있으며, 실험 참가자들을 대상으로 장기간 관찰하는 방법이 있다. 마지막으로 검색 시스템을 이용한 이용자들의 이용 기록을 분석하는 로그파일(log file)분석방법이 있는데 로그파일 기록은 이용자 성향을 조사하는 객관적인 자료로 이용되어 진다. 도서관의 OPAC 시스템을 대상으로 이용자 검색에 관한 연구(Ballard 1994)가 로그파일 분석방법을 통해 진행되었으며, Jansen et al.(1998)은 새로운 정보검색 환경인 인터넷을 대상으로 51,000개의 질의를 대상으로 조사, 분석하였다. 보다 대규모적인 실험으로는 Silverstein et al.(1998)에 의해 진행된 검색엔진 Altavista의 로그파일분석이 있다. 그는 6주간 대략 2억 8500만명의 이용자들이 남긴 280GB의 로그파일을 분석하였는데, 장기간에 걸친 대규모 실험이라는 점에서 신뢰할 수 있는 여러 유용한 정보를 얻을 수 있었다.

여러 방법들 가운데 가장 객관적인 분석 방법인 로그 파일분석을 통한 이용자분석에서는 검색 결과에 대한 이용자 만족도에 대해서는 아무런 기록정보를 얻을 수 없는 관계로 이에 관한 분석이 불가능하며, 직접조사나 실험 관

찰을 통한 별도의 연구가 필요하다. 관련 연구로서 예제를 통해 검색한 결과를 비교한 검색엔진 비교에 관한 연구(Chu and Rosenthal 1996; 정영미, 김성은 1997)나 이용자 검색 만족도를 측정하는 방법에 관한 연구(Schlichting and Nilsen 1998) 등 이미 다양한 연구가 진행된 바 있다. 하지만 대부분의 연구가 이용자 검색 평가성향조사를 통해 얻은 정보를 조사, 분석하는 이용자분석의 측면보다는 이미 운영되고 있는 검색 시스템의 성능평가에 초점을 맞추어 연구가 진행되어져 왔다. 따라서 본 연구에서는 네 가지 서로 다른 검색시스템을 비교하는데 있어 이용자 검색 평가성향조사에 초점을 맞추어 연구를 진행하였으며, 평가결과 분석 및 실험참가자와 인터뷰를 통해 인터넷 이용자 성향중 일부인 이용자 평가성향을 파악하는데 중점을 두어 연구를 진행하였다.

### 3. 인터넷 이용자 평가성향

동물에 대한 전문 지식베이스를 갖춘 지능형 정보 에이전트 HIIA-1a(이용현 1999)와 검색엔진 Altavista, Lycos, Excite들간의 검색성능 비교데이터들을 보다 광범위한 관점에서 분석하는 작업을 통해 웹 정보검색 이용자들의 평가성향을 분석해 보았다. 평가 인원은 홍익대 전자과 대학원생 19명을 대상으로 실시하였으며, 모든 응답자들은 기존에 웹 정보검색에 대해 교육을 받은 적이 없으며, 웹 정보검색 수행 능력은 대부분 중급이하의 일반 이용자들이로 구성하였다. 평가자는 동물에 관

한 질의 21개(부록 참조)에 대해서 각각 4개의 웹 정보검색 시스템으로부터 최대 100개까지 문서의 적합성여부를 평가하였다. 평가 작업수행 중 발생하게 되는 돌발상황이나 의문점에 대해서는 이메일을 통해 관리자에게 수시로 문의토록 하였고, 최종적인 결과 수집 및 분석에 있어 발생하는 의문사항에 대해서도 이메일을 통해 각 평가자들과 의견교환을 하였으며 이를 통해 보다 정확한 평가가 이루어지도록 하였다.

#### 3.1 평가 시스템

다음은 본 연구에서 평가를 위해 사용된 네 개의 시스템인 HIIA-1a, Altavista, Lycos, Excite에 대해서 간단한 소개와 주요 특징, 차이점에 대해 살펴본다(Eagan, Bender 1996; Chu, Rosenthal 1996; 권근오, 유명기 1998). 실험을 위한 평가 시스템으로서 일상생활에서 자주 이용되는 검색엔진중 세 개의 검색시스템을 선정하였으며, 성능비교의 기준으로 동물에 대한 전문 지식베이스를 갖춘 HIIA-1a를 선정하였다.

##### 1) HIIA-1a

HIIA-1a는 홍익대학교 전자계산학과 에이전트연구실에서 개발한 지능형 정보 에이전트로 동물 영역에 관한 정보와 질의와 관련된 웹 문서를 제공하는 시스템이다. 즉, 사용자로부터 주어진 질의에 대한 질의 정보를 제공하고 더불어 질의와 관련된 웹 문서를 제공해주는 검색의 기능을 포함하고 있다. 검색 시스템의 관점에서 HIIA-1a의 특징은 다음과 같다.

첫째, HIIA-1a의 특징으로는 사용자가 보다 나은 질의식을 구성할 수 있도록 단순 불리언 연산자 AND, OR에 추가적인 연산자로 OF, MIN, MAX, LT, LE, GT, GE를 제공하고 있다. 이러한 연산자의 사용은 사용자가 좀 더 구체적인 질의를 쉽게 작성할 수 있도록 하여 사용자의 의도를 보다 정확히 명시할 수 있는 기능을 제공한다.

둘째, 웹 문서의 검색에서 보다 나은 웹 문서제공을 위하여 주어진 질의에 대한 질의 확장 단계를 처리한다. 이 처리 과정에서는 HIIA-1a가 가지고 있는 지식베이스를 이용하여 주어진 질의에 대한 객체의 계층 정보, 속성 정보, 지리 정보, 단위 정보를 추가하여 문서의 검색 과정에 질의 정보로 사용하게됨으로서 문서 검색의 효율성을 높이도록 하였다.

셋째, 웹 문서 검색은 본 시스템이 가지고 있는 데이터베이스를 이용하여 문서 검색을 처리한다. 검색에는 사용자가 주어진 질의를 기반으로 질의 확장 단계에서 추가된 질의들과 함께 질의 set을 구성하게 된다. 관련 문서의 추출은 주어진 질의와의 정합(matching)을 통해 관련 문서를 추출하고, 이렇게 추출된 문서들을 대상으로 질의 set과 각 관련문서간의 유사도를 계산한다. 이 유사도 값에 의해 사용자에게 제공할 문서의 순위를 결정하게 된다. 이런 유사도의 계산에는 단어 가중치를 위한 TF 방식을 이용하고, 유사도 값을 계산하기 위하여 질의 set과 문서간에 P-norm model을 적용하여 처리하고 있다.

마지막으로 HIIA-1a는 웹 문서의 제공을 위해서는 웹 문서를 대상으로 인덱스 처리 과

정을 통하여 자체 데이터베이스를 가지고 있다. 이 작업은 웹 로봇을 이용하여 웹 문서를 가져오면 이 문서가 동물 영역에 적합한 문서 인지를 판정하게 된다. 적합 판정을 받은 문서를 대상으로 시스템은 문서를 데이터베이스에 색인하는 작업을 처리한다. 또한 색인/저장이 된 문서에 대해서 변화여부(존재유무)를 지속적으로 감시하는 기능을 가지고 있어서 항상 최신의 문서들을 저장하고 있다.

## 2) Altavista

알타비스타는 DEC사에서 운영하는 것으로 '슈퍼 스파이더'라는 자동 검색 로봇을 통해 자료를 수집하는 방식을 채택하였는데, 뛰어난 검색능력과 정확성을 가진 검색엔진이다. 알타비스타의 특징은 검색능력의 정확성과 전문검색 지원으로 검색 속도 및 정확성에 있어서 다른 검색엔진의 추종을 불허하고 있으며 전문검색을 지원하기 때문에 문서 전체의 내용을 인덱스하여 데이터베이스로 구축하고 있다. 또한, 단순 검색과 고급 검색의 지원으로 알타비스타의 검색은 단순 검색과 고급 검색으로 구분하여 검색을 수행하고 있고, 이 두가지 검색의 공통 요소로는 단어, 구, 큰따옴표, 영문자의 대소문자, 절단기능, 제한검색이 있고 차이점은 단순검색은 연산 기호를 지원하지 않지만 고급 검색은 검색 연산자를 지원하고 있는 것이다. 우선 순위의 선정으로 알타비스타는 출력 우선순위 시스템(rating system)에 의해 검색결과가 출력되는데, 검색어 출력 우선 순위는 점수 매김 시스템에 의해 공통으로 적용되는 순위 추정을 의미한다. 단, 고급 검색에서는 순위 검색어를 지원하는데, 순위

검색어 입력란에 특정한 검색어를 입력하지 않으면 출력 우선순위에 따라 검색결과가 나타난다.

알타비스타의 검색결과는 일반적으로는 출력 우선순위 시스템에 의해 검색을 수행하지만, 고급 검색에서는 순위 검색어에 검색어를 입력하면 입력한 단어에 따라 출력 순위를 변경할 수 있다. 고급 검색어의 순위 검색어는 검색식에 사용된 검색 단어를 입력하거나 일반적인 단어를 입력할 수 있다. 알타비스타 출력 우선순위의 기준으로는 첫째, 검색한 단어가 검색단어의 표제(title) 또는 페이지의 앞부분에 나타나는 경우, 둘째, 검색한 단어들 페이지에서 서로 인접하고 있는 경우, 셋째, 검색한 단어의 사용 횟수가 많이 나타난 경우, 넷째, 검색한 단어 및 구들이 보다 많이 나타나는 경우 등이다. 알타비스타는 이러한 출력 우선순위 시스템에 의해 검색결과가 출력되는데, 검색어 출력 우선 순위는 점수 매김 시스템(scoring system)에 의해 공통으로 적용되는 순위 측정을 의미한다. 단, 고급 검색에서는 순위 검색어를 지원하는데, 순위 검색어 입력란에 특정한 검색어를 입력하지 않으면 출력 우선순위에 따라 검색결과가 나타난다. 또한, 라이브토픽스(LiveTopics)는 알타비스타 검색에 사용된 검색어 결과를 분석할 수 있는 도구로 라이브토픽스를 통해서 검색에 사용된 키워드를 직접 입력하지 않고 마우스로 클릭하여 선택하거나 검색어와 관련된 내용을 검색 대상에서 제외할 수 있다. 라이브토픽스의 특징은 첫째 일반적인 검색어를 정확한 검색어로 추가 선택하는 기능, 둘째 많은 검색결과 중 유용하지 못한 검색결과를 제외

하는 기능, 셋째 검색어와 유사한 주제별로 검색하는 기능 등을 가지고 있다.

### 3) Excite

익사이트(Excite)는 1993년 스탠포드 대학원의 학생들에 의해서 개발된 검색엔진으로 개발 당시에는 아키텍트 소프트웨어사(Architect Software Co.)로 출발하였으나 현재는 Excite Co.로 바뀌었고, 익사이트는 월드와이드웹을 비롯하여 유즈넷 뉴스까지 검색하며, 로이터통신사에서 제공하는 최신 뉴스와 웹사이트에 대한 평가 등을 제공하고 있다.

익사이트는 키워드 검색서비스(Excite Search)와 주제별 카탈로그 서비스(Channels by Excite)를 기본적으로 제공하고 있으며, 키워드형 검색엔진인 익사이트의 특징은 다음과 같다. 첫째, 익사이트는 키워드 검색기능과 주제 검색기능을 제공하는데, 익사이트 서치는 검색기능을 제공하는 것으로 익사이트의 홈페이지에서 검색할 수 있다. 익사이트 채널은 야후처럼 14가지 주제별 분류 목록을 제공하고, 홈페이지에서 'Channels by Excite'의 분류목록을 클릭하여 이용할 수 있다. 일반적인 검색엔진과 비슷하게 익사이트에서도 검색결과를 검색신뢰도 점수에 의해 백분율(%)로 표시하고 있다. 라이코스에서는 100%의 점수가 출력되지만 익사이트에서는 보통 85% 정도가 최대치의 점수로 나타난다. 키워드 입력에는 영문자의 대소문자를 구분하지 않고, 연산자를 이용하여 검색할 수 있다. 연산자로는 연산기호(+, -)와 연산자(AND, OR, NOT), 괄호 등을 사용하여 검

색할 수 있다. 일반적으로 웹을 통한 출력형태별 검색결과는 점수(%), 검색된 문서의 제목, 'More Like this', URL, 요약문 형태로 제공되며, 'More Like this'는 검색결과와 연결되어 있는 곳에 연결되었다는 것을 의미한다.

#### 4) Lycos

라이코스(Lycos)는 미국의 카네기 멜론 대학의 Michael L. Mauldin에 의해서 개발된 검색엔진으로 현재는 라이코스라는 회사가 설립되어 독립 법인으로 운영되고 있다. 라이코스는 'SCOUT Indexer'라는 검색로봇을 활용하여 웹페이지를 검색한다. 검색방법은 라이코스 18가지의 주제별 분류 및 주제별 검색서비스와 검색어를 입력한 후 검색하는 방법으로 구분되어 있고, 라이코스의 특징은 키워드형 검색엔진이면서 주제별 카탈로그 서비스와 연결되어 라이코스는 키워드 입력을 통해 정보를 찾아내는 키워드형 검색엔진이면서 a2z, 포인트 리뷰 서비스와 연계되어 있다. a2z는 인터넷의 자원을 16가지 주제별로 분류하여 제공하는 서비스이며, 포인트 리뷰는 포인트 사이에서 선정한 서비스로 상위 5% 사이트 등과 연결되어 있다. 익사이트에서는 검색결과를 출력할 때 검색신뢰도 점수를 백분율로 표시된 점수를 제공하는데, 라이코스에서는 검색된 결과 점수를 백분율로 표시한다. 라이코스에서는 최대 100%로부터 시작하여 점수를 부여한다. 검색수준의 정밀도는 검색결과의 정확성을 측정하는 기준으로 즉, 입력한 검색어가 검색된 페이지 내에서 많이 나타나거나 인접한 경우 높은 점수가 제공되는 것

으로 검색결과에서 검색수준을 설정할 수 있다. 즉, 검색수준 중 'loose match'는 가장 많은 검색건수를 보여주지만 신뢰성이 떨어지는 결과이며, 'strong match'는 검색건수는 적지만 신뢰성이 높은 결과를 의미한다.

### 3. 2 평가 성향

본 연구에서는 인터넷 정보검색시스템간의 성능비교의 목적보다는 검색 결과에 대한 이용자 만족도와 이용자 평가성향을 파악하는데 중점을 두어 조사를 진행하였다. 평가 작업을 수행해가는 과정에 있어 밝혀진 이용자 평가 성향 중 가장 두드러진 특징으로는 검색시스템으로 넘겨받은 표제(title)와 기술(description) 정보로부터 평가자가 표시한 만족/불만족의 평가값이 실제 웹 문서를 확인한 다음에 정확한 평가가 이루어진 결과와 큰 차이를 보이지 않는다는 점이다. 다시 말해 이용자가 검색시스템으로부터 넘겨받은 표제(title)와 기술(description)정보, URL을 포함한 검색결과만을 보고 별 가치 없으리라 여긴 웹 문서는 실제 그 문서를 방문한 후에도 별 가치 있는 정보를 얻을 확률이 극히 희박하다는 점이다.

웹 정보 검색결과에 대한 이용자들의 평가 결과를 수집 및 분석해 가는 과정에서 밝혀진 가장 큰 특징은 각기 개인적 성향의 차이를 보이나, 웹 정보 검색을 수행한 이용자들간에는 어느 정도 공통된 평가성향이 존재한다는 점이다. 검색된 특정 웹 문서에 대해 한 이용자가 불만족을 표시하는 경우, 다른 대부분의 이용자들 역시 불만족스럽게 여길 확률이 컸



〈표 1〉 검색시스템 평가(HIAA-1a, Altavista)

검색시스템 Query	HIAA-1a						Altavista				
		total	적합	비적합	애매	통신장애	total	적합	비적합	애매	통신장애
carnivora	평균값	100	67.231	22.538	5.7692	4.4615	100	46.923	39.923	9.1538	4
	중양값		69	24	3	4		47	45	5	3
cetacea	평균값	100	69.462	16.462	5.8462	8.2308	100	64.077	25.692	6.6923	3.5385
	중양값		72	13	5	9		72	23	2	3
chimpanzee	평균값	100	50.154	32.615	7.5385	9.6923	100	39.385	44.923	8.1538	7.5385
	중양값		52	32	9	11		37	49	6	8
giant panda	평균값	100	74.615	10	9.6923	5.6923	100	55.308	22.231	14	8.4615
	중양값		83	4	7	5		56	19	14	10
gorilla	평균값	100	33.769	53	6.5385	6.6923	100	18.231	65.615	8.6154	7.5385
	중양값		31	48	5	7		12	70	4	8
polar bear AND giant panda	평균값	51	31.769	7.9231	2.6154	4.1538	55	20.615	25.615	3.2308	5.7692
	중양값		32	5	2	5		18	19	2	5
cheetah AND chimpanzee	평균값	22	8	6.4615	5.1538	2.3846	22	5.7692	9.6923	3.8462	2.6154
	중양값		7	7	5	3		6	10	4	2
dog AND cat	평균값	100	44.538	36.154	11.692	7.6154	100	17.538	60.231	15	7.2308
	중양값		38	34	13	10		12	58	20	10
giraffe camel- opardalis OR giraffe	평균값	100	44.846	39.538	11.769	3.8462	100	38.231	38.231	9.6923	13.846
	중양값		42	38	11	3		44	30	9	17
polar bear OR giant panda	평균값	100	57.846	23.769	7.8462	6.0769	100	40.308	45.615	8.7692	5.3077
	중양값		60	26	6	4		34	47	5	3
lion OR tiger	평균값	100	41.692	41.769	9.1538	7.3846	100	30.308	54.154	5.1538	10.385
	중양값		43	52	12	10		24	61	5	10
baby OF vampire bat	평균값	20	6.9231	8.5385	1.3846	3.1538	20	5.6923	8.9231	1.6154	3.7692
	중양값		6	7	1	2		7	8	1	4
length OF giant panda	평균값	100	27.846	61.154	6.3077	31.769	100	16.462	68	6.7692	8.7692
	중양값		21	71	3	5		10	79	2	8
life span OF giant panda	평균값	100	24.846	65.308	5.9231	3.9231	100	15.769	63	7.3077	9.6154
	중양값		21	74	3	3		12	74	1	10
MAX size animal AMONG mammal IN land	평균값	100	31.077	58.692	4.3846	5.8462	100	24.538	64.846	5.7692	4.8462
	중양값		25	69	0	6		23	71	1	4
MIN offspring animal AMONG carnivora IN land	평균값	85	27.769	42.538	6.8462	5.9231	85	11.231	62.308	5.8462	4.4615
	중양값		28	43	1	6		10	64	1	4
heavy bear	평균값	100	74.231	13.769	7.3846	4.6154	100	16.154	68.308	5.9231	9.6154
	중양값		77	13	6	4		7	78	1	12
few offspring ursidae	평균값	100	64.231	19.538	10.462	5.7692	100	27.615	56.462	10	6.0769
	중양값		60	21	13	5		32	53	10	5
middle felidae	평균값	100	46.692	42.538	8.1538	2.6154	100	21.231	66.231	7.5385	5
	중양값		45	45	3	2		16	73	6	3

〈표 2〉 검색시스템 평가 (Lycos, Excite)

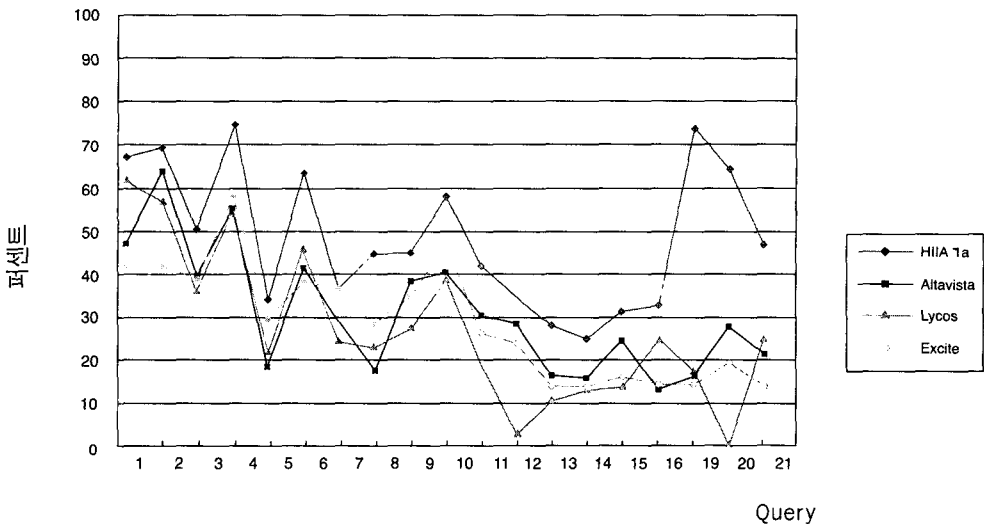
Query	Lycos						Excite				
		total	적합	비적합	애매	통신장애	total	적합	비적합	애매	통신장애
carnivora	평균값	100	61,538	21,154	8,3846	8,9231	100	41,923	42,077	8,3846	8,3846
	중양값		65	18	6	9		40	48	5	7
cetacea	평균값	100	56,615	31,231	8	4,1538	100	41,615	34,154	8,4615	15,769
	중양값		57	34	4	3		42	33	4	19
chimpanzee	평균값	100	35,769	37,231	8,1538	18,846	100	38,769	46,692	8	6,5385
	중양값		33	39	6	15		38	49	6	7
giant panda	평균값	100	57,846	23,923	6,692	11,538	100	58,154	29,231	6,846	5,7692
	중양값		60	21	2	10		60	31	3	5
gorilla	평균값	100	21,615	64,923	6,8462	6,6154	100	29,538	57,077	8,3077	5,0769
	중양값		18	68	1	7		28	62	5	5
polar bear AND giant panda	평균값	55	22,846	23,615	3,5385	5,2308	55	19,154	23	6,6923	6,3846
	중양값		20	17	2	5		19	18	6	6
cheetah AND chimpanzee	평균값	22	4,8462	11,154	3,1538	2,8462	22	7,3077	10,923	1,9231	1,8462
	중양값		3	12	3	3		5	13	2	2
dog AND cat	평균값	100	22,846	54,154	16,231	6,7692	100	28,385	44	20,154	7,4615
	중양값		22	53	19	6		30	44	20	7
giraffe camel- opardalis OR giraffe	평균값	100	27,154	53,154	12,231	7,4615	100	34,769	41,846	14,308	9,0769
	중양값		23	52	11	8		33	40	14	9
polar bear OR giant panda	평균값	100	38,385	44,538	9,2308	7,8462	100	46,846	29,692	8,6154	14,846
	중양값		29	54	8	9		49	26	7	15
lion OR tiger	평균값	93	17,769	53,077	4,0769	18,154	100	26,231	50,154	8,2308	15,077
	중양값		10	51	0	16		20	48	6	17
baby OF vampire bat	평균값	20	2,6923	11,769	1,2308	4,3077	20	4,7692	7,7692	2	5,4615
	중양값		1	13	1	4		5	5	1	8
length OF giant panda	평균값	100	10,385	70,846	7,1538	11,615	100	14	67,615	9,6154	8,7692
	중양값		8	77	2	13		13	70	6	9
life span OF giant panda	평균값	44	6,5385	39,923	3,4615	6,9231	100	13,923	65,846	7,3077	12,923
	중양값		6	32	0	7		8	73	2	15
MAX size animal AMONG mammal IN land	평균값	100	13,692	73,692	5,3846	7,2308	100	15,923	71,385	4,6154	8,0769
	중양값		10	80	2	7		14	76	0	7
MIN offspring animal AMONG carnivora IN land	평균값	25	4,9231	15,846	1,9231	2,3077	37	5,3077	23,538	2,3077	5,4615
	중양값		4	17	0	2		4	24	0	5
heavy bear	평균값	100	17,462	54,308	6,3077	21,923	100	14,385	62,461	9	14,154
	중양값		8	60	3	25		3	68	3	12
few offspring ursidae	평균값	1	0	0,9231	0	0,0769	100	19,231	67,231	7,6154	5,9231
	중양값		0	1	0	0		11	77	5	7
middle felidae	평균값	67	16,615	42,077	5,3077	3,3846	100	13,769	71,923	4,6154	9,6923
	중양값		21	53	5	2		8	80	1	9

다. 이러한 사실을 근거로 각 정보검색시스템에 대한 평가자들의 평가 결과를 평균값과 중앙값으로 나누어 정리하였다. 평균값은 최대값에 영향을 받아 평균치로서 의미를 상실할 수 있으므로 중앙값을 함께 계산하여 비교하였다. 평가 결과를 정리한 <표 1>과 <표 2>에서는 원 질의 21개중 17번째와 18번째 질의에 대한 결과는 제외되었는데 이는 HIIA-1a를 제외한 나머지 정보검색시스템에서는 충분한 검색결과를 얻을 수 없었던 관계로 결과에서 제외되었다.

<표 1>과 <표 2>를 살펴보면 대부분의 경우 평균값과 중앙값의 차이가 심하지 않음을 살펴볼 수 있는데, 이는 언급한 바대로 동일한 주제로 웹 문서를 검색 시 검색된 문서에 대한 이용자들의 만족도는 유사한 패턴을 보이고 있다. <그림 2>는 각 질의(Query) 별로 정확률의 변화를 보여주고 있으며, 각 검색시스

템 별로 정확률은 정규화과정을 통해 계산하였다. 각 검색시스템 별 정확률의 변화에서 모든 시스템에서 공통적으로 21개의 질의 Set중 12번째부터 16번째까지 질의에 대한 검색 결과에서는 나머지 질의들과 대조적으로 대부분 평가자들이 낮은 검색 만족도를 표시하였음을 살펴볼 수 있다.

이는 12번째부터 16번째까지 질의들이 대부분 구체적인 검색 목적을 가지고 검색을 시도하였다는 가정 하에 검색 결과에 대한 평가가 이루어졌기 때문이다. 일례로 3번째 질의의 경우 막연히 "침팬지에 관한 내용을 가지고 있는 문서"를 찾는 것을 목적으로 하며, 평가자는 침팬지와 관련된 문서가 검색되면 검색 결과에 대해 만족을 표시하게 된다. 이에 반하여 12번째 질의의 경우 "황금박쥐(vampire bat)가 낳는 새끼의 수에 관한 문서"라는 구체적인 검색 목적을 가지고 있다. 전자의 경우



<그림 2> 각 Query별 정확률의 변화

검색 목적이 뚜렷하지 않은 대략적인 검색을 시도하여 검색된 문서들 가운데 관심 있는 문서를 찾아내는 것으로 'loose match'에 해당하며, 구체적인 검색 목적을 가지고 검색을 시도하여 검색목적에 부합되는 문서를 찾아내는 후자의 경우는 'strong match'에 해당한다. 실제 인터넷 정보검색 시에는 위 두 가지 상황이 모두 발생하게 되는데, 초보자일수록 정확한 검색식을 작성하여 구체적인 웹 문서를 검색해 가는데 어려움으로 인하여 낮은 검색 만족도를 얻게 된다.

실험 결과는 모든 검색시스템에서 이용자가 'strong match'를 시도하게 되는 경우 검색 정확률이 평균 40%이하임을 살펴볼 수 있으며, 이는 실제 인터넷 정보검색 수행시 대부분의 이용자들이 정보검색 결과에 대해 대체적으로 불만족함을 보여주는 근거를 제시하고 있다. 이러한 결과는 이용자들이 인터넷 정보 검색 시스템을 사용하는데 있어 매번 자신이 찾고자하는 목적에 부합되는 웹 문서가 검색 결과 중 상위에 위치하기를 기대하는 정보검색 태도는 바람직한 인터넷 정보검색 방식이 아님을 보여주고 있다. 인터넷 정보검색 시스템에서 높은 검색 정확률을 기대할 수 없는 현 상황에서는 인터넷 정보검색 시스템에 전적으로 의존하는 대신 인터넷 정보검색 시스템을 사용하여 자신이 찾고자하는 주제와 부합되는 관련 웹 문서들을 찾아내고, 웹 서핑을 통해 자신의 검색 목적에 부합되는 웹 문서를 찾아내는 방법이 초보이용자들에게는 보다 효과적인 정보검색의 방법일 수 있겠다.

#### 4. 인터넷 이용자 성향분석

일반적으로 대다수의 이용자들이 인터넷을 이용하는 수준은 음료 자동판매기를 이용하는 과정에 비유될 수 있다. 대부분 자신이 원하는 상품을 고르는 수준에서 검색어 한 두 개를 고르며, 이를 통해 얻은 검색 결과에 흡족하지 않더라도 더 이상의 재검색을 시도하지 않는다(대부분이 선택한 음료수가 이용자 기호에 잘 맞지 않더라도 그냥 참고 마시는 것과 같다). 이보다 높은 수준에서 인터넷 정보검색을 수행하는 과정은 약간 더 복잡한 커피 자동판매기를 이용하는 경우에서처럼 커피의 종류를 선택하고, 설탕을 추가하고 프림을 빼는 몇 가지 추가적인 단계를 거치는 것과 유사하게 진행되어 진다. 즉, 단순 검색에 비해 보다 신중하게 검색어를 선정하며, 검색연산자를 추가하거나 구절검색(phrase)을 통해 보다 효과적인 검색이 되도록 유도한다.

검색 결과가 불만족스러운 경우 검색어를 약간 수정하는 수준에서 재검색을 시도하기도 한다. 하지만 대다수 이용자의 수준이 초보인 상황에서 이용자 수준에 따른 검색의 차이는 그다지 변별력을 가지지 못한다. 대부분 이용자들은 정보검색 수행시 적극적으로 문제를 해결해가기 보다는 상황에 따른 수동적인 검색 패턴을 보이며, 정보 검색 전략 역시 매우 단순하다. 이는 인터넷 정보검색이라는 특수한 상황에서는 정보검색 시 이용자들의 행동 양식을 분석하는 과정을 통해 이용자의 내면적 심리상태나 요구사항을 파악할 수 있음을 의미한다. 선행된 연구들의 결과와 검색시스템 평가결과를 바탕으로 먼저 새로운 정보검

색 환경인 인터넷과 인터넷 이용자의 특성을 외면적 행동 양식의 측면과 숨겨진 내면인 이용자 성향 측면에서 다음과 같이 정리할 수 있다.

#### 4. 1 인터넷의 특징

- 내용적 특성 : 인터넷의 내용적인 측면에서 가장 큰 특징은 주제의 다양성을 들 수 있다. 인터넷은 특정 주제를 대상으로 한 서지 데이터베이스와는 달리 거의 모든 주제를 대상으로 하며 항목별로 정리되어 있지 않다. 이것은 이용자의 정보 요구에 부합되는 적합한 정보를 찾기가 보다 힘들어짐을 의미한다.

- 형식적 특성 : 인터넷의 형식적인 측면에서 가장 큰 특징은 매체의 다양성을 들 수 있다. 텍스트 위주의 서지 데이터베이스 검색과는 달리 인터넷의 웹 문서는 텍스트, 그림, 사운드, 동영상 등 다양한 매체들로 구성되어 있으나, 아직까지 그림, 사운드, 동영상 등 다양한 매체로 구성된 웹 문서들을 효과적으로 검색하는 방법이 개발되어 있지 못하다.

- 중복성 : 인터넷 정보검색 시 보통 검색된 문서의 30%정도가 서로 유사하거나 동일한 중복 문서이다.

- 유사성 (Resemblance) : 사용자들이 입력하는 검색식의 상당수가 서로 동일하거나 유사한 경우가 많았다.

- 인터넷의 규모 : 인터넷에서 검색 가능한

정보의 양은 일반적인 서지 데이터베이스에 비해 훨씬 더 많다.

- 웹 문서의 유효기간 : 한번 방문한 웹 문서를 일정기간이 경과한 후 재방문시 내용이 바뀌었거나, 인터넷상에 존재하지 않을 가능성이 높다.

- 웹 문서의 품질 : 대부분의 웹 문서들은 출판되기까지 과정에 있어 내용과 형식적인 측면에서 아무런 검열과정을 거치지 않는다. 따라서 서로 다른 웹 문서들 상에서 동일 대상을 가리키는 수많은 어휘적인 표현이 사용되어 질 수 있으며, 철자의 오류를 포함할 가능성 역시 서지 데이터베이스에 비해 훨씬 높은 편이다.

#### 4. 2 인터넷 이용자의 행동 양식

- 질의어 수(short query) : 보편적으로 인터넷 이용자들은 검색식 작성 시 1개내지 2개 정도의 검색어를 중심으로 하여 검색식을 작성한다. 선행연구에서도 나타났듯이 Papka와 Allan(1998)의 연구에서 조사된 평균 질의어 수는 2.3개였으며, Jansen et al.(1998)의 연구에서는 2.35개로 대체적으로 질의어 수는 1개내지 2개임을 확인할 수 있었다. 이는 Silverstein et al.(1998)의 연구 결과에서 보다 명확해지는데 전체 질의 중 약 70%이상 질의어 수가 2개 이하였음이 밝혀졌다.

- 세션 당 평균 질의어 수 : 이용자가 검색 목적을 가지고, 처음 검색을 시작하여 검색을

종료하기까지의 일련의 과정을 하나의 세션이라 부른다. 한 세션 당 평균 질의어 수 조사에서 전체의 약 80%가 1개였음이 밝혀졌다.

- 재검색 시도 : 검색식을 재조정하여 재검색하는 경우가 거의 없었다.

- 상위 10개(Top 10) : 대부분 이용자들은 검색된 문서들 가운데 상위 10위 내에 문서들만을 방문하는 것으로 조사되었다.

- 평균 연산자 수 : 한번의 질의에서 사용되는 평균 연산자수는 0.41개로 대부분 이용자들은 검색식을 작성하는 과정에 있어 검색 연산자를 잘 사용하지 않는 것으로 밝혀졌다.

- 재검색 방식 : 이미 앞에서 언급한 바대로 대부분의 이용자는 검색식을 재조정하지 않는다. 검색식을 재조정하여 재검색을 하는 경우에도 키워드나 검색 연산자를 추가, 삭제하는 적극적인 검색행동은 전체 중 10%정도에 불과하며, 대부분은 원래 키워드를 재수정하는 정도에서 재검색을 시도하거나 질의식을 새로 작성하여 새로운 검색을 시도하게 된다.

#### 4. 3 인터넷 이용자 성향

- 정보 검색 수준 : 대체적으로 웹 정보 검색을 수행하는 이용자들의 정보검색 수행능력은 그다지 뛰어나지 않으며, 실제 정보 검색시 그다지 주의 깊게 정보 검색을 수행하고 있지 못하다. 대부분 이용자들의 웹 정보 검색수행은 기존의 전자가전제품을 이용하는 수준에

불과하며, 합리적인 정보검색 행동(rational motion)을 보이기보다는 과거의 경험에 따른 반응적 행동(reactive motion)을 취하는 경향이 있다.

- 단순한 검색전략 : 정보검색 초보자의 전략은 비교적 단순하다. 이는 정보검색 숙련자도 역시 별반 큰 차이를 보이지 않는다. 정보사냥 대회에 참가한 숙련자의 검색행동의 분석 결과도, 검색어의 수, 연산자의 종류와 사용빈도, 검색어 유형 등에서 초보자의 결과와 크게 다르지 않았다. 질의문 작성시 초보자와 숙련자 모두 (최초의) 한두 검색어를 중심으로 하여 다른 검색어들을 교체하곤 하였다.

- 수동적 검색성향 : 대부분의 이용자들은 충분한 문제 해결 능력을 갖추고 있지 못하며, 따라서 시스템 스스로 문제를 해결해주기를 바란다. 이러한 사실은 기존의 불리안 방식검색엔진에서 자주 자연어검색이 시도되어 진다는 점에서 확인할 수 있다.

- 이용자정보 : 이용자로부터 이용자정보(User Profile)를 얻을 수 있는 횟수는 사실적으로 1회로 제한된다. 지속적인 추가정보요구는 이용자에게 시스템 이용을 회피하게 만드는 역효과를 초래한다.

- 인터넷 이용자 행동성향 : 서로 다른 성격을 지닌 이용자그룹간의 정보 서비스이용 형태는 별다른 큰 차이점을 보이지 않는다고 보고하고 있다. 이용자들은 대부분 적은 숫자의 단어로 구성된 간단한 질의문을 선호하며,

이용자의 컴퓨터에 관한 충분한 사전지식의 유무를 떠나 질의문 작성시 사용하는 연산자에 대해 별다른 주의를 기울이는 않는다.

• 인터넷 이용자 평가성향 : 인터넷 정보 검색 시스템 검색 결과 평가데이터를 분석한 결과 다음과 같은 사실을 확인할 수 있었다. 웹 정보 검색결과에 대한 이용자들의 평가는 각기 개인적 성향의 차이를 보이나, 웹 정보 검색을 수행한 이용자들간에는 공통된 평가성향이 존재한다는 점이다. 다시 말해 특정한 웹 정보 검색결과에 대해 한 이용자가 불만을 높게 표현하는 경우, 다른 대부분의 이용자들도 동일 검색 결과에 대해 불만족스럽게 여길 확률이 크다.

• 정보검색전략 : 추가적인 질문을 통해 실험 참가자들의 정보 검색 유형을 확인해본 결과, 검색엔진선정이나 검색식 선정에 있어 자신의 검색 목적을 확인하여 검색 전략을 수립하기보다는 과거의 경험에 의존하여 정보 검색을 수행하는 모습을 살펴볼 수 있었다.

#### 4. 4 인터넷 정보검색시스템 설계를 위한 가이드라인

앞에서 살펴본 바와 같이 대부분 이용자들은 웹 정보검색 수행 시 단순한 검색전략 방식을 취한다. 이는 웹정보검색 이용자들의 정보검색 성향을 관찰을 통해 보다 이용자 친화적(User friendly)인 시스템의 설계가 가능함을 시사하고 있다. 실제 이용자들은 인터넷 정보검색시스템의 작동원리와 같은 기술적인 측

면에 미숙하며 자신이 가지고 있는 개념을 검색식으로 구체화하는데 어려움을 가지게 된다. 따라서 인터넷 정보검색시스템은 이용자에게 최대한 부담을 주지 않는 방향에서 주어진 제한된 정보를 효과적으로 활용하여 작업을 수행해나가야 한다.

대부분 이용자들은 2개 이하의 매우 적은 수의 검색어 만을 입력하며, 검색식을 작성하는 과정에 있어 검색연산자를 잘 사용하지 않는 사실에서 이용자들의 검색 방식이 단일 질의나 AND 연산자(주요 검색엔진에서 연산자가 생략된 경우 디폴트 연산자는 AND임)만으로 구성된 짧은 검색식을 통해 실제 정보검색을 수행함을 유추할 수 있다. 이는 대부분의 인터넷 정보 검색 시스템이 이용자로부터 얻을 수 있는 정보가 단일 질의 형태이거나 AND 연산자와 2개 내지 3개 정도의 검색어로 구성된 짧은 검색식으로 한정된다. 또한 이용자로부터 추가적인 정보(user profile)를 획득할 수 있는 횟수가 극도로 제한되는 현실에서 인터넷 정보 검색 시스템은 제한된 정보를 효과적으로 활용하여 작업을 수행해나가야 한다. 이용자에게 이용자 정보(user profile)를 1회 이상 요구하는 것은 사용자에게 시스템 이용을 회피하게 만드는 역효과를 초래할 수 있다. 따라서 대부분의 정보 검색엔진들에서도 이용자별 맞춤 서비스를 제공하고자 하는 경우, 처음 등록시 이용자로부터 개인 정보를 얻는 방식을 채택하고 있다. 매번 이용자 피드백(user feedback)정보를 제공받기를 요구하는 시스템은 그다지 현실적으로 높은 이용자 피드백 회수율을 기대할 수 없으며, 결국 이용자 피드백 정보를 이용한 학습이나 이를 통한

검색식 재조정은 실용적이지 못하다.

동일주제에 대해 여러 이용자들의 비슷한 정보검색요구가 반복하여 발생할 가능성은 상당히 높으며, 이를 위해 대부분의 인터넷 정보 검색시스템들은 웹 캐쉬(Web Cache)를 마련하여, 동일 질의에 대해 보다 빠른 서비스를 제공한다. 또한 이용자의 이용기록(log file)을 상세히 분석하는 과정을 통해 이용자의 이용실태를 조사할 수 있으며, 이를 보다 나은 서비스로 제공하기 위한 자료로 활용할 수 있다. 이용자의 이용기록을 활용한 실제 사례로 이용자가 검색된 문서들 가운데 관심 있어하는 문서를 기준으로 하여 이와 유사한 문서들을 찾아주는 효과적인 검색 기법으로 'Search by Example'이란 기술이 사용되어지고 있으며, Netscape 검색엔진이 제공하는 'What's related'가 이에 해당한다. 'What's related'의 방법은 이용자가 관심을 보인 문서 D1에 대해 과거의 이용자들이 문서 D1 다음으로 방문한 문서 D2를 찾아내어 연관성 있는 문서로 제공하게 된다. 따라서 관련 있는 유사한 문서들을 찾기 위해서 이용자로부터 별도의 검색어 입력을 필요로 하지 않으며, 초보자라도 쉽게 이용할 수 있는 유용한 도구로 활용되어질 수 있다. 이는 인터넷 이용자 성향 중 '수동적 검색 성향'의 측면을 제대로 반영한 효과적인 인터넷 검색 시스템 설계에 해당한다.

현재까지 대부분의 인터넷 정보검색시스템은 대부분 낮은 검색 정확률을 보이고 있다. 실제 웹 정보 검색을 수행하는 과정에 있어 검색된 문서들 가운데 상위 10위 내의 문서들 중에서 자신이 찾고자하는 문서가 존재하는

경우는 극히 희박하다. 앞으로도 혁신적인 웹 정보검색 기술이 개발되지 않는 이상 인터넷 정보검색시스템은 높은 검색 정확률을 제공할 수 없으며 이는 이용자의 부담으로 작용하게 된다. 따라서 이용자 스스로 자신의 정보검색 능력을 향상시키기 위한 노력을 필요로 하고 있다. 그중 하나로 검색된 문서들 가운데 상위 10위 내에 문서들만을 살펴보는 검색 방식은 매우 비효율적인 검색 방식이며 이러한 검색 습관은 고쳐져야 한다. 대부분의 웹 정보 검색 가이드라인에서는 상위 30위까지의 문서들을 살펴보는 것을 권장하고 있으며 스스로 적극적인 문제해결이 어려운 일반 이용자들의 수동적인 검색 성향을 반영해볼 때, 검색 시스템 스스로 이용률이 낮아 빠른 서비스가 가능한 시간대에는 한 페이지에 제공하는 검색 문서 수를 반드시 10개가 아닌 상황에 맞추어 20개 내지 30개까지 유동적으로 늘리는 방법도 고려해 볼 수 있다.

선행연구조사 및 이용자 검색 평가성향조사에 초점을 둔 실험연구를 토대로 일반이용자들이 보다 쉽게 사용할 수 있는 인터넷 정보 검색시스템을 설계하기 위한 가이드라인은 다음과 같다.

- 이용자 중심의 시스템 설계 : 실제 이용자의 요구사항을 조사, 분석하는 과정을 통해 시스템 이용자들의 보편적인 성향 파악이 가능하며 이를 토대로 시스템 설계가 이루어져야 한다.

- 이용자의 관점 vs. 설계자의 관점 : 대부분의 인터넷 이용자들은 정보검색교육을 받은



적이 없는 초보자이며 인터넷 정보검색 및 정보검색시스템에 작동원리에 대해 잘 알고 있지 못하다. 시스템 설계자가 쉽게 여기는 부분도 실제 이용자는 어렵게 느껴질 수 있으며, 정보검색시스템의 다양한 기능을 제대로 활용하지 못하는 경우가 대부분이다. 시스템 설계자는 이러한 사용자들의 특성을 반영하여 이용자의 관점에서 시스템을 설계하여야 한다.

- 검색과정의 간소화 : 인터넷 정보검색에 숙련되지 않은 일반이용자들은 검색 만족도 못지 않게 이용의 편리성을 중요시한다. 대부분 다양한 검색연산자를 이용한 복잡한 검색식을 구성할 능력이 없는 일반이용자들은 AND나 OR 정도의 간단한 연산자를 사용하거나 검색연산자 없이 키워드만으로 검색을 시도하게 된다.

- 이용자 부담 최소화 : 직접적인 키워드 입력 검색방식 대신 이용자 관심사항을 토대로 시스템 스스로 연관 있는 웹 문서를 찾아주는 자동화된 검색시스템을 구현할 수 있다. 하지만 이용자 관심사항을 파악하기 위해 자주 이용자에게 연관된 피드백정보를 얻는 방법은 이용자에게 부담으로 작용하며, 이용자로 하여금 시스템 이용을 회피하게 되는 역효과를 초래하기 쉽다.

- 달라진 정보검색 환경 : 인터넷 정보검색 환경은 문헌검색 위주의 정보검색과는 여러 면에서 다르다. 따라서 시스템 설계자는 달라진 정보검색 환경에 대한 충분한 이해를 토대로 시스템 설계에 임해야 한다.

## 5. 결 론

본 연구에서는 네 가지 서로 다른 검색시스템을 검색 결과 비교를 통해 이용자 검색평가 성향을 조사하였으며, 이를 통해 지금까지 잘 알려지지 않았던 검색결과 평가의 측면에서 인터넷 이용자 성향을 파악하였다. 선행된 연구들의 연구결과와 검색시스템 평가결과를 바탕으로 새로운 정보검색 환경인 인터넷의 특징을 살펴보았으며 인터넷 사용자들의 특징을 외면적 행동 양식의 측면과 숨겨진 내면인 이용자 성향 측면에서 분석해보았다. 또한 밝혀진 여러 단편적인 사실들을 종합, 분석하여 보다 상세한 분석을 시도해보았다. 분석결과는 인터넷 이용자들은 과거의 정보검색자들과 다른 검색 행동양식을 갖고 있기 때문에, 보다 나은 정보서비스를 하기 위해서 새로운 사용자들의 특성을 활용하여 효율적인 인터넷기반 정보시스템 설계를 위한 가이드라인으로 활용될 수 있을 것이다.

인터넷의 급격한 팽창 및 높은 보급 증가율에 비하여 아직까지 인터넷에 관한 연구는 초기 단계에 머무르고 있는 실정이다. 또한 인터넷을 이용하는 이용자들은 그들이 이용하는 인터넷 정보검색 시스템에 대해 잘 알지 못하며, 시스템 설계자 역시 사용자들의 요구사항을 제대로 반영하지 못하고 있다. 하지만 대다수의 이용자들이 인터넷 정보검색 교육을 거의 받을 수 없는 현 상황에서 이용자들에게 인터넷과 인터넷 정보검색 시스템에 대해 스스로 잘 이해해서 제대로 사용하기를 바라는 것은 거의 불가능에 가깝다. 따라서 시스템 설계자에게는 사용자들의 요구사항을 파악하여

이를 새로운 인터넷 정보검색 시스템의 설계에 반영하는 적극적인 자세가 요구되어 진다. 하지만 인터넷 정보검색에 대한 새로운 패러다임이 정립되지 않은 현 상황에서는 이용자나 설계자 모두가 어려운 상황에서 작업을 진행해 나갈 수밖에 없으며, 앞으로도 장기간에 걸쳐 혼란은 계속되어질 것이다. 하지만 어느 시기에 이르면 새로운 정보검색 환경인 인터넷을 위한 새로운 패러다임이 정립되어질 것이며, 본 연구 역시 새로운 패러다임이 조속

히 정립되기를 기대하는 일련의 기초연구 중 하나로 진행되어 진 것이다.

\* 후기

본 연구를 위한 실험에 참가해준 홍익대학교 전자과 대학원생, 에이전트 연구실 조교, 숙대 문헌정보학 실습실 조교 여러분들과 연구에 많은 도움을 주신 홍익대학교 정보전산원장 변영태 교수께 감사사를 드린다.

참 고 문 헌

권근오, 유영기. 1998. 『인터넷 정보검색사 한번에 끝내기』. 서울: 한글과 컴퓨터.

박창호, 박민규. 1998. 인터넷의 정보검색에서 탐색학습과 사용자의 전략. 『한국인지과학회』, 9(4): 17-32.

박창호, 염성숙, 이정모. 1999. 정보검색엔진에서 홈페이지 분류구조 및 학습. 『1999년도 한국인지과학회 춘계 학술대회』, 1999년 5월 29일. [서울: 고려대학교].

이응봉. 1999. 인터넷 웹사이트 문서의 평가 기준 및 방법-내용적인 측면을 중심으로 『한국도서관·정보학회지』, 30(1): 151-169.

이용현. 1999. 『정보통신망에서 지능형 정보 에이전트와 특정 영역에서의 구현』, 박사학위논문, 홍익대학교 대학원, 전자공학과.

정영미, 김성은. 1997. WWW 탐색도구의 색인 및 탐색 기능 평가에 관한 연구. 『한국문헌정보학회지』, 31(1): 153-184.

황상규, 오경묵, 변영태. 1999. 어휘의미중의성이 인터넷기반 정보검색에 미치는 영향. 『1999년도 한국정보관리학회 학술대회』, 1999년 8월 18-19일. [서울: 연세대학교].

Abdulla, G., Liu, B. and Fox, E. A. 1998. Searching the World-Wide Web: Implications from studying different user behavior. WebNet98, Orlando. <www.microsoft.com/usability/webconf.htm>

Bains, S. 1997. End-user searching behavior: considering methodologies. The Katharine Sharp Review. ISSN 1083-5261, No. 4.

- <<http://edfu.lis.uiuc.edu/review/winter1997/bains.html>>
- Ballard, T. 1994. Comparative searching styles of patrons and staff. *Library Resources and technical Services*, 38(3): 49-72.
- Beavers, A. F. 1998. Evaluating search engine models for scholarly purposes(A Report from the Internet Applications Laboratory). *D-Lib Magazine (The Magazine of Digital Library Research)*. ISSN 1082-9873. <<http://www.dlib.org>>
- Chu, Heting and Rosenthal, M. 1996. Search engines for the World Wide Web: A comparative study and evaluation methodology. *American Society for Information Science*. <<http://www-diglib.stanford.edu/cgi-bin/WP/get/SIDL-WP-1997-0055>>
- Eagan, A. and Bender, L. 1996. Spiders and Worms and Crawlers, Oh my: searching on the World Wide Web. <<http://www.library.ucsb.edu/untangle/eagan.html>>
- Microsoft. 1996, Empirical Studies. In the second conference in the human factors and the Web/HTML Conference series. <<http://www.microsoft.com/usability/webconf.htm>>
- Henzinger, M. R., Heydon, A., Mizenmacher, M. and Najork, M. 1999. Measuring index quality using Random Walks on the Web. *Proceeding of the 8th International World Wide Web Conference*, Toronto, Canada.
- Josef, F., Alfred, K. and Andreas, N. 1997. Adaptable and adaptive information access for all users, including the disabled and the elderly. *The Sixth International Conference on User Modeling*, Springer Wien, New York.
- Jansen, B. J., Spink, A., Bateman, J. and Saracevic, T. 1998. Real life information retrieval: a study of user queries on the Web, *SIGIR FORUM*, 32(1): 5-16.
- Pollock, A. and Hockley, A. 1996. What's wrong with Internet searching. In *Designing for the Web empirical studies*. Microsoft. <<http://www.microsoft.com/usability/webconf.htm>>
- Papka, R. and Allan, J. 1998. Document classification using Multiword features. *Proceedings of the ACM 7th international conference on information and*

knowledge management, 124-131.

Schlichting, C. and Nilsen, E. 1998. Signal detection analysis of WWW search engines. <<http://www.microsoft.com/usability/webconf/schlichting/schlichting.htm>>

Silverstein, C., Henzinger, M., Marais, H. and Moricz, M. Analysis of a very large AltaVista query log. SRC Technical Note 1998-014. <<http://gatekeeper.dec.com/pub/DEC/SRC/technical-notes/abstracts/src-tn-1998-014.html>>

<부록>

- 21개의 Query List

No	Query	Search Engine Query	의미	Tip.
1	carnivora	左同	육식동물에 관한 내용을 가지고 있는 문서	육식동물
2	cetacea	左同	고래에 관한 내용을 가지고 있는 문서	고래
3	chimpanzee	左同	침팬지에 관한 내용을 가지고 있는 문서	
4	giant panda	左同	육식동물에 관한 내용을 가지고 있는 문서	
5	gorilla	左同	고릴라에 관한 내용을 가지고 있는 문서	
6	polar bear AND giant panda	左同	북극곰(polar bear)과 giant panda 둘 다에 관한 내용을 가지고 있는 문서	
7	cheetah AND chimpanzee	左同	치타(cheetah)와 침팬지(chimpanzee) 둘 다에 관한 내용을 가지고 있는 문서	
8	dog AND cat	左同	개(dog)와 고양이(cat) 둘 다에 관한 내용을 가지고 있는 문서	
9	giraffe camelopardalis OR giraffe	左同	기린에 관한 학문적인 내용을 가지는 문서를 찾기위해서 기린(giraffe)과 기린의 학명(giraffe camelopardalis)을 OR 처리함.	기린
10	polar bear OR giant panda	左同	북극곰(polar bear)이나 giant panda 둘 중 하나의 내용이라도 포함한 문서	
11	lion OR tiger	左同	사자(lion)와 호랑이(tiger) 둘 중 하나의 내용이라도 포함한 문서	
12	baby OF vampire bat	baby, vampire bat	황금박쥐(vampire bat)가 낳는 새끼의 수에 관한 문서	
13	length OF giant panda	length, giant panda	자이언트 팬더(giant panda)의 키(길이)에 관한 문서	
14	life-span OF giant panda	life span, giant panda	자이언트 팬더(giant panda)의 수명에 관한 문서	
15	MAX size animal AMONG mammal IN land	+largest +mammal +land	땅에 사는 포유류 중에서 크기가 가장 큰 동물은?	African elephant

16	MIN offspring animal AMONG carnivora IN land	+offspring +carnivora +land	땅에 사는 육식동물 중 새끼를 가장 적게 낳는 동물은?	felis lynx acinonyx jubatus lynx lynx canis familiaris martes flavigula enhydra lutra phoca groenlandica giant panda polar bear asiatic black bear
17	GT 5m length animal AMONG mammal			
18	LT 30 day gestation animal			
19	heavy bear	heavy bear	곰과에 속하는 동물 중 무거운 편에 속하는 동물은?	brown bear polar bear
20	few-offspring ursidae	few, offspring, ursidae	곰과(ursidae)에 속하는 동물 중 무거운 편에 속하는 동물은?	giant panda brown bear
21	middle felidae	middle, length, felidae	고양이과(felidae)에 속한 동물 중 크기가 중간정도인 동물은?	felis concolor acinonyx jubatus panthera pardus