

碩士學位論文

# 에이전트 기술을 이용한 AQL시스템



濟州大學校 大學院  
제주대학교 중앙도서관  
情報工學科 LIBRARY

田 炳 輝

110 507

2000年 12月

# 에이전트 기술을 이용한 AQL시스템

指導教授 郭 鎬 榮

田 炳 輝

이 論文을 工學 碩士學位 論文으로 提出함



田炳輝의 工學 碩士學位 論文을 認准함

審査委員長

안 기 중

審査委員

곽 호 영

審査委員

변 상 용



濟州大學校 大學院

2000年 12月

# Automatic Question Link System Using Agent Technology

Byung-Hwi Jun

(Supervised by professor Ho-Young Kwak)



제주대학교 중앙도서관  
JEJU NATIONAL UNIVERSITY LIBRARY

A THESIS SUBMITTED IN PARTIAL  
FULFILLMENT OF THE REQUIREMENTS FOR  
THE DEGREE OF MASTER OF ENGINEERING

2000. 12.

Department of Information Engineering  
GRADUATE SCHOOL  
CHEJU NATIONAL UNIVERSITY

# 목 차

SUMMARY .....	1
I. 서 론 .....	1
II. 관련연구 .....	3
1. 에이전트(AGENT) .....	3
2. 사용자 경향 분석과정 .....	8
3. 정보 검색에 관한 기존 연구 .....	11
III. AQL시스템 설계 .....	18
1. 기존 시스템의 문제점 .....	18
2. AQL시스템의 제안 .....	19
3. AQL시스템 구조설계 .....	23
4. 시스템 동작과정 .....	25
IV. 구현 및 결과 고찰 .....	26
1. 시스템 모듈 구현 .....	26
2. 시스템 검증 및 결과 .....	36
V. 결론 및 향후 과제 .....	41
참고 문헌 .....	43

## SUMMARY

Many information in the internet service causes the increase in the requirement of the efficient information search. Many systems are developed to that purpose, such as search engine, question and answer system, and so on. Conventional search systems have many drawbacks, which cause the low efficiency of information search, the overhead on the server system, and the increase of the network traffic etc.

In the paper AQL(Automatic Question Link) system is proposed to support the automatic information search. The system uses the agent technology for efficient information search. The system processes the query statement and manages the user information and the search history DB for the next search stage. The preprocessed keywords are used to search information on the conventional search engine and the user gets the result information after postprocessing stage.

The prototype of the suggested system is implemented and operates correctly. The system helps the reliability of the search information and the efficient use of network traffic. The next study will be the natural processing step on the QPM(Question Processing Module).

# I. 서 론

최근 인터넷 사용의 대중화와 더불어 정보의 양도 역시 급격하게 증가하고 있으며, 이러한 변화는 사용자의 요구에 맞는 양질의 정보를 제공할 수 있는 서비스를 더욱 필요로 하고 있다.

네트워크상의 정보를 제공하기 위한 서비스로 일반적으로 많이 이용되는 것은 정보검색 시스템이다. 정보검색시스템은 검색대상에 따라서 서버 중심, 프록시 서버 중심, 그리고 사용자 중심의 시스템으로 분류할 수 있다(김 등 1996). 이러한 검색시스템을 통하여 사용자는 원하는 정보를 세부적인 정보의 보유위치에 대한 정보 없이도 검색이 가능한 반면에 정보의 양과 더불어 양질의 정보를 접근하는데는 한계가 있다. 이러한 한계를 극복하기 위하여 많이 이용되는 질의응답시스템인 경우, 양질의 정보를 접할 수 있다는 장점이 있으나 사용자가 요구한 질문에 대한 정보를 얻기까지는 상당한 시간이 요구된다는 단점을 함께 내포하고 있다.

본 논문에서 기술하는 시스템은 기존의 정보검색 시스템과 질의응답 시스템의 장, 단점을 적절히 결합하여 사용자가 요구하는 가능한 한 빠른 시간 내에 양질의 정보를 제공하기 위한 목적으로 설계하였다. 이를 위하여 기존의 다양한 응용분야에서 활용되는 에이전트 기법을 활용한다.

에이전트는 사람을 대신하여 문제를 해결하기 위해 작업을 처리하는 프로그램의 일종이다(Franklin 등 1996). 제안된 시스템은 기존 시스템들의 문제를 해결하기 위해 에이전트를 이용하여 각 검색엔진이 가진 정보를 활용할 수 있도록 한다. 이를 위해서 크게 사용자 등록정보, 질의처리, 키워드 생성 에이전트, 그리고 자바기반의 검색 에이전트로 구성된 전체 시스템을 설계하였다. 이 과정에서 사용자의 정보 및 이전단계의

검색에 대한 히스토리를 활용하여 정보검색에 요구되는 시간을 줄일 수 있도록 하며, 부가적으로 네트워크의 부하를 줄일 수 있는 전체 시스템을 구현하였다.

본 논문은 다음과 같이 구성된다. II장에서는 에이전트 기법 및 검색시스템, 질의응답 시스템과 연관된 기존의 연구결과 및 개발사례에 대해 살펴보고, III장에서는 에이전트를 이용한 AQLS 모델 및 전체 시스템의 동작과정에 대해 설명하다. 4장에서는 시스템의 세부설계 및 구현을 하며, 5장에서 결론 및 향후 연구방향에 대해 서술한다.



## II. 관련연구

본 장에서는 기존에 수행되었던 에이전트를 이용한 정보검색에 관련된 연구 및 사례에 대해 기술한다. 먼저 에이전트의 전반에 대해 설명하고, 지금까지 연구되어 온 검색방법에 대해 알아보고, 그들의 장단점에 대해 알아본다.

### 1. 에이전트 (AGENT)



#### 1) 에이전트의 특성

에이전트 개념은 수많은 분야의 이론들에 의해 영향을 받았으나, 실질적인 기원은 인공지능, 특히 분산 인공지능에서 찾을 수 있다. 이 개념은 1980년대 초부터 사용되기 시작했는데, 'Object that think', 즉 자신만의 지능(지식+추론)을 가지고 행동하는 자동적인 소프트웨어 단위를 뜻하는 목적으로 쓰였다. 에이전트는 다음의 성질들에 의해 구분된다(Magedanz 등 1996).

#### (1) 지능(Intelligence)

에이전트는 지능을 가지기 위해, 미리 정의된 규칙에서부터 자체 학습 인공지능 추론기계에 이르기까지 수많은 방법을 사용하여 다양한 수준의 지능을 가지게 된다.



## (2)비동기적 수행(Asynchronous Operation)

에이전트는 자신의 사용자나 다른 에이전트들과는 무관하게 자신의 일을 수행한다. 에이전트는 수많은 시스템 자원과 다른 에이전트들과 통신하며 일을 수행하면서, 사용자와 독립적으로 동작이 가능하다.

## (3)통신(Communication)

수행 도중인 에이전트들은 많은 시스템 자원들, 그리고 사용자들과 통신한다. 사용자들과 직접적으로 접촉하는 에이전트들은 때때로 개인비서나 인터페이스 에이전트로 불린다. 에이전트의 관점에서 볼 때, 자원들은 지역적이거나 원거리시스템에 있다.

## (4)상호협조(Cooperation)

에이전트 시스템이 에이전트들 간의 상호협조를 가능하게 하는 것을 뜻한다. 이것은 단지 클라이언트/서버(Client/Server) 수준에서부터, 스피치액트 이론에서부터 도출된 여러 가지 인공지능적인 방법에 기초한 상호 협조 수준까지 다양하다. 이러한 상호 협조는 지식 정보의 교환을 필요로 하게 되고, 이것이 다중 에이전트 시스템의 기초 조건이 된다.

## (5)이동(Mobility)

어떤 특정한 일을 수행하기 위해, 에이전트 자신이 네트워크를 타고 원거리의 지점으로 이동하여 그 곳의 환경에서 실행될 수 있다. 일반적으로, 에이전트의 이동성은 프로그램 자신과 데이터가 원거리의 시스템으로 이동되어 실행되는 원거리 실행과 에이전트 자신의 수행을 중지하고, 네트워크상의 다른 지점으로 이동 후 수행을 재개하는 이주의 두 가지로 나누고 있다.

## 2) 에이전트의 분류

앞 절에서 설명한 에이전트의 특성들을 이용하여 다양한 에이전트 기술들을 구분하면 Table 1과 같이 분류할 수 있다(Magedanz 등 1996).

Table 1 The Feature of Agents

Single-Agent Systems		Multi-Agent Systems	
Local Agents	Networked Agents	DAI-based Agents	Mobile Agents
Personal Assistants Advisory Assistants	Personal Assistants Smart Mailboxes	Distributed Problem Solving	Telecommunication Pers Communication
Meeting Schedules	Retrieval Agents Process Automation		Netw. Management Electronic Market

단일 에이전트 시스템에서는 에이전트가 사용자나 어떤 프로세스를 위해 일을 수행한다. 일을 수행하는 동안 에이전트는 지역 또는 원격지의 시스템 자원 및 사용자와의 통신을 수행할 수 있으나 다른 에이전트들과의 통신은 불가능하다. 반면, 다중 에이전트 시스템에서는 자신의 목적을 위해 다른 에이전트들과 교신이 가능하다.

좀 더 세부적으로 에이전트의 특성에 따라 지역 에이전트, 네트워크 에이전트, 분산 인공지능 기반 에이전트, 그리고 이동 에이전트로 구분할 수 있다.

### (1) 지역 에이전트

지역 에이전트는 단지 지역 자원에만 접근이 가능하다. 이들은 통상적으로 사용자의 일상적인 일을 도와주는 조연자 에이전트나 개인비서처럼 동작한다. 이들의 주된 목표는 사용자와 에이전트의

상호교류에 있으므로 이런 종류의 에이전트를 인텔리전트 인터페이스(Intelligent Interface) 또는 인터페이스 에이전트(Interface Agent)라고 부른다.

이들은 사용자를 위해 일을 대신 처리할 수도 있고, 비동기적으로 행동하면서 사용자를 훈련시키고 가르칠 수도 있다. 지역 에이전트가 도움을 줄 수 있는 분야는 지역정보 검색 및 필터링, 지역 메일관리, 미팅계획관리등 외에도 가상적으로는 무제한적이라고 할 수 있다. 그러나 실제적으로 위와 같은 수준의 에이전트는 아직 만들어지지 못하고 있다.

## (2)네트워크 에이전트

지역 에이전트에 반해, 네트워크 에이전트는 지역자원뿐만 아니라 원격자원에도 접근할 수 있고, 그러기 위해서 네트워크 내부 구조와 가능한 서비스 등에 대해 좀더 자세한 정보를 가지고 있다. 네트워크 에이전트와 다중 에이전트 시스템과의 주된 차이는, 이들은 서로 상호 협조할 수 없다는 점이다. 지금까지 개발된 네트워크 에이전트는 대개 개인 비서와 비슷한 기능을 한다. 우편함이나 검색 엔진과 같은 것이 가장 널리 알려진 이 부류의 에이전트들이다. 이들은 사용자에게 지능적인 인터페이스를 제공할 뿐만 아니라, 네트워크 상의 수많은 서비스들을 사용한다. 검색엔진은 사용자의 기호에 맞추어 네트워크 상의 유용한 정보들을 수집한다. 이러한 에이전트들은 'Knowbot' 또는 'Softbot'으로 불리며, 수집된 정보들을 사용자에게 보여줄 때, 네트워크 내부의 복잡한 구조나 프로토콜(Protocol)등의 하부구조에 대해서는 사용자에게 알리지 않는다. 네트워크의 크기가 점점 커짐에 따라 네트워크 상의 가용 정보들을 관리하는 이러한 종류의 네트워크 에이전트들에 대한 수요가 점점 늘어나고 있다.

### (3) 분산 인공지능 기반 에이전트

이 부류의 에이전트 시스템에서의 주된 관심사는 지능을 가진 에이전트들이 어떻게 그들의 지식, 목표기술, 그리고 계획 등을 서로 나누어서 상호 협력하는가에 있다. 이들은 다양한 분야에 적용될 수 있는데, 예를 들면 분산 자동차 모니터링, 자연어 처리, 전송 계획 작성, 그리고 특정한 망 관리 등에 적용될 수 있다.

이 부류의 에이전트는 법칙 기반 시스템(Rule-Based systems)과 상황이나 실례 기초추론과 같은 인공지능 기술을 이용하여 개발된다. 적용되는 지식표현방법이나 추론메커니즘은 기본적으로 지식 기반 시스템의 그것과 같다. 또한, 이 부류의 에이전트들은 사용자, 시스템 자원, 그리고 나아가서 다른 에이전트들과 통신할 수 있다.

### (4) 이동 에이전트(Mobile Agents)

이동 에이전트는 다른 에이전트와는 달리 자신이 직접 네트워크를 통해 이동하면서 실행되는 에이전트이다(Magedanz 등 1996, Lawton). 이 부류의 에이전트들은 Safe-Tcl SUN Microsystem사에서 개발된 Java언어, General Magic사에서 개발된 Telescript와 같은 스크립트 언어(Script Language)로 개발된다. Java나 Safe-Tcl은 주로 이동 에이전트의 이동성에 중점을 두어 개발을 한 반면, Telescript는 전자 상거래를 위한 플랫폼 위에서 응용될 수 있는 이동 에이전트의 구현에 중점을 두었으므로 에이전트간의 통신이 가장 중요한 구현 목표이다. 그러나 실제로 복잡한 에이전트간의 상호 협조를 구현하는 데는 어려움이 있다. 그러나 역시 가장 중요한 이동 에이전트의 특성은 이동성이므로, 모든 언어들이 이 성질을 가장 기본적으로 구현하고 있다.

이동 에이전트의 응용분야는 대단히 광범위한 것으로 생각되지만, 아직까지는 용어도 제대로 정의되지 않았을 정도로 새로운 분

야이다. 지금까지 언급되어 온 분야는 정보검색(Information Retrieval), 네트워크 관리(Network Management), 전자 상거래(Electronic Market), 이동 컴퓨터(Mobile Computing)등이 있다.

이동 에이전트의 구조는 코드(Code), 실행 데이터(Data), 속성(Attribute) 등으로 이루어져있다(Lawton). 지금까지 여러 종류의 에이전트에 대해 알아본 것을 하나의 표로 나타내면 Table 2와 같이 나타낼 수 있다(Magedanz 등 1996).

Table 2 The differences among Agents

Agent Attribute	Local Agents	Networked Agents	DAI-based Agents	Mobile Agents
Intelligence	○	○	○	○
Async. Operation	○	○	○	○
Communication		○	○	○
Cooperation			○	○
Mobility				○

## 2. 사용자 경향 분석과정

### 1)사용자의 경향 파악

인터넷을 사용하는 사용자의 Bookmark나, Cache, Log파일, History파일, 그리고 Cookie파일들을 분석함으로써 해당 사용자의 관심과 사용성향을 알아낼 수 있으며, 사용자 등록 시 사용자의 해당직종이나 관심분야를 기록함으로써 카테고리를 줄여서 좀더 사용

자에게 필요한 정보를 제공할 수 있다. 현재 많은 인터넷 쇼핑몰 사이트에서는 일대일 고객마케팅을 위해 사용자의 웹 로그(Web log)파일이나, 북마크(Bookmark) 등의 파일을 분석해 고객의 소비성향을 파악하고 있다.

## 2) 자연언어 처리

사용자가 질의를 할 경우 질의에 대한 키워드를 생성하기 위해 해당질의에 대해 자연어처리과정을 거쳐야 한다. 자연언어를 처리하는 컴퓨터시스템에 관한 연구는 1950년대에 시작되어 지금까지 계속되고 있으며 시기별로 응용분야의 변화를 보이고 있다. 1950년대와 1960년대 초기까지의 연구대상은 주로 기계번역이었으며, 1960년대 중기부터 1970년대 중기까지는 질문응답시스템(LUNAR, LSP, PROTOSYNTHESIS 등)에 관한 연구가 활발하였다. 1970년대 후기에 들어서는 실용화된 데이터베이스시스템에 자연언어로 접근할 수 있는 진단시스템들(REQUEST, INTELLECT, PLANES 등)이 개발되었다. 이외에 자연언어 텍스트를 언어학적 기법에 의해 처리하는 자동색인에 관한 연구가 1970년대 이후 지속되고 있으며, 최근에는 지식베이스를 이용하는 지식기반 색인시스템에 관한 연구가 수행되었다. 1980년대에 들어서는 특히 자연언어 인터페이스를 제공하는 전문가시스템과 지능형 정보검색시스템에 관한 연구가 활발히 수행되고 있다. 국내에서도 한국어 처리에 관한 연구가 꾸준히 진행되고 있으며, 특히 질문응답 시스템과 데이터베이스 진단시스템, 자동색인 시스템 등에 자연언어 처리기술을 응용한 예를 많이 찾아볼 수 있다.

자연언어 처리시스템들은 초기에는 주로 특정한 단어(구)의 도식대조 방식에 의해 문장을 분석하는 초보적인 시스템들이었으나, 후에 ATN이나 변형문법 등을 이용하면서 본격적인 구문분석을 시도

하였으며, 구문분석과 의미분석을 병행한 시스템들도 나타났다. 실제로 자연언어 처리시스템을 설계할 때 기본적으로 고려할 점으로는 구문분석의 수준(완전한 분석, 부분적 분석)과 사용할 문법의 선택이 있다. 또한 의미분석의 수준도 고려할 수 있는데, 불완전한 구문분석을 보조하는 수준으로서 사용하는 경우와, 완전한 구문분석을 수행하되 의미자질을 이용하여 보다 완전한 처리를 목적으로 하는 경우에 각각 다른 수준의 의미분석을 수행하는 시스템이라고 자연언어의 모든 현상을 완벽하게 처리하기는 힘들다. 따라서 처리 가능한 구문의 범위가 시스템에 따라 다르게 설정되며, 실제로 개발된 시스템들은 제한된 형식의 문장만을 처리하는 시스템에서부터 내포문의 처리, 복잡한 명사구의 처리, 관계대명사나 생략문의 처리 등이 가능한 시스템에 이르기까지 다양하다. 또한 대부분의 시스템이 소규모의 어휘사전과 의미사전이나 의미해석규칙 등을 이용하여 특정한 주제영역만을 처리대상으로 하고 있기 때문에 시스템의 범용성 내지는 이식성이 떨어지는 문제가 제기된다. 현재로는 자연언어 처리에 있어서 분석대상이 되는 주제영역의 폭과 언어학적 분석의 깊이는 상호 조정관계에 있는 것으로 보인다. 다시 말해 주제영역의 폭이 넓으면 심층적이고 완벽한 언어학적분석은 기대하기 어렵다는 것이다.

빅커리(Vickery)등이 제안한 자연언어 인터페이스 시스템이 갖춰야 할 기능 가운데 특히 지능형 정보검색 시스템에서 고려해야 할 기능들을 열거하면 다음과 같다(정 1990).

- 생략문, 숙어, 문법적 오류, 단편적 문장 등의 처리가 가능한 융통성 있는 구문분석 기능
- 대화의 정확한 전달과 해석이 가능한 강력한 커뮤니케이션 기능
- 시스템이 ‘무엇을’ ‘왜’ 할 수 있는지, 했는지, 하고자 하는지에

관한 설명기능

- 설명을 통해 대상물을 식별할 수 있는 기능
- 이용자의 수준을 진단하고 이용자모형을 구축하는 기능
- 이용자가 입력오류를 수정할 수 있는 기능
- 이용자 편의시스템으로서 기능
- 이용자를 위한 지도 기능
- 적절한 수준의 시스템 응답시간

### 3. 정보 검색에 관한 기존 연구

정보검색이란 정보자료의 내용을 분석, 가공하여 구축한 정보 데이터베이스로부터 이용자의 정보요구에 적합한 정보를 탐색하여 찾아내는 일련의 과정을 의미한다. 여기서는 정보검색을 위해 구현된 기존의 여러 가지 시스템에 관해 알아본다. 기존의 시스템은 크게 세 가지 정도로 나누어 볼 수 있다. 첫 번째, 기존의 검색 엔진들과 같은 서버중심의 시스템이 있다. 두 번째로는 프록시 서버를 이용하는 시스템이다. 세 번째는 사용자 중심의 에이전트를 이용하는 시스템이다. 기타에서는 이동 에이전트를 이용하여 구현된 IBM사의 Aglet Workbench로 구현된 검색 프로그램을 소개한다.

#### 1) 서버 중심의 검색 시스템

서버 중심 시스템은 일반적인 예로써 검색 엔진을 들 수 있다. 하나의 검색 엔진이 모든 네트워크 상의 정보를 저장하고 있다가 사용자의 질의가 들어오면 그 질의에 맞는 정보를 검색하여 보여주는 방식이다. 이러한 시스템들은 실제로 가장 많이 사용되고 있고,



Fig. 1과 같은 형태를 취하고 있으며, 가장 대표적인 시스템들로 Altavista, Yahoo, MetaCrawler등이 있다.

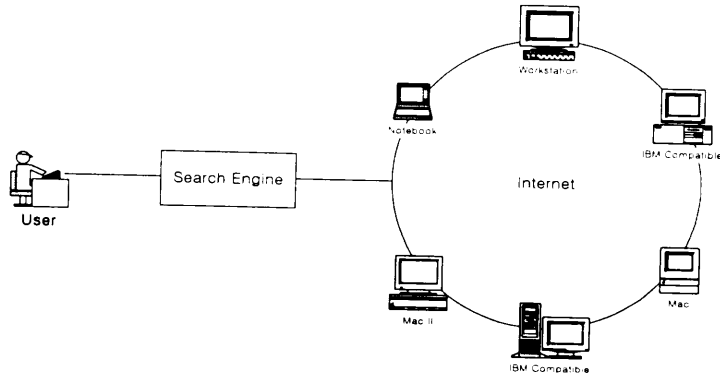


Fig. 1 Server Oriented System

이러한 서버중심의 검색 시스템이 가지는 문제점들은 다음과 같다. 첫째 모든 작업이 서버에게 집중되어 서버의 부담이 너무 크다. 그러므로 네트워크의 크기가 커질수록 점점 더 고성능의 서버가 필요하게 된다. 둘째, 네트워크의 트래픽이 증가하는 단점이 있다. 서버에 대한 사용자의 연결요청, 웹 로봇의 정보검색과정 등에서 발생하는 네트워크 트래픽은 네트워크 서비스의 중대한 성능저하요인이 된다. 세 번째, 검색된 내용에 따라서 불필요한 정보의 검색이 가능하기 때문에 실제로 원하는 정보를 얻는데는 어려움이 발생할 수 있다.

## 2)프록시 서버를 이용한 검색 시스템

프록시 서버 중심의 검색 시스템은 프록시 서버가 검색 엔진의 역할을 한다고 볼 수 있다. 사용자들이 지역(Domain)밖에 있는 네트워크로 연결되기 위해서는 프록시 서버를 거치게 된다. 또한 밖에서 안으로 들어오는 정보들 또한 프록시 서버를 거친다. 결국 안에서 밖으로, 또는 밖에서 안으로 들어오는 정보들이 모두 프록시

서버를 거쳐가므로, 프록시 서버에서 이러한 정보들을 필터링 (Filtering)하여 저장해 둔다. 사용자가 프록시서버에 질의를 하면, 프록시 서버는 저장된 정보들을 검색하여 결과를 사용자에게 보여 주게 된다. 연구소처럼 같은 지역 내에 있는 사용자들은 대개 비슷한 분야의 일을 하고 또한 비슷한 분야에 관심이 있을 것이라는 아이디어에서 나온 시스템이며, Fig. 2와 같은 형태를 취하고 있다.

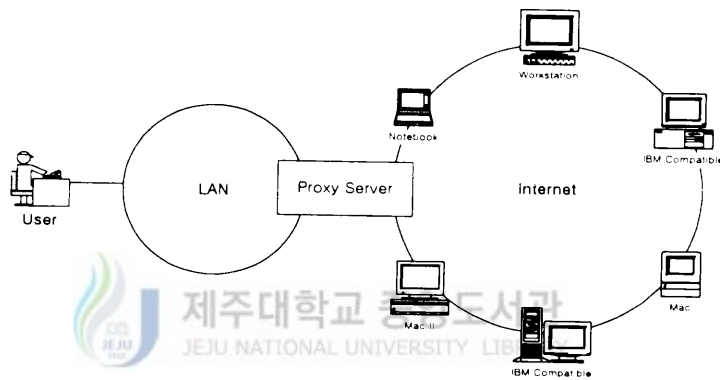


Fig. 2 Proxy Search Oriented System

프록시 기반 검색 시스템의 예로는 OSF WAIBA를 들 수 있다. WAIBA(Wide Area Intelligent Browsing Assistance)는 대부분의 사용자들은 어떤 회사나 연구소와 같은 조직에 속해 있으며, 같은 조직의 사람들은 같은 형태의 정보를 사용한다는 측면에서, 각 조직 당 그룹 서버가 있고 여기에서 그룹의 사람들이 관심을 가지는 정보에 대한 분석과 예측을 하며, 이를 바탕으로 필요한 정보를 미리 가져다 놓을 수 있게 하는 구조이다. 개인의 특성은 그룹의 특성을 이루고, 이것은 또 전체 특성을 이룬다는 점을 이용해 네트워크 사용과 사용자의 행동에 대한 모델을 만들고 이를 통해 미래의 사용을 좋은 방향으로 도모할 수 있다는 아이디어를 가진다.

프록시 중심의 검색 시스템은 서버 중심의 시스템보다는 서버의

부담이 많이 줄어들고 또한 네트워크의 트래픽도 많이 줄어든다. 그러나, 저장하고 있는 정보가 너무 편중되어 있기 때문에, 즉 사용자들이 이용하는 정보들이 너무 특정 분야에만 깊이 관련된 정보일 확률이 높기 때문에 넓은 분야의 정보를 얻을 수 없다는 단점이 있다.

### 3) 사용자 중심의 검색 시스템

사용자 중심의 검색 시스템은 사용자가 하나의 에이전트를 두어 자신의 기호(Preference)에 맞게 웹 브라우저를 관리하는 시스템이다. 에이전트는 자신의 사용자의 기호를 여러 가지 방법으로 분석하여, 관련된 웹사이트를 찾는다면 자주 방문하는 웹사이트의 URL을 북마크 파일에 포함시키는 등의 작업을 함으로써 사용자들이 좀 더 편리하고 빠르게 정보를 검색할 수 있게 한다. 사용자 중심의 검색시스템은 Fig. 3과 같은 형태를 취하고 있으며, 대표적인 예로 NetBook을 들 수 있다. Net Books는 사용자가 좀 더 효율적인 웹사이트의 접근을 할 수 있도록 관리하는 시스템이다. 모든 사이트들을 여러 개의 Book단위로 분류하여 관리한다. 예를 들어 사용자가 빨리 접근하고자 하는 웹 페이지를 모아 하나의 책으로 구성하고 이 페이지들은 지역 환경에 저장해 두었다가 빠른 결과를 보여주며, 자주 접근하는 웹사이트들의 변화를 검색하여 항상 최신의 정보로 바꾸어 주는 일도 할 수 있다. 그 외에도 검색 엔진에서 나온 결과를 이용하여 좀 더 효율적인 검색을 도와주는 등의 사용자의 검색 환경을 개선시키고자 많은 노력을 하였다.

그러나 사용자 중심의 검색 시스템은 진정한 의미의 검색 시스템으로 볼 수 없다. 이것은 단지 사용자에게 속한 하나의 에이전트가 사용자의 기호에 맞게 검색환경을 구성하여 사용자에게 좀 더 빠르고 편리한 웹 브라우징 환경을 제공하려는 것이 목적이다.

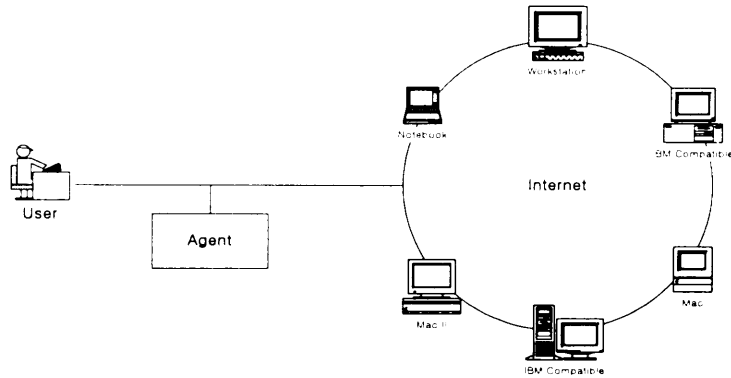


Fig. 3 User oriented system

#### 4)기타 - Aglet Sample Program

Java는 현재 이동 에이전트를 구현할 수 있는 언어들 중 하나이다. IBM사에서는 Java로 이동 에이전트를 구현할 수 있는 Aglet Workbench라는 환경을 제공하고 있다. Java로 작성된 이동 에이전트를 Aglet이라고 이름지었으며, 자세히 말해 Aglet은 이동 에이전트의 Object이다. Aglet을 구현하기 위해서 필요한 Interface들과 Class들을 Java Aglet API(J-AAPI)로서 제공하고 있다. IBM사에서는 이 환경을 이용하여 몇 가지 간단한 예제 프로그램을 구현하였는데, 그중 이동 에이전트가 지정된 곳으로 이동하여, 주어진 키워드가 이름에 포함되는 파일을 디렉토리 구조를 검색하여 찾아서 결과로 보내는 Searcher라는 프로그램이 있다. 이 프로그램은 클라이언트/서버 모델을 기반으로 하였으며 지역 환경에 고정되어 있는 Searcher 에이전트가 사용자로부터 키워드와 가야할 곳들의 주소를 입력받는다. 그리고 이동 에이전트인 Searcher\_slave 에이전트에게 이 정보들을 주어 이동시킨다. 이동 에이전트는 주어진 주소들을 차례로 방문하면서 같은 작업을 계속 반복하며, 그 검색 결과를 하나의 Searcher\_info라는 이름의 객체를 생성하여 저장한

다. 마지막 주소까지 모두 방문을 끝마친 이동 에이전트는 결과 객체를 가지고 지역 환경으로 돌아오게 된다.

#### 5) 기존 시스템들에 대한 비교 분석

이제까지 소개한 여러 가지의 정보 검색 시스템들을 처음에 제시한 문제점들을 중심으로 비교해 보면 Table 3과 같다. 제시된 4가지의 평가 기준들은 다음과 같다.

- 서버의 부담 - 시스템의 서버에 가해지는 부담
- 네트워크 트래픽 - 시스템으로 인해 야기되는 트래픽
- 검색정보의 효율성 - 검색 결과 중에서 유용한 정보가 차지하는 비율
- 검색 정보의 양 - 검색되어 나온 전체 정보의 양

Table 3 The analysis of Before system

	Server Overhead	Network Traffic	Efficiency	Amount
Search Engines	X	X	X	●
OSF WAIBA	▲	●	▲	▲
Net Books	●	●	●	X

X: Bad, ▲:Normal, ●: Good

서버중심 시스템인 검색 엔진은 서버의 부하, 네트워크 트래픽, 그리고 검색정보의 효율성 면에서 문제점이 많으나, 검색 시스템의 가장 중요한 기능인 정보의 양적 측면에서 다른 두 시스템에 비해

서 우수하다. 프록시서버 중심의 시스템인 WAIBA는 검색 엔진이 세 가지 문제점을 많이 보완했으나 원래의 구현 목적이 정보의 그룹화를 위한 것이므로, 정보의 양이 비교적 적고 편중되어서 한정된 분야의 정보에 대한 검색에만 좋은 효과를 나타낸다. 사용자 중심의 시스템인 Net Books는 검색시스템보다는 사용자의 검색환경 관리 시스템이라고 할 수 있으므로 직접적인 비교는 불가능하지만, 사용자가 자주 이용하는 사이트에 대한 정보는 가장 효율적으로 관리할 수 있다. 그러나 정보의 양이 너무 적고, 실질적인 검색 기능은 제공하지 않는다는 치명적인 단점이 있다. 그러나 에이전트가 사용자의 미래의 행동을 예측하고 앞으로 사용자가 원할 것 같은 정보를 미리 가져오는 시스템을 설계할 수 있다면 이러한 단점이 많이 보완될 수 있을 것이다.



### III. AQL시스템 설계

본 논문의 목표는 에이전트 기술을 이용하여 사용자 질의에 대한 적절한 키워드 생성과 기존의 질의연결 히스토리와 여러 검색엔진의 정보공유를 통해 전문가의 지식을 이용하는 사용자 중심의 검색 방법을 제안함으로써 기존의 질의응답방식과 검색방법이 가지고 있는 문제점들을 해결할 수 있는 새로운 AQL시스템(Automatic Question Link System)을 제안하는 것이다. 여기서 좀 더 효과적인 결과를 얻기 위해 사용자 데이터와 질의연결 히스토리를 사용한다. 이 장에서는 새로운 질의연결 방법을 제안하고, 전체적인 시스템 구조설계와, 시스템 동작과정을 설명하였다.

#### 1. 기존 시스템의 문제점

II장에서 알아본 것과 같이 사용자질의에 대한 기존의 질의응답 시스템의 문제점은 다음의 네 가지로 압축할 수 있다.

- 질의를 한 후 무한정 누군가의 답변을 기다린다.
- 답변이 없을 경우 그냥 지나쳐버린다.
- 결과의 신뢰성이 떨어진다.
- 과도한 대기시간을 요한다.

또한, 기존의 검색시스템의 문제점은,

- 서버의 부담이 너무 크다.
- 네트워크의 트래픽을 야기시킨다.
- 결과에 불필요한 정보가 너무 많아 정보의 질이 떨어진다.

기존의 사용자질의에 대한 검색시스템은 앞에서 언급한 것과 같이 세 가지로 분류할 수 있는데, 첫 번째 서버 중심 시스템은 가장 일반적으로 쓰이고 있으나 위의 문제점을 모두 가지고 있다. 두 번째는 프록시서버 중심의 검색 시스템으로써 이 시스템은 첫 번째의 시스템보다는 서버의 부담이 적고 네트워크 트래픽도 줄일 수 있으나 정보의 범위가 한정되어 있다는 문제가 있다. 세 번째 시스템은 사용자 중심 시스템인데 이것은 단순히 브라우저를 관리하여 사용자가 보다 효율적으로 정보를 얻도록 하는 에이전트 시스템으로써 새로운 분야의 정보를 찾는 데는 전혀 적합치 않다.

이러한 기존 시스템의 문제점을 해결하기 위해 다음 2절에서는 AQL시스템에 대한 모델을 제시한다.

## 2. AQL시스템의 제안

### 1) 키워드생성에 따른 정보검색 방법

#### (1) 사용자 경향분석과 질의연결을 위한 정보수집

사용자가 원하는 정보의 종류를 세분화하기 위해 사용자의 아이디, 직업과 관심분야, 질의를 데이터베이스화하는 단계이다. 이를 위해서 사용자등록 폼, 질의입력 폼, 입력된 정보를 데이터베이스화하는 데이터베이스 처리 등으로 구성되며, 이러한 자료는 검색의 결과로 나오는 정보에 신뢰성을 더해준다.

#### (2) 에이전트



사용자의 관심사를 학습하고 이전의 질의연결 히스토리 데이터베이스를 분석해 사용자 질의를 정확히 파악하여 검색을 위한 키워드와 카테고리 생성해내는 KGA(Keyword Generator Agent), 사람을 대신하여 정보를 검색해주는 SA(Search Agent)로 구성된다. 이를 위해서 문서수집 모듈과, 자연어처리에 기반 하여 비정형문서를 분석할 수 있는 문서분석모듈, 분석된 문서를 데이터베이스화하는 데이터베이스 처리 모듈, 그리고 지능형 웹 로봇에 기반 한 메타검색모듈 등으로 구성된다. 문서 분석 모듈은 형태소 분석기 및 품사 태거를 이용하여 질의를 이루고 있는 문장을 분석한 후 키워드의 추출을 처리하게 된다. 일반적으로 띄어쓰기 단위로 어절을 분리한 뒤 불용어 리스트를 이용하여 제거대상 어절을 없애고 나머지를 모두 키워드로 간주한다. 그러나 이러한 방법은 첨가어라는 한국어에 적합하지 않는 방법이다. 예를 들어 “절약과”, “저축만이”, “살”, “길이다.”라는 어절이 모두 키워드가 된다. 따라서 사용자가 “절약”, “저축”이라는 키워드를 입력하면 검색이 되지 않는다. 이는 한국어가 첨가어적인 성격을 가지고 있으므로 어절 내의 의미를 포함하는 부분과 기능을 포함하는 부분을 분리해야 하는 형태소 분석과정을 포함하지 않았기 때문이다.

형태소 분석과정을 거치면 “절약과”는 “절약\_명사 + 과\_접속조사”로 “저축만이”는 “저축\_명사 + 만이\_조사”, “살”은 “살다\_동사 + 르\_관형어어미”, “길이다.”는 “길\_명사 + 이다\_종결어미 + .”등으로 분석될 수 있으며 키워드 추출과정 시 더욱 문서를 잘 표현하는 키워드를 추출할 수 있다.

또한 명사뿐만 아니라 조사, 어미 등 다른 품사에 대해서도 키워드화 할 수 있는 근거가 마련되므로 향후, 구문분석, 의미분석 등의 자연어질의 처리할 수 있다. 반면, 기존의 자연어처리 정보검색은 이러한 형태소적 분석을 통한 한국어 분석이라기 보다는 명사 사전을 이용한 명사추출 시스템이라고 볼 수 있으며, 용언에 대한 분석 능력을 가

지고 있지 않다. 따라서 “금융실명제에 대해서 검색하라”라는 질의와 “금융실명제에 대해서 검색하지 말라”라는 질의를 제대로 구별할 수 없다. 이러한 자연어처리과정을 하는 분석모듈은 이미 개발된 모듈을 사용한다.

본 연구개발을 통하여 인터넷 회선 점유에 의한 네트워크 과부하를 최소화하면서도 사용자가 원하는 정보를 효과적으로 제시할 수 있도록 사람을 대신하여 인터넷을 검색하고 그 결과를 제시하는 에이전트가 개발될 것이다.

## 2) 전문가의 지식을 이용하는 검색 방법

기존의 검색 엔진에서는 검색결과 불필요한 결과로 너무 많이 나온다는 문제가 있다. 주어진 키워드로 검색된 결과만으로는 어떤 웹사이트가 정말 중요한 정보를 가지고 있어서 사용자들이 많이 방문하는 곳인지를 알 방법이 없다. 단지 검색 엔진에 간단히 나타난 정보만으로 판단하여야 한다.

만약 어떤 분야에 대한 전문가들이 많이 방문하는 웹사이트를 알아낼 수 있다면 아주 좋을 것이다. 우리가 어떤 작업을 수행하는데 있어서 비슷한 관심 분야를 가진 사람들이 가까이 모여서 어떤 그룹 단위로 일을 하는 경우가 많다. 이런 경우 우리는 누가 어떠한 분야에 대한 전문가라는 것을 알 수 있다. 만약 내가 그 사람의 정보를 공유하여 얻을 수 있다면 중요한 웹사이트를 찾는 데 도움이 될 수 있다. 즉, ‘내가 알고자 하는 정보에 대한 전문가의 지식을 이용하자’는 것이 아이디어이다.

예를 들어 어떤 사용자 A가 B라는 분야에 대한 전문가이고 그 분야에 대해 아주 관심이 많다고 해 보자. 이젠 컴퓨터를 이용하는 사람이라면 누구나 웹을 필수적으로 이용한다고 생각할 수 있다. 이때 우리는 그 사람이 웹 브라우저를 통해 방문하는 사이트들 중에 B에 관련된 중요한 정보가 아주 많을 것이라고 생각할 수 있다.

그러므로, 그 사용자가 웹 브라우저를 통해 그 동안 방문한 웹사이트들에 대한 정보를 공유하며, B에 관련된 중요한 정보들을 가지고 있는 곳들을 알아낼 수 있다. 우리는 그러한 정보를 결과로 받아 직접 그 웹사이트에 접속함으로써 B에 대한 지식을 얻을 수 있다.

그러나 이러한 검색 방법은 내가 원하는 정보에 대한 전문가는 누구인가에 대한 정보는 있어야 한다는 전제조건이 따른다. 그러나 이러한 정보는 사실상 인터넷상에서는 개인이 알아내기는 불가능한 정보이다. 이러한 정보는 주변에 있는 사람들에게 대해서만 알 수 있을 뿐이다. 그러므로 앞에서 제안한 검색 방법은 소규모의 집단이 사용하는 네트워크, 즉 인트라넷 규모의 환경에서는 유용하게 쓰일 수 있을 지 몰라도, 인터넷 규모의 커다란 네트워크 상에서 이러한 검색 방법을 사용하기 위해서는 사실상 개인이 앞에서 언급한 정보를 얻어낸다는 것은 매우 어렵기 때문에 그러한 정보들을 계속적으로 기록해 나간다면, 점차 신뢰도가 증가할 수 있다.

### 3) 새로운 질의연결 방법의 제안

사용자질의에 대한 키워드생성에 따른 검색과 기존의 질의응답시스템을 혼합하는 새로운 질의연결시스템을 제안할 수 있다. 사용자가 웹 브라우저를 통해 사용자 등록 시 사용자의 아이디, 직업, 관심분야를 기록하게 하여 사용자의 경향을 분석한다. 입력된 사용자의 질의를 분석하여 가능성 있는 예비키워드들을 생성한 후 사용자의 경향과 기존의 결과를 분석해 검색키워드와 카테고리를 정하여, 여러 검색엔진들을 검색하여 구축된 데이터베이스에 대한 2차 적인 질의를 하여 검색된 정보를 통해 최종사용자에게 가장 적합한 응답을 만들어내고, 이를 기반으로 가장 정확한 정보를 가지고 있는 위치로 연결한다.

제안된 모델에서 정확한 정보를 가진 웹사이트를 평가하는 기준

으로는 다음과 항목을 선정하였다

- 검색키워드와 사용자의 관심분야와의 연관성
- 검색키워드에 대한 검색기록 존재여부

이러한 기준들을 가지고 페이지들을 분석하며 사용자는 자신이 제공한 키워드와 관련 있는 내용들을 최종검색결과로 얻게 된다. 또한 관련 사이트들의 URL을 얻게 됨으로서 그 URL들에 직접 접속하여 그 분야에 관한 정보들을 얻을 수 있으며, 기존의 질의응답 시스템과의 병행사용으로, 그 질의에 대해 잘 알고 있는 전문가로부터 응답을 받을 수도 있다.

### 3. AQL시스템 구조설계

제안하는 AQL시스템의 전체구조는 다음 Fig. 4와 같다

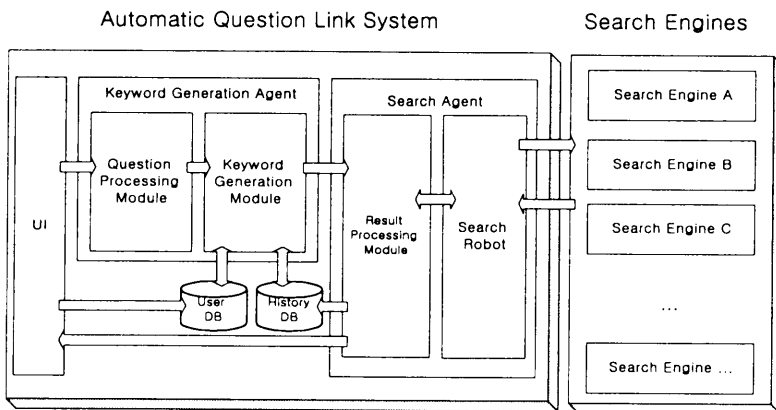


Fig. 4 The Architecture of AQLS

KGA은 검색을 위한 검색키워드와 카테고리를 결정하기 위한 에이전트이며, SA는 검색을 위한 에이전트이다. 다음은 AQL 시스템 각 모듈에 대한 요약이다.

- AQLS(Automatic Question Link System)  
사용자의 질의를 자동으로 검색결과에 연결해주는 시스템
- UI(User Interface)  
사용자로부터 입력을 받고, 결과를 보여주기 위한 GUI
- KGA(Keyword Generation Agent)  
사용자의 질의에 대한 키워드와 카테고리를 생성하는 에이전트
- QPM(Question Processing Module)  
사용자의 질의에 대한 예비키워드를 생성하는 모듈
- KGM(Keyword Generation Module)  
예비키워드와 사용자데이터베이스로부터 키워드와 카테고리를 생성하는 모듈
- SA(Search Agent)  
사용자 질의에 대해 검색을 대행해주는 에이전트
- RPM(Result Processing Module)  
SR로부터 얻어진 결과를 검색하여 연결을 제공하기 위한 모듈
- SR(Search Robot)  
각 검색엔진에 질의를 하고 결과를 받아오는 모듈
- UDB(User Database)  
사용자의 정보를 저장하는 데이터베이스
- HDB(History Database)  
질의연결결과를 저장하는 히스토리 데이터베이스
- SDB(Search Database)  
SR로부터 얻어진 결과를 저장하기 위한 데이터베이스

#### 4. 시스템 동작과정

제안된 시스템의 동작과정은 Fig. 5와 같다. 먼저 질의가 시작되면 사용자로부터 미리 정의된 사용자인터페이스를 통해 사용자정보를 입력받아 사용자 데이터베이스인 UDB에 저장하고, 사용자로부터 질의내용을 입력받아 불필요한 문장을 제거하여 가능한 예비키워드를 추출해낸 후, 예비키워드와 UDB에 저장되어있는 사용자의 정보, 그리고 기존에 검색된 질의연결 히스토리의 정보인 HDB를 참조하여 검색을 위한 키워드와 카테고리를 정한다. 인터넷에 존재하는 검색엔진들을 검색하는 자체검색로봇이 검색을 수행하여 구축된 데이터베이스인 SDB로부터, 상위 단계에서 생성된 키워드와 카테고리를 이용하여 검색을 한 후 검색결과를 정리하여 검색결과를 만들어낸 후 질의과정을 마치게 된다.

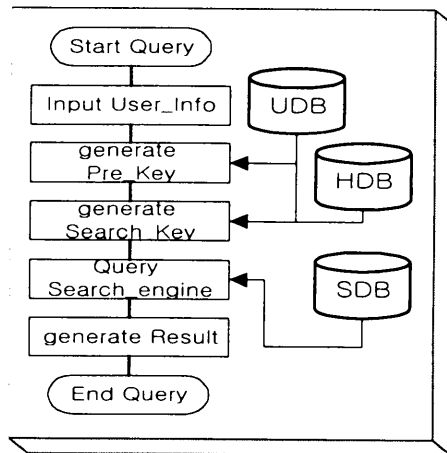


Fig. 5 The System flow

## IV. 구현 및 결과 고찰

본 장에서는 위에서 제안한 모델을 구현하기 위해 구현환경 및 각 모듈에 구현에 필요한 기술, 그리고 각 모듈의 기능에 대하여 기술한다.

### 1. 시스템 모듈 구현

본 논문에서 제안된 에이전트기술을 이용한 AQL시스템은 인터넷상에 존재하는 여러 검색엔진들을 가지고 있는 검색사이트를 대상으로 검색하였다. 시스템 구현환경은 다음과 같다.

- OS: Windows 98 Second Edition
- Programming Language : SUN's JDK 1.3, ASP, HTML
- Servlet Platform : SUN's JSDK 2.1
- Database : MS Access 7.0

#### 1) 사용자인터페이스(User Interface)

사용자에게 보여질 인터페이스는 인터넷에서 정보를 표현하기 위해 많이 이용되는 HTML과 ASP를 이용하여 작성되어 있으며, VB스크립트를 이용하여 사용자등록 폼을 구성하였다. 사용자등록 폼에서는 사용자의 아이디, 나이, 직업, 3개의 관심분야정보와 사용자의 질의내용을 입력받는다.

## 2)KGA(Key Generation Agent)

키워드를 생성해내기 위해서는 사용자 질의에 대해 자연어처리를 하여 불필요한 문장을 제거함으로써 키워드들을 생성하고, 키워드 중 사용자에게 알맞은 검색키워드를 찾는 기능으로, 질의를 처리하는 모듈과 사용자에게 알맞은 검색키워드를 분석하는 모듈이 필요하다.

### (1)QPM(Question Processing Module)

이미 개발된 모듈을 이용하며, 입력받은 사용자의 질의를 문장으로 받아들여 불필요한 문장을 제거하고 가능한 예비키워드를 생성한다. 사용자로부터 자연어 질의가 입력되면 먼저 형태소 분석과정을 거쳐 품사태깅을 수행한다. 형태소 분석기는 형태소 사전과 단어절어 사전, 그리고 형태소 패턴사전을 이용하여 등록어와 미등록어의 형태소를 동일한 방법으로 분리하고 원형을 복원하며 접속감사표를 이용하여 접속을 검사한다. 접속이 이루어진 형태소들은 부분적인 형태소그래프를 형성하고 이 그래프에 대하여 통계정보와 Viterbi 탐색을 이용하여 태깅을 한다. 이때 미등록어의 통계정보는 유절 Trigram을 이용하여 구해진다. 생성된 형태소그래프는 다시 구문분석을 거쳐 구문 트리를 생성하고 이를 기반으로 키워드를 생성한다. 이러한 처리의 전체구조는 Fig. 6과 같다.

자연어 질의로부터의 키워드를 생성하는 알고리즘의 구체적인 단계는 Fig. 7과 같다. 구문분석 결과트리로부터 키워드와 연산자 정보를 추출하여 트리를 간략화 한 후 명사구 합성 및 키워드의 가중치를 부여하여 키워드를 생성한다. 블리언트리를 중위순회(Inorder Traverse)하면서 키워드를 생성한다.



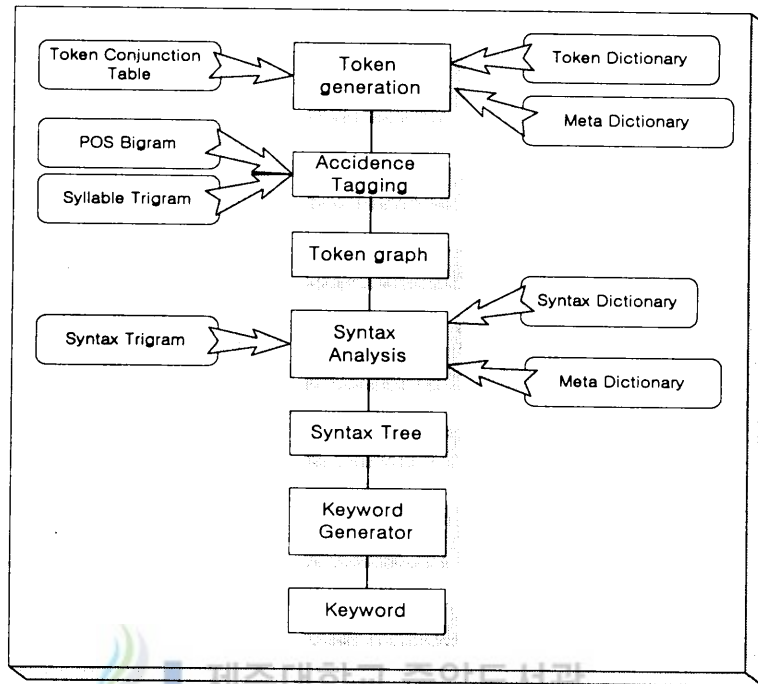


Fig. 6 The Architecture of QPM

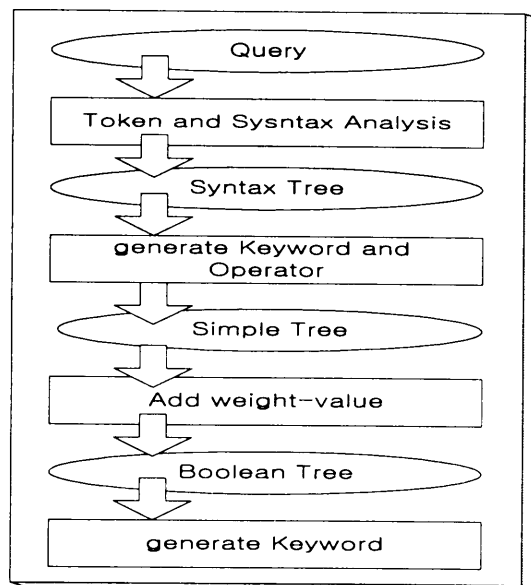


Fig. 7 The flow of QPM

## (2)KGM(Keyword Generation Module)

QPM에서 생성한 키워드를 바탕으로 사용자 정보와 기존의 질의 연결히스토리 정보인 HDB(History DataBase)를 분석하여 검색키워드와 카테고리를 생성하는 기능을 담당하며, Fig. 8과 같은 흐름으로 동작한다. 먼저, 가능한 키워드들을 중심으로 사용자의 직업과 비교한다. 직업과 관련이 있는 키워드에 가중치를 두고, 검색키워드의 순서를 재배치한다. 해당사용자가 입력한 관심분야를 UDB로부터 찾아보고, 있으면 관련분야코드와 키워드를 SA로 넘겨주고, 없다면 HDB에서 이전에 같은 키워드로 검색된 카테고리과 키워드를 SA로 넘겨주며, Java언어를 이용하여 구현하였다.

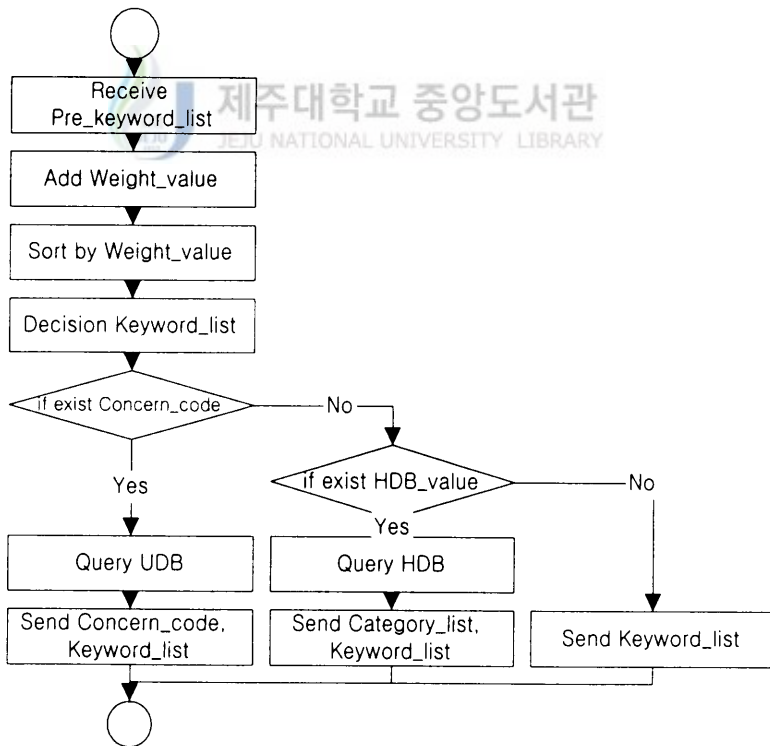


Fig. 8 The of KGM module

### 3)SA(Search Agent)

#### (1)SR(Search Robot)

검색엔진의 역할을 하며 여러 검색엔진들이 가지고 있는 사이트들의 정보를 카테고리별로 분류하여 페이지들을 저장하며, Fig. 9와 같이 동작한다. 검색작업이 시작되면 먼저, 사이트들의 정보를 저장하기 위해 데이터베이스의 연결을 시도하고, 데이터베이스 연결에 오류가 발생하면 오류메시지를 표시한다. 다음 단계로 여러 검색엔진들의 주소로 검색을 시작하여 각 검색엔진들이 가지고 있는 사이트들의 정보인 주소, 제목, 분류, 내용, 기타를 데이터베이스에 저장하는 기능을 구현하였다.

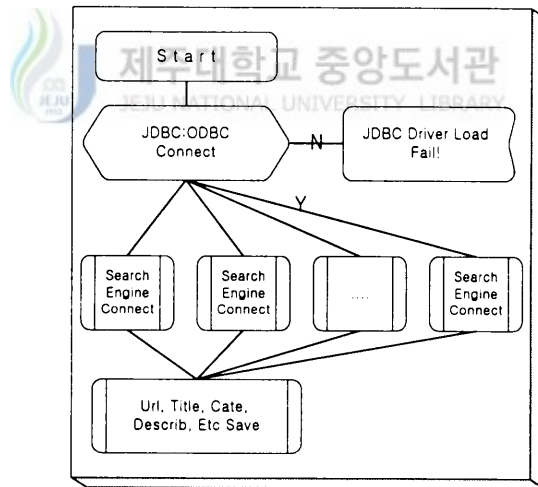


Fig. 9 The structure of SR

#### (2)RPM(Result Processing Module)

검색이 시작되면 SR에 의해 저장된 사이트들의 정보를 가지고 사용자의 질의에 대한 키워드와 카테고리를 이용하여 검색 후, 최

중사용자에게, 질의에 대한 가장 효과적인 정보를 가지고 있는 사이트로 연결한다. 먼저 관심분야 즉, 카테고리가 3가지 등록되어 있으므로, 첫 번째 카테고리에 키워드를 이용하여 검색하고, 두 번째 카테고리에 키워드를 이용하여 검색하고, 세 번째 카테고리에 키워드를 이용하여 검색한 후, 결과로 나온 페이지 중 중복되는 페이지를 제거한 후 결과를 HDB에 저장하고 사용자에게 제공하기 위해 HTML 문서로 변환하여 검색을 마친다. Fig. 10은 RPM의 구조를 보여준다.

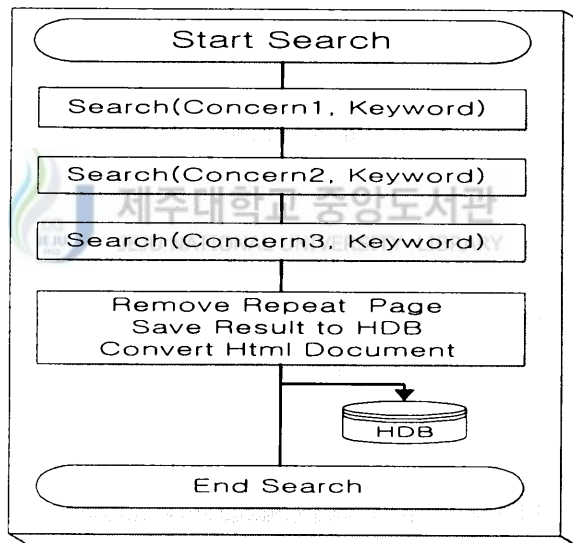


Fig. 10 The structure of RPM

#### 4) 데이터베이스 설계

사용자의 정보와 질의, 그리고 질의를 통해 연결된 결과와 검색 결과를 저장하는 데이터베이스가 필요하므로, 사용자DB와 히스토리DB 그리고 검색결과DB를 구성한다.

##### (1) 사용자데이터베이스(User Database)-UDB

사용자인터페이스를 통해 입력받은 정보와 질의내용을 데이터베

이스에 저장하며, 사용자데이터베이스의 테이블은 User Table과, Question Table이다. User Table은 사용자의 정보인 아이디, 직업, 관심분야1, 관심분야 2, 관심분야 3을 저장하기 위한 필드를 가지고 있으며, 아이디는 텍스트로 된 10글자, 직업코드는 2자리의 정수, 관심분야코드 3개는 관심분야표의 대분류와 소분류를 합한 4자리 정수를 사용한다. Question Table은 사용자 아이디와 해당 사용자의 질의내용을 저장할 수 있도록 텍스트 10글자의 아이디, 질의내용을 저장할 수 있도록 텍스트 50글자로 정의된 필드를 가지고 있다. Fig. 11은 UDB의 테이블 구조이다.

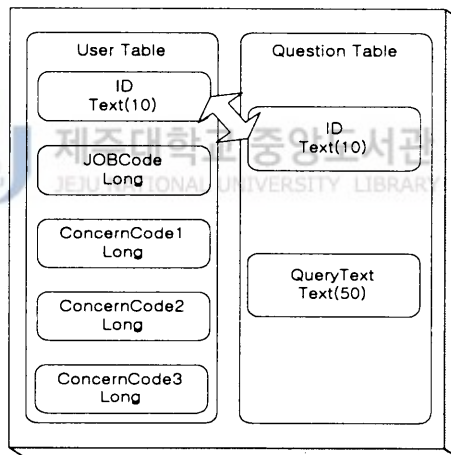


Fig. 11 User Database

- 직업코드 : 01 -> 회사원(컴퓨터) , 19 -> 공무원
- 관심분야코드 : 0102 -> 방송, 0402 -> 인터넷, 0501 -> 여행

(2)질의연결히스토리 데이터베이스(History DataBase)-HDB  
 검색엔진으로부터 검색한 결과를 저장하며, HDB의 테이블은 Search Table과 Cate Table이다. Search Table은 검색키워드

대한 검색결과를 연결한 정보를 저장하기 위한 테이블로, 키워드를 저장하기 위한 텍스트 10자, 페이지의 주소를 저장하기 위한 80자, 제목을 저장하기 위한 100자, 분류를 저장하기 위한 4자, 내용을 저장하기 위한 메모, 기타를 저장하기 위한 100자의 필드를 가지고 있고, Cate Table은 해당페이지의 분류를 정의하기 위한 테이블로 4자리 정수로 된 분류코드와 해당분류의 분류명인 20자리 필드로 구성되었다. Fig. 12는 HDB의 테이블 구조이다.

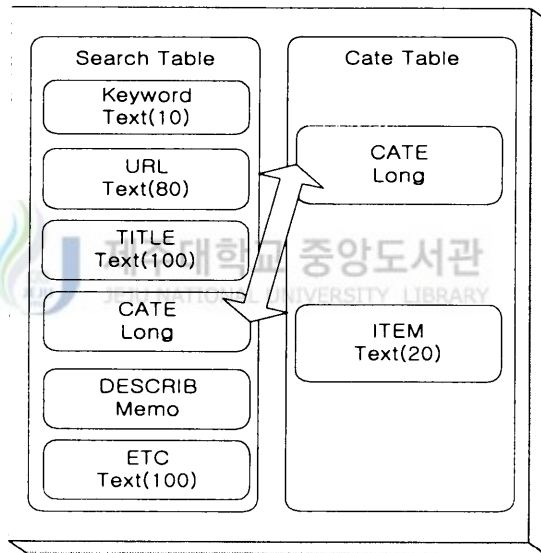


Fig. 12 History Database

### (3) 검색결과 데이터베이스 - SDB

SR에 의해 검색된 페이지의 정보를 저장하기 위한 테이블로, SEI table은 검색을 한 검색엔진의 코드 4자리와, 80자의 주소, 100자의 제목, 4자리의 분류코드, 메모형식의 내용, 100자의 기타를 저장하기 위한 필드를 가지고 있고, SEU table은 검색엔진의 코드 4자리정수와, 20자의 이름, 80자의 주소를 저장하기 위한 필드로

구성되었다. Fig. 13은 SDB의 테이블구조이다.

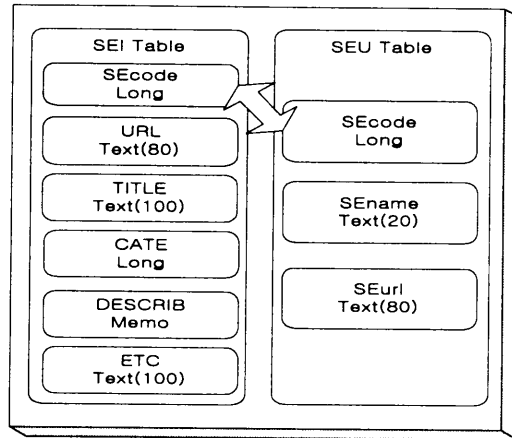


Fig. 13 Search Database

다음의 Table 4와 Table 5는 사용자의 관심분야와 직업등록을 위한 직업코드표와 관심분야 코드표 이다.

Table 4 Job table

	직업	코드	직업	코드	직업	코드
회사원	컴퓨터	01	초등학생	11	법조인	21
	인터넷/온라인	02	중학생	12	의료인	22
	전자/전기/통신	03	고등학생	13	언론인	23
	광고/마케팅/홍보	04	학원생	14	예술인	24
	제조/생산	05	대학생	15	농/수/축산/광업	25
	금융/증권/보험/회계	06	대학원생	16	연예/스포츠	26
	유통/물류	07	주부	17	군인/경찰	27
	출판	08	자영업	18	종교인	28
	건설	09	공무원	19	부직	29
	운송/교통	10	교직자/학원강사	20	기타	30

Table 5 The concern table

대분류	소분류	
뉴스&미디어 01	신문(01)	잡지/웹진(03)
	방송(02)	인터넷방송(04)
예술&엔터테인먼트 02	영화(01)	문학(03)
	음악(02)	디자인(04)
비즈니스&경제 03	산업용품(01)	광고(03)
	금융/투자(02)	보험(04)
컴퓨터/인터넷 04	게임(01)	멀티미디어(03)
	인터넷(02)	통신(04)
레크레이션 05	여행(01)	유머(03)
	자동차(02)	레저/야외(04)
생활&여성 06	여성(01)	요리(03)
	주택(02)	결혼(04)
쇼핑 07	영화/음반/공연(01)	컴퓨터(03)
	도서(02)	
스포츠 08	야구(01)	골프(03)
	축구(02)	농구(04)
건강&의학 09	의학(01)	대체의학(03)
	질병/증상(02)	병원(04)
국가정보&지역정보 10	대한민국(01)	유럽(03)
	북아메리카(02)	아시아(04)
문화&사회 11	사람들(01)	종교(03)
	정부/정치(02)	장애(04)
교육&참고자료 12	교육(01)	사전(03)
	도서관(02)	통계(04)
과학&학문 13	인문과학(01)	공학(03)
	사회과학(02)	
취업정보 14	종합취업정보(01)	학원/훈련기관(03)
	자격증(02)	



## 2. 시스템 검증 및 결과

### 1) 사용자 인터페이스

본 논문에서는 사용자의 정보와 질의내용을 입력받기 위해 웹상에서 볼 수 있는 HTML 문서와 MS ASP를 사용하여 작성되었으며, Fig. 14는 검색어를 생성을 하기 위한 데이터를 입력받기 위한 화면이다.

* 아이디	compueye
* 나이	27 세
* 직업	회사원(컴퓨터)
* 관심분야	인터넷방송
* 관심분야	영화
* 관심분야	통신

Fig. 14 User Interface A

이를 위해 아이디, 나이, 직업, 3개의 관심분야를 입력받는데, 실험데이터로 입력되는 데이터로 아이디는 'compueye', 나이는 27세, 직업은 컴퓨터에 관련된 회사원, 관심분야는 '인터넷방송', '영화', '통신'이 선택되었다.

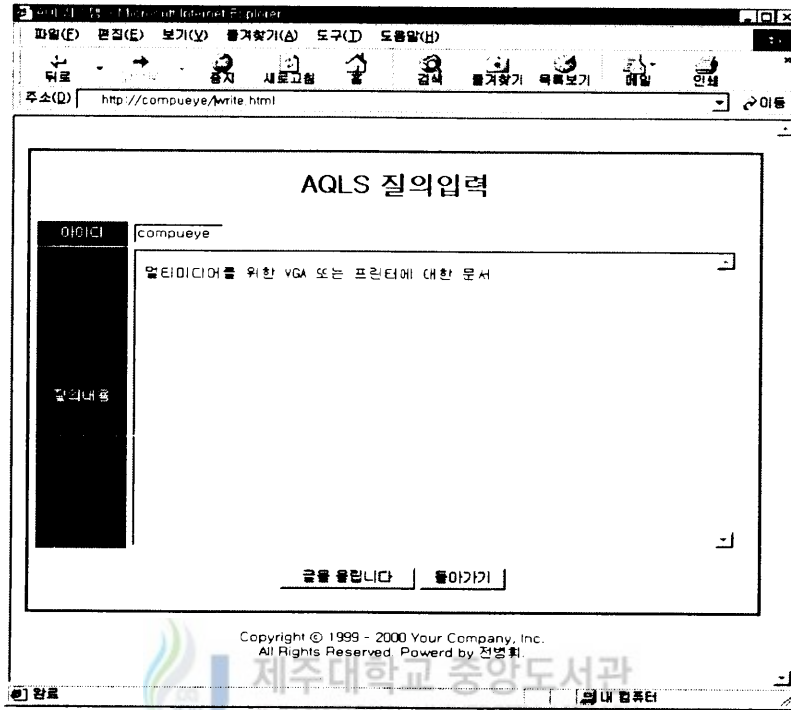


Fig. 15 User Interface B

Fig. 15는 사용자의 질의를 입력받기 위한 화면으로, 사용자의 아이디와 질의내용을 입력받도록 하였다. 입력된 아이디는 'compueye'이고, 질의내용은 '멀티미디어를 위한 VGA 또는 프린터에 대한 문서'이다.

## 2) 자연어 질의에서 키워드 생성

사용자가 입력한 문장을 통해 검색어를 생성하기 위해 자연어 처리를 거쳐 가능한 키워드를 생성해 내었다.

\*\*\*\*\*  
 자연어 질의 : 멀티미디어를 위한 VGA 또는 프린터에 대한 문서  
 \*\*\*\*\*

```

(0)3.956e-87 /complete/end---s<문장시작>(I)
(0)3.956e-87 complete
(0)3.956e-87 \ /np---MC*<멀티미디어>(멀티미디어)
(0)3.956e-87 \ /np/np
(0)3.956e-87 \ / \ (np/np)\np---jc<을>:s<#>:DI여<위하>:eCNMG<ㄴ>
(0)3.956e-87 \ /np
(0)3.956e-87 \ / \ /np---sf<VGA>(VGA)
(0)3.956e-87 \ / \ /np/np
(0)3.956e-87 \ / \ / \ (np/np)\np---BM<또는>(또는)
(0)3.956e-87 \ / \np
(0)3.956e-87 \ / \np---MCF<프린터>(프린터)
(0)3.956e-87 \ /np/np
(0)3.956e-87 \ / \ (np/np)\np---jO<에>:s<#>:DI여<대하>:eCNMG<ㄴ>
(0)3.956e-87 \ /np
(0)3.956e-87 \ / \np---MCC<문서>(문서)
(0)3.956e-87 \end (0)3.956e-87 \end\X---s<문장끝>(I)

```

\*\*\*\*\*  
 예비검색어 : (((멀티미디어[0.335156] & (VGA[0.541241] | 프린터[0.412040]))  
 & 멀티미디어VGA[0.876397] & 멀티미디어프린터[0.747196])

위의 결과 중 가중치가 가장 높은 '멀티미디어VGA'와 '멀티미디어프린터'를 중점키워드로 하여 검색을 시작한다.

### 3)검색결과화면

사용자 질의가 등록되고, 질의에 대해 기존의 질의응답방식으로 응답이 가능한 기능과 자동으로 질의에 대한 검색어를 사용하여 검색한 결과를 보여주는 화면이다.

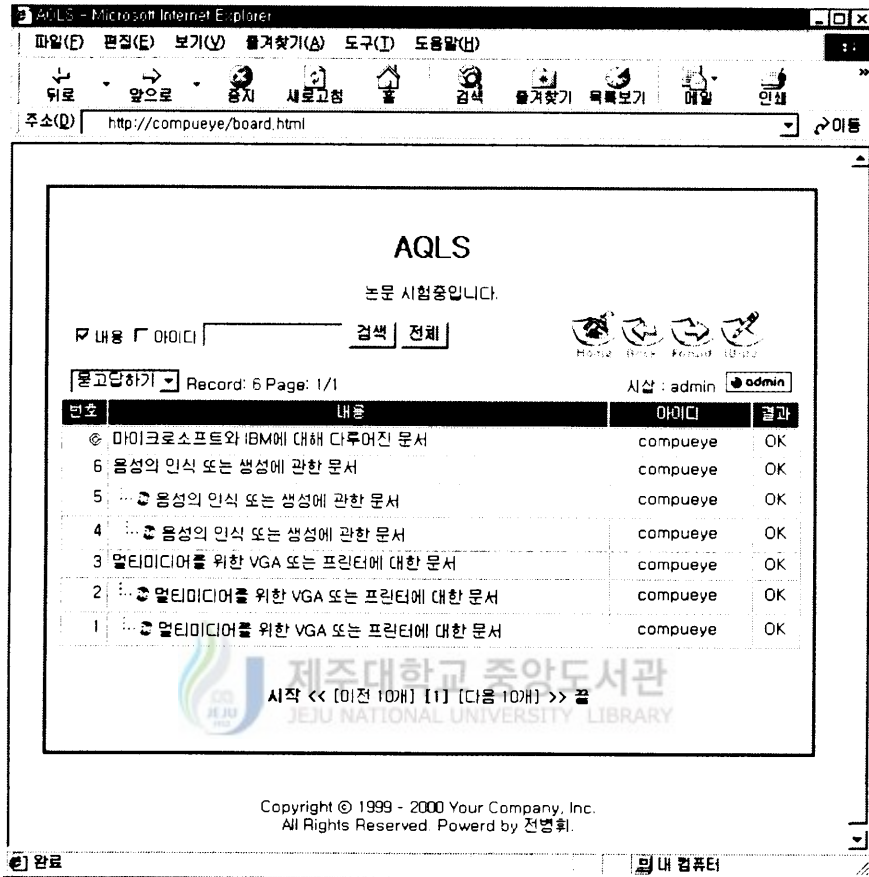


Fig. 16 User Interface C

Fig. 16은 기존의 질의응답방식과 비슷한 화면을 구성하고 있으며, 다른 기능은 질의내용에 대해 자동으로 연결한 페이지의 정보를 담고 있다는 것이다.

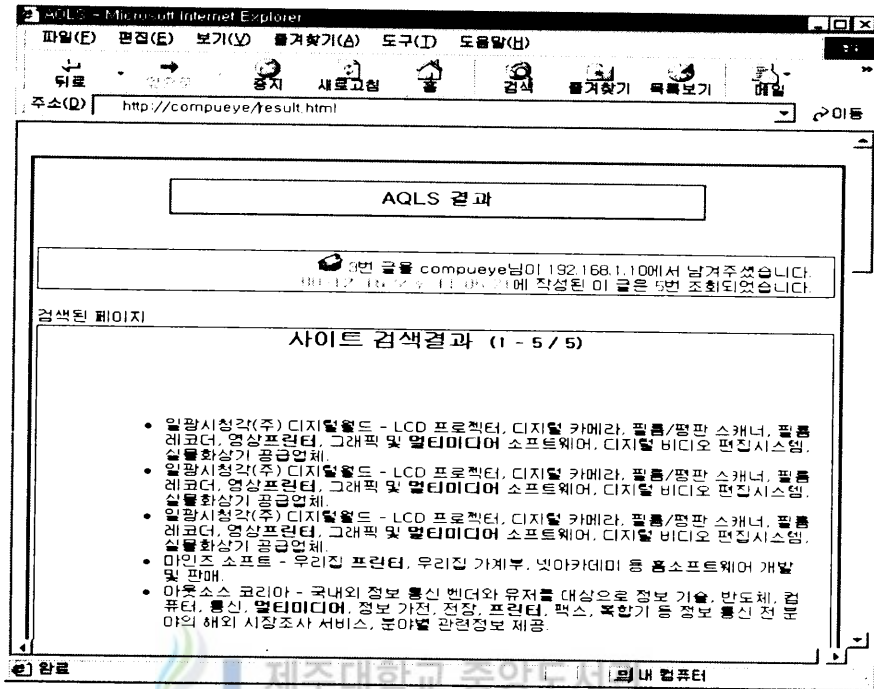


Fig. 17 The result of AQLS

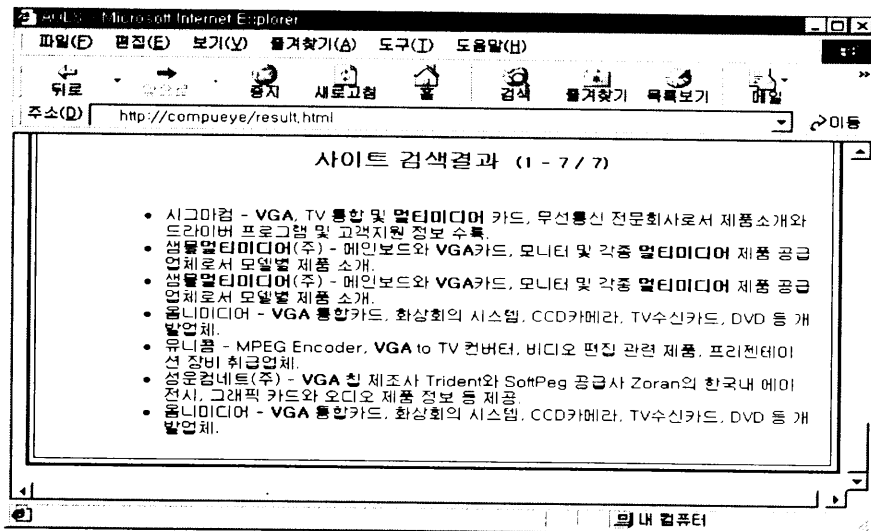


Fig. 18 The result of AQLS

Fig. 18과 Fig. 19는 AQL시스템의 결과를 보여주고 있다.

## V. 결론 및 향후 과제

본 논문에서는 네트워크의 급격한 성장으로 헤아릴 수 없이 많은 정보들 중 특정한 정보를 검색하는 방법은 여러 가지가 있으나 원하는 정보에 대한 접근이 어렵다든지 네트워크의 과부하와 같은 몇 가지의 문제점들을 문제점을 해결하기 위해 정보 공유 환경에서 에이전트를 이용한 자동질의연결시스템을 제안하였다.

기존의 질의응답시스템의 경우 질문에 대해 아는 전문가의 조언을 상세히 들을 수도 있다는 장점은 있지만, II장에서 언급한 바와 같은 문제점이 나타날 수 있다. 새로운 질의연결 방법은 III.1장에서 언급한 네 가지 문제점을 해결할 수 있으며, 기존의 방식과도 병행할 수 있다. 제안된 시스템은 기존의 질의응답방식의 형태를 취하고 있으나, 질의 후 검색을 시작하여 어떠한 결과를 검색 후 바로 제공하기 때문에, 질의에 대한 응답을 무한정 기다리지 않아도 되며, 답변이 없어 그냥 지나치는 경우가 없어진다. 또한 결과로서 제공되는 정보는 자신의 직업과 관심분야, 그리고 기존의 질의에 대한 응답히스토리과 비교하여 제공되므로, 결과의 신뢰성은 기존의 방식보다 월등하다. 그리고 검색어 생성시간과 검색에 걸리는 시간, 그리고 화면에 보여주기 위해 변환되는 시간만이 소요되기 때문에 과도한 대기시간을 요하지 않아도 된다.

제안된 시스템에서는 에이전트를 이용, 추가적인 지식을 공유하는 방법을 사용하여 검색을 수행함으로써 사람이 해야 할 단순하고 반복적인 작업을 효과적으로 처리하여 업무 효율과 시간적 비용을 절감해 줄 수 있다. 또한 제안된 시스템은 기존의 시스템이 가지고 있는 문제점들을 완화시킬 수 있다. 서버 중심 구조를 탈피함으로써 서버의 부담을 덜고, 서버에 집중되던 네트워크 트래픽을 줄일

수 있다. 또한 정보 공유를 통해 더 많은 정보를 이용할 수 있게 되며, 사용자에게 좀 더 관련된 결과를 얻게 되므로 중요하지 않은 웹사이트들의 정보가 많이 나오지 않아서 양질의 정보를 얻을 수 있게 되고, 마지막으로 자신에게 맞는 검색을 수행해주는 에이전트를 가질 수 있게 되어 같은 키워드에 대해서도 사용자마다 다른 결과를 얻을 수 있게 된다.

연구결과, 기존의 질의연결방식에 비해 대부분의 경우, 결과로서 얻어지는 데이터가 사용자의 질의에 응답이 가능하며, 그 결과로 응답 결과에서 중요한 정보를 찾기 쉽게 된다. 이러한 시스템은 사용자의 질의의도를 파악하는 검색환경을 제공하여 보다 정확한 정보검색 서비스를 구축함으로써 정보검색 서비스 및 IP분야나, 부서간의 회의나 문서전달, 검색 등에 필요한 기업 내 인트라넷 환경, 사용자들의 기호나 성향을 파악, 소비자 취향에 맞는 광고전달 및 제품정보를 제공하는 전자상거래, 사용자 수준에 맞는 멀티미디어 교육분야 등에 사용될 수 있다.

향후과제로 이러한 에이전트 시스템의 표준화에 대한 연구와 한글 자연언어처리부분의 발전이 필요하다.

## 참고 문헌

S. Franklin, A. Grasser, "Is it an Agent, or just a Program?: A Taxonomy for Autonomous Agents," University of Memphis, 1996.

Charles J. Petrie, "Agent-Based Engineering, the Web, and Intelligence," Working Draft in IEEE Expert, 1996.

L. Foner "What's an Agent Anyway"?,

D. Richman, "Distributed Computing: Let Your Agent Handle It," Techweb, 1995

T. Magedanz, K. Rothermel, S. Krause, "Intelligent Agents: An Emerging Technology for next Generation telecommunications?", "INFOCOM '96, San Francisco, 1996.3.

D. Gilbert et al., "Intelligent Agent Strategy : A White Paper," IBM Corporation, 1995.

C. Guilfole, E. Warner, "Intelligent Agents: the New Revolution in Software," Technical Report, OVUM Limited, 1994.

B. Hermans, "Intelligent Software Agent on the Internet,"



Tilburg University, Tilburg, the Netherlands, 1996.

M. Wooldridge, N. Jennings, "Intelligent Agents : Theory and practice," Knowledge Engineering Review, Vol.10, No.2, 1995.

P. Maes, "Agents that Reduce Work and Information Overload," pp.31-40, Communications of the ACM, Vol.37, No.7, 1994.7.

J.E. White, "Telescript Technology: The Foundation for the Electronic Marketplace," General Magic Paper, 1994.8.

G. Lawton, "Agents to roam the Internet," Sunworld Online.

Sun Microsystems, "The Java Language Environment: A White Paper," white papaer,  
<http://java.sun.com/docs/white/langenv/> 2000.3.

H. J. Song, S. Franklin, A. Negatu, "SUMPY: A Fuzzy Software Agent," Proceedings of the International Conference on Intelligent Systems, 1996.7.

F. Menczer, W. Willuhn, R. K. Belew, "An Endogenous Fitness Paradigm for Adaptive Information Agents," CIKM '94 Workshop,

Charles L. Brooks, "Wide Area Information Browsign Assistance : Final Technical Report," The Open Group Research Institute, 1996.4

"Net books" <http://www.frontier-software.com/>.1998.7.

Terry A. Gray, "How to Search the Web-A Guide to Search Tools,"1996.

"Eureka! Internet Search Engine from Mentor Marketing Services,"1996.

"IBM Aglet Workbench Aglet Sample Program,"IBM Corporation, 1996.

강동현, 박재현, "Core Java,"1997.9.

김병학, 이광형, 조충호, "지능형 웹 브라우저 에이전트," HCI '96, 1996.2.

전응휘, "인터넷 정보검색," WWW-KR, 1996.11.

정영미 "우리말 정보자료를 처리하는 지능형 정보검색시스템의 설계", 연세대, 1990