

정보 검색에서 대량의 검색결과를 효과적으로 볼 수 있는 “결과 렌즈”의 설계

* ** * **
강현규, 홍성혁, 김영섭, 이용석
* **

ETRI 자연어처리 연구실, 전북대 전산통계학과

A Design of “Results Lens” for Effective Viewing of Massive Results in Information Retrieval

* ** * **
Hyun-Kyu Kang, Seong-Hyuk Hong, Young-Sum Kim, Yong-Seok Lee
* **

Natural Language Processing Section/ETRI, Dept. of Computer Science/Chonbuk National University

요약

정보검색 시스템이나 웹(Web)이 방대해지고 초보적 수준의 사용자들이 늘어남에 따라 간단한 조작만으로 원하는 정보를 얻어낼 수 있는 도구의 개발이 점점 중요해지고 있다. 일반적으로 정보검색 시스템 (검색엔진 포함)의 검색되는 결과의 수는 정보검색 시스템 별로 다양하고 많은 시스템들이 검색을 못하거나 방대한 결과를 제시한다. 본 논문은 방대한 결과를 여러가지 다양한 view로 볼 수 있고 사용자의 피드백에 의해 정보를 재 검색 함으로서 검색하고자 하는 문서를 보다 정확하게 검색할 수 있도록 도와주는 **결과 렌즈 (Results Lens: RL)**를 제안한다. 본 논문에서 제시한 **결과 렌즈**는 다양한 view로의 결과를 수정(refine) 할 수 있으며 웹을 기반으로하는 플러그인(plugin)으로 제공함으로써 유용성 및 확장성이 크다.

1. 서론

주요한 정보원으로서의 인터넷의 잠재적인 능력은 인터넷상에서 사람들이 어떻게 원하는 정보를 빠르고 정확하게 찾아낼 수 있는가에 달려있다. 웹(Web)이 방대해지고 초보적 수준의 사용자들이 늘어남에 따라 간단한 조작만으로 원하는 정보를 얻어낼 수 있는 도구의 개발이 점점 중요해지고 있다. 사용자들은 컴퓨터에게 원하는 분야를 말하고 어떤 결과를 얻는다는 개념에는 익숙치 않다. 예를들어 사용자가 도서관에 가서 사서에게 정보를 요청한다는 것은 이해할 수 있지만 인터넷에서 원하는 정보를 찾아내기 위해 검색엔진을 사용하는것은 비록 두 절차가 논리적으로 동일하다고 해도 사용자들은 쉽게 이해할 수 없다.

일반적으로 정보검색 시스템 (검색엔진 포함)의 검색되는 결과의 수는 정보검색 시스템 별로 다양하고 많은 시스템들이 검색을 못하거나 방대한 결과를 제시한다. 그러나 일반 사용자들은 물론, 검색 목적이나 사용자에 따라서 다르지만 일반적으로 검색해 준 결과의 많지 않은 수 (본인의 경우 약 20개 정도 내외) 만을 검토하게 된다. 만일 동일 데이터베이스(정보)를 다른 정보검색 시스템이 검색하는 경우 해당 정보검색 시스템의 색인이나 검색 모델에 다소 차이가 있으나 많은 수의 정보를

검색한다고 가정했을시는 검색 순위에 다소 차이가 있을지는 모르지만 실제 필요로하는 문서는 전체적인 검색 문서 안에 포함되어 있을 수 있다. 따라서 방대한 검색 결과를 다양한 viewing을 통하여 원하는 정보를 검색할 수 있는 방법이 필요하다.

본 논문에서는 방대한 결과를 여러가지 다양한 view로 볼 수 있고 사용자의 피드백에 의해 정보를 재 검색 함으로서 검색하고자 하는 문서를 보다 정확하게 검색할 수 있도록 도와주는 **결과 렌즈 (Results Lens: RL)**를 제안한다.

제 2장에서는 결과 렌즈의 요구사항에 대해서 설명하고, 제 3장에서는 결과 렌즈 설계시 고려사항에 대하여 나열한다. 제 4장에서는 설계된 결과 렌즈의 화면을 중심으로 설명한다. 마지막으로 제 5장에서 결론을 맺는다.

2. 결과 렌즈의 요구사항

다음은 현재의 정보검색 시스템이나 웹상의 검색 엔진들에 있어서 개선되어야 할 요구사항들에 대하여 나열한다.

- **결과의 view**
 검색 결과의 갯수(set)를 명시할 수 있어야하며 여러가지 시각적인 view를 통하여 사용자가 검색된 결과를 한눈에 알아볼 수 있도록 해야한다. 또한 현재 상태의 정보를 알 수 있는 도움이 될만한 메시지(message)를 제공해야 한다.
 예
 왜 주어진 문서가 검색 되었는지.
 검색이 하나도 안되거나 너무 많은 검색에 대하여 다음에 어떻게 해야하는지를 제시
- **방대한 결과 내지는 소수의 결과를 처리할 수 있어야 한다.**
 질의 “정보 검색 시스템”에 대한 일반 정보검색 엔진의 결과 수 (1997년 6월 4일 실행)
 웹글라이더 : 3,268 건
 정보탐정 : 1,809 건 (1997.6.11. : 2,056 건)
 애니서치 : 1,522 건
 유니 파인더 : 1,522 건
 줌(ZOOM) : 1211 건
 와카노 : 1507 건
 코시크(KOR-SEEK) : 500 건
 알타비스타(AltaVista) : 200 건
 아이보트(iBot) : 183 건
 미스 다찾니 : 46 건
 % 애니서치, 정보탐정, 심마니, 코시크, 알타비스타의 메타 검색엔진
 % 검색엔진 각각 10위내 문서 검색 가능
 심마니 : 35 건
 넷서치 : 6 건
 짚(ZIP) : 4 건
- **표준 결과 포맷(format)을 지원해야 한다.** 즉, 일반 검색엔진의 질의 및 검색 결과를 수용할 수 있어야 한다. 일반 검색엔진으로의 질의 리스트를 call 할 수 있고 또한 일반 검색엔진의 결과 정보를 수용할 수 있어야 한다.
 심마니 : 순위, 제목, 정확성, URL, 요약
 웹 글라이더 : 순위, 제목, 중요도, URL, 요약
 와카노 : 순위, 제목, 요약, URL, 크기
 유니 파인더 : 제목, 중요도, 요약, URL, 크기
 넷서치 : 제목, URL, 요약
 코시크 : 제목, 요약, URL
 애니서치 : 제목, 중요도, 요약, URL, 크기
 정보탐정 : 순위, 중요도, 제목, URL, 요약
 미스 다찾니 : 순위, 제목, 요약, URL, (검색엔진 정보)
 알타비스타 : 제목, 요약, URL, 크기, 날짜
- **검색 결과를 볼 수 있는 편리한 인터페이스여야 한다.**
 현 검색엔진의 검색 결과 :
 쪽당 검색 갯수 지정 : 가, 불가
 1,2,3, 4, ... 표시 : 일반적으로 10 개 단위
 <이전> 1, 2, 3,... <다음> 제공
- **검색 결과의 순서화가 있어야 한다.**
 문서의 중요도에 따른 순서화가 이루어져야 한다.
- **검색어의 기여도를 알 수가 있어야 한다**
 검색어의 참여도를 보여 줌으로써 쉽게 검색어의 기여도를 알 수 있어야 한다.
- **다양한 view로 검색 결과를 볼 수 있어야 한다.**
- **검색 실행**

명시적으로 검색을 실행하는 경우에 검색을 하며 내제적으로 검색어가 변할때마다 계속하여 보여 주고 수정할 수 있어야 한다.

- **결과의 수정(refinement)**
 검색 결과에 대한 사용자 피드백을 할 수 있어야 한다.
 긍정적 적합 판단(positive relevance judgement)이 현실의(working query)에 영향을 미칠 때마다 새로운 용어들을 보여주고 또한 존재하는 용어들의 중요도를 조정해서 새롭게 재 순서화 함에 의해 즉각적인 피드백(feedback)을 수행해야 한다.
 ⇒ 검색 결과에 대한 약간의 점증적 변화(small incremental change)는 검색자로 하여금 더욱 미묘한 수정을 할 수 있도록 함으로써, 추출된 용어들을 어떤 문서들과 연관(associate)하는 것을 더 쉽게 한다.
- **가능한 한 색(color)나 크기(size)를 다양하게 유지하여 한눈에 알아볼 수 있도록 도움을 주어야 한다.**
- **보다 미세한 검색어의 가치치(중요도)를 조정할 수 있어야 한다.**
- **검색 문서의 갯수를 지정할 수 있어야 한다.**

3. 결과 렌즈 설계 고려사항

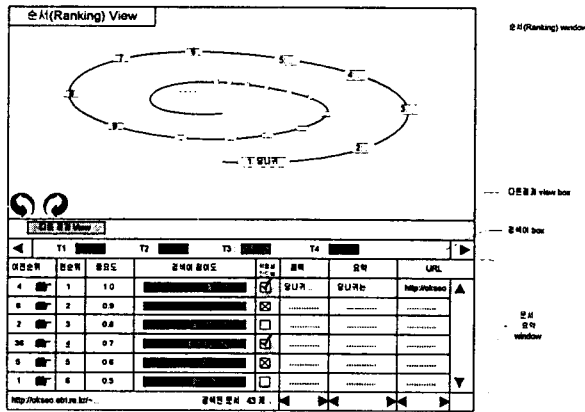
검색 결과의 정확성을 향상하거나 검색 결과를 효과적으로 보여줌으로써 사용자가 검색하고자하는 정보를 제공할 수 있는 방법(도구)이 있어야 하겠다. 쉽게 정보를 얻을 수 있도록 다음과 같은 설계 고려 사항을 만족해야 한다.

검색된 결과에 대하여 다양한 결과 viewing을 한다.

- 검색 결과의 다양한 방법에 의한 viewing
- informative하고 편리한 사용자인터페이스 view (Information Visualization)
- 다양한 결과 view에 따른 재 순서화
- 적합성 피드백에 의한 검색 수정
- 문서의 선택 후 색인기를 통하거나 또는 검색어 확장
- 사용자 인터페이스는 단순 명료하고 편리해야 한다.
- 메뉴나 아이콘등의 이름은 일반 사용자도 쉽게 이해할 수 있도록 전문 용어가 아닌 의미가 동일한 일반 용어를 사용해야 한다.
- 검색 결과들은 가능한 빨리 사용자에게 보여줘야 한다. “진행 중”과 같은 지시자는 사용자에게 어떤 것이 현재 일어나고 있다는 것을 보여주기 때문에 검색의 효율성에 있어서 신뢰(confidence)를 향상시키는데 도움을 준다.
- 사용자에게 대한 인식의 부담(cognitive load)을 줄인다.
- 시스템의 기능을 시각화(visible)하게 한다.
- 전체 화면에 여러 영역을 나눔으로써 한번에 모든 필요한 정보를 시각화할 수 있게 하여야 한다.
- 현 상태의 정보를 곳곳에서 알 수 있어야 한다.
- 여타의 검색엔진에서도 사용할 수 있도록 플러그-인(plug-in)으로 구현 되어야 한다.
- 확장성을 지원해야 한다.

4. 결과 렌즈 화면 설계

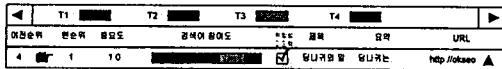
결과 렌즈 구성은 크게 검색 문서들의 순위를 볼 수 있는 '순서(rank) 윈도우', 검색 문서의 순위, 중요도, 검색어 참여도, 제목등을 볼 수 있는 '문서 요약 윈도우(summary 윈도우)', 여러가지 다른 결과 view를 볼 수 있는 '다른결과 view box', 검색어를 나타내는 '검색어 box' 그리고 검색 문서 갯수등을 나타내는 '상태정보'로 나뉘어진다. 화면 1은 결과에 대한 viewing 방법을 나타낸다. Viewing 윈도우와 문서 요약 윈도우는 서로의 정보 변화를 1:1 대응으로 반영한다. 예를 들면, 가장 순위가 높은 위치(꼭 1위의 문서가 아니라 1위의 위치에 해당)를 기준으로 1:1로 반영된 화면이다.



화면 1 : 결과 렌즈 (Results Lens) 기본 화면

순서 View는 기본적으로 나선형 모델을 따르며 검색 결과로 순위가 높은 문서들이 바깥쪽에 위치하며 회전 버튼의 클릭에 의해서 회전되며 회전에 따라 자동적으로 요약 윈도우의 내용이 수정(update) 된다. 각 노드를 double-click하면 문서의 내용을 브라우저(browser) 윈도우를 통해 볼 수 있다.

다음 그림은 문서요약 윈도우의 일부이다.



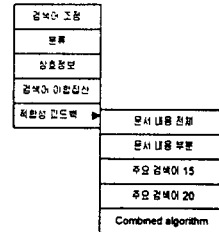
결과 문서(검색된 문서)의 이전 순위와 method를 달리하여 검색한 현순위의 변화의 이동을 손가락 아이콘을 중심으로 나타내고 있다. 현재까지 순위가 변화한(높은 쪽으로 5단계 이상) 문서의 현순위를 coloring(Red, 붉은색)하여 강조한다.

결과 문서의 내장된 결과식에 의해 산출된 중요도는 질의에 대해서 어느정도의 정확률을 갖는 문서인지 검색시스템이 계산하여 보여주는 결과이다. 검색어 참여도는 검색된 문서에 대해 각 검색어(term)들에 대한 기여도(참여도)를 그래프로 나타낸다. 각 검색어는 색깔로써 구분되어진다.

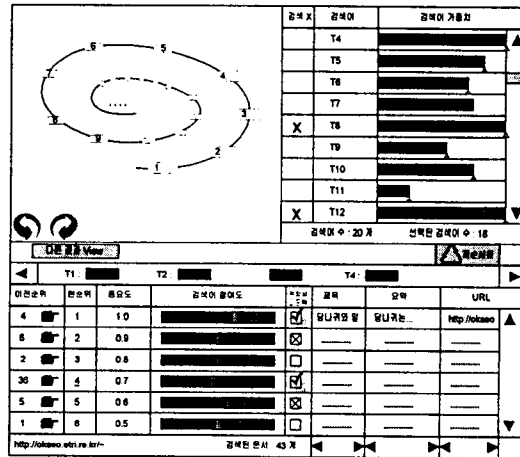
적합성 피드백은 결과 문서에 대해 사용자의 판단을 check한다. 적합 부적합 모르겠음 이다. 나머지 열은 각각 검색된 문서의 제목, 요약, URL 위치이다. 검색된 문서의 해당

행에 대해서 double-click을 하면 문서의 내용을 가져와서 브라우저 윈도우를 통해 보여준다.

'다른결과 View' 버튼은 검색의 다른 결과 view의 풀다운 메뉴로써 다음과 같으며 검색어 조정 메뉴부터 차례로 설명한다.



화면 2는 검색어 조정 메뉴를 선택했을 때 나타나는 윈도우이다.



화면 2 : 검색어 조정 선택 화면

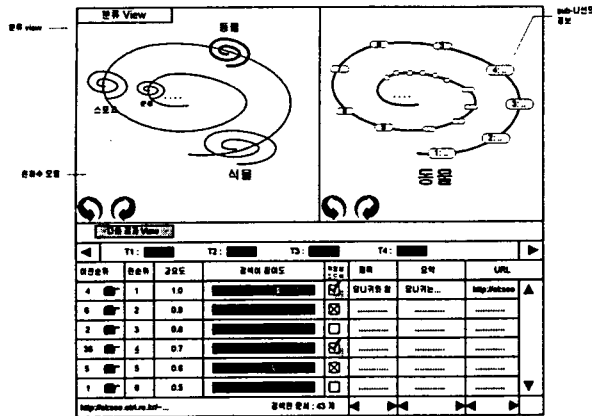
검색어 박스는 사용된 검색어에 가중치를 부여하는 인터페이스이다. 좌측의 '검색 X' 항목에 표시를 받은 검색어는 검색시에 이 검색어를 포함하는 문서는 검색 결과로 '절대 나오지 말라'는 의미를 가진다. 검색어 가중치는 0에서 1까지의 범위를 가지며 '슬라이드 바'로 조정된다. 질의 검색어의 가중치는 검색시스템의 내장된 알고리즘의 계산에 사용된다.

재순서화 버튼은 1차 검색된 결과에 대하여 새로 조정된 검색어의 가중치를 가지고 다시 재순서화하는 실행 버튼을 나타낸다.

화면 3은 메뉴의 분류를 선택했을 때 나타나는 화면이다. 좌측의 분류 view는 시스템의 분류 알고리즘에 의해 분야별로 분류된 검색 문서들의 집합의 형태를 은하수 모델로써 나타내고 있다.

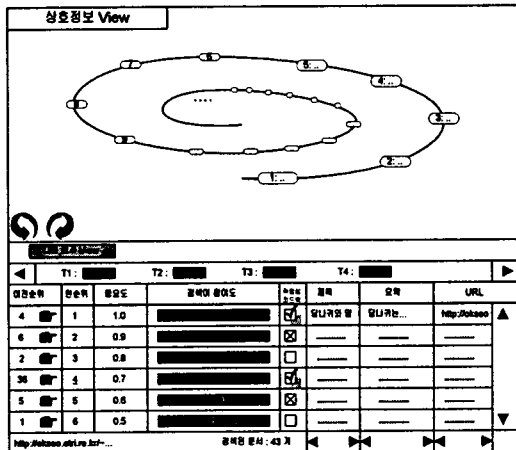
윈도우의 좌측 하단 회전버튼은 전체 은하수 나선을 회전시킨다. 마우스를 이동함에 따라 커서에서 가장 가까운 sub-나선의 색깔이 바뀌고 은하수 윈도우에서 sub-나선을 마우스로 선택하면 오른쪽 윈도우에 sub-나선의 정보를 보여준다. Default

sub-나선은 은하수 모델의 가장 바깥쪽 sub-나선이다. 문서 요약 윈도우에는 선택된(예: 동물), 나선의 내용을 나타낸다.



화면 3: 분류 선택 화면

화면 4는 상호 정보 메뉴를 선택했을 때 나타나는 화면이다.



화면 4: 상호정보 선택 화면

상호 정보 view는 검색 결과에 대해 검색 시스템이 상호 정보를 적용하여 자동 재순서화한 검색 결과를 보여준다.

검색어 이합집산 메뉴는 검색어 집합(term set)으로 정의하여 검색어 집합 별 재순서화를 지원한다. 화면은 분류 선택 view와 유사하다.

적합성 피드백의 문서 내용 전체 메뉴는 사용자가 문서의 내용을 살펴볼 수 있으며 개개의 문서에 대해서 적합성 피드백을 수행한다. 적합은 현재 문서가 적합하다는 적합성 피드백, 부적합은 현재 문서가 적합하지 않다는 적합성 피드백, 모르겠음은 현재 문서가 적합한지 적합하지 않은지 모르겠다는 적합성 피드백을 의미한다.

사용자가 결정한 결과는 요약 윈도우의 적합성 피드백 check box에 자동적으로 마킹된다.(V : 적합, X : 부적합, No marking : 모르겠음)

적합성 피드백의 문서 내용 부분 메뉴는 적합한 문서의 부분을 마우스로 드래그(drag)한다. 적합성 피드백을 사용자가 선택한 문서 내용의 부분(passage)에 대해서 수행한다.

적합성 피드백의 주요 검색어 15, 주요 검색어 20, Combined algorithm 메뉴는 사용자가 적합성 피드백을 각 문서에 대해서 결정을 내리면 검색시스템의 내장된 주요 검색어 15, 주요 검색어 20, combined algorithm은 사용자의 적합성 피드백에 의한 질의문 자동 수정이 수행되며 문서의 재순서화를 제공한다.

적합성 피드백 View는 '문서 내용 전체 / 문서 내용 부분 / 주요검색어 15 / 주요검색어 20 / 알고리즘 조합'의 각 결과를 보여주며 내용은 결과 렌즈 순서 View와 동일하다.

5. 결론

본 논문에서는 현재의 정보검색 시스템 및 웹(web)상의 검색엔진들에 대한 개선되어야 할 요구사항 및 고려사항들을 정의하였다. 또한 정보검색 시스템의 문제점을 지적하고 방대한 결과를 여러가지 다양한 view로 볼 수 있고 사용자의 피드백에 의해 정보를 재 검색 함으로서 검색하고자 하는 문서를 보다 정확하게 검색할 수 있도록 도와주는 결과 렌즈 (Results Lens: RL)를 제안하였다. 구체적으로 화면을 중심으로 기능들에 대한 규격을 정의하였다.

검색어 조정, 분류에 의한 조정, 상호정보, 검색어 이합집산, 적합성 피드백을 통한 다양한 view로의 결과를 refine 할 수 있다. 일반 검색엔진 보다는 메타 검색엔진이 유용한 경우가 많은데 이들의 검색 결과 단점을 보완할 수 있는 하나의 방법이다. 검색 결과의 정확성에 향상을 꾀할수 있다.

현재 웹을 기반으로 하는 검색엔진 및 정보검색 시스템에 적용할 수 있도록하기 위하여 넷스케이프 (Netscape)를 기반으로하는 플러그 인 (Plug-in)으로 구현 중 이다. 클라이언트와 서버사이의 검색을 의뢰하고 검색결과를 주고 받기위해서 기본적으로 서버는 IIS(Internet Information Server)를 이용하고 그 결과를 넷스케이프 기반 클라이언트의 플러그인에서 받아 기능을 수행한다.

참고 문헌

- [1] 강현규, 장호욱, 이승률, 박세영, "옥서에서의 표제어와 자연어 검색의 설계 및 구현", 1994년도 한국정보과학회 가을 학술발표 논문집(A), pp.633-636, 1994.
- [2] 이수현, 박동인, "한글 인터넷 정보검색 시스템의 기술방향", 제8회 한글 및 한국어 정보처리 학술대회, pp. 59-62, 1996.
- [3] 강현규, 장호욱, 전미선, 박세영, "인터넷 기반 멀티미디어 정보 검색 시스템 : 옥서 '95의 색인 및 검색", 제8회 한글 및 한국어 정보처리 학술대회, pp. 85-92, 1996.
- [4] 노현철, 권혜진, 이근배, 이종혁, "한국어 담화 분석을 이용한 Session 기반의 웹 검색 시스템: AIR-Web", 1996년도 한국정보과학회 가을 학술발표논문집(A), pp.525-528, 1996.

- [5] 박세진, 강상배, 권혁철, “Relevance Feedback을 이용한 정보검색 시스템의 검색 효율 향상”, HCI '97 학술대회 발표논문집, pp.3-8, 1997.
- [6] 김평, 주원균, 맹성현, “디지털도서관 클라이언트를 위한 사용자 및 통신 인터페이스 개발”, 1997년도 한국정보과학회 봄 학술발표논문집(B), pp.547-550, 1997.
- [7] Hyun-Kyu Kang, Key-Sun Choi, “Two-level Document Ranking Using Mutual Information in Natural Language Information Retrieval,” *Information Processing & Management*, Vol. 33, No.3, pp. 289-306, 1997.
- [8] J.H. Lee, “Combining Multiple Evidence from Different Properties of Weighting Schemes,” SIGIR '95, Seattle, WA, USA, pp. 180 ~ 188, 1995.
- [9] A. Pollock and A. Hockley, “What's Wrong with Internet Searching,” <http://www.dlib.org/dlib/march97/bt/03pollock.html>, D-Lib Magazine, March 1997.
- [10] L. T. Nowell, R. K. France, et. al, “Visualizing Search Results: Some Alternatives To Query-Document Similarity,” SIGIR '96, Zurich, Switzerland, pp. 67 ~ 75, August 1996.
- [11] D. Cutting, D. Karger, and J. Pederson, “Constant Interaction-Time Scatter/Gather Browsing of Very Large Document Collections,” SIGIR '93, Pittsburgh, PA, USA, pp. 126 ~ 134, 1993.
- [12] M. Hearst, “TileBars: Visualization of Term Distribution Information in Full Text Information Access,” CHI '95, Denver, Colorado, USA, pp. 67 ~ 73, May 1995.
- [13] M. Hemmje, C. Kunkel and A. Willett, “LyberWorld - A Visualization User Interface Supporting Fulltext Retrieval,” SIGIR '94, Dublin, Ireland, pp. 249 ~ 259, 1994.
- [14] M. Hearst, G. Kopec, and D. Brotsky, “Research in Support of Digital Libraries at Xerox PARC,” <http://www.dlib.org/dlib/june96/hearst/06hearst.html>, D-Lib Magazine, June 1996.
- [15] J. Allan, “Relevance Feedback With Too Much Data,” SIGIR '95, Seattle, WA, USA, pp. 337 ~ 343, 1995.
- [16] G. Salton, *Automatic Text Processing*, Addison-Wesley Publishing Co., Reading, MA, 1989.
- [17] ZIP 웹 디렉토리, <http://www.zip.org/>
- [18] 웹글라이더(WebGlider), <http://www.infoglider.com/>
- [19] 정보탐정, <http://aistar.kotel.co.kr/>
- [20] 애니서치, <http://www.anysearch.com/>
- [21] 미스 다찾니, <http://www.mochanni.com/>
- [22] 까치네, <http://kachi.taegu.ac.kr>
- [23] 코시크, <http://kor-seek.chungnam.ac.kr/cgi-bin/korea>
- [24] 심마니, <http://simmany.hnc.net/>
- [25] 유니파인더, <http://www.unifinder.com/>
- [26] 알타비스타, <http://www.altavista.digital.com/>
- [27] 와카노, <http://www.keimyung.ac.kr/wakano/>
- [28] 줌, <http://zoom.cyso.net/>