

히스토리 정보를 활용한 사용자 지향의 검색 시스템 설계 및 구현

전 병 휘* · 김 정 희** · 곽 호 영***

Design and Implementation for User Oriented Search System using the Information of History

Byung-Hwi Jun*, Jeong-Hee Kim** and Ho-Young Kwak***

ABSTRACT

User oriented search system is proposed and implemented using the information of history. For offering more qualified information to user without the redundant search results and unnecessary information of legacy search system, this system uses user information and generates a keyword and category. Comparing user favorite category with search result of legacy search system through the agent oriented search engine, the result close to the user category is only supplied. At the same time, search result is saved as the history databases according to each category to be used for later search work. As a result, the redundant information of search result was efficiently removed and the information close to the user favorite category was obtained.

Key Words : Agent, Keyword, Category

1. 서론

최근 인터넷 사용의 대중화와 더불어 정보의 양도 급격하게 증가하고 있으며, 이러한 변화는 사용자의 요구에 맞는 양질의 정보를 제공할 수 있는 서비스의 중요성을 한층 높이고 있다[1]. 이에 네트워크상의 정

보를 제공하기 위한 서비스로 일반적으로 많이 이용되는 것은 정보검색 시스템이며, 이 시스템은 검색대상에 따라서 서버 중심, 프록시 서버 중심, 그리고 사용자 중심의 시스템으로 분류할 수 있다. 하지만, 현재 제공되는 검색시스템을 이용하여 사용자는 원하는 정보가 존재하는 위치에 대한 정보 없이도 세부적인 정보검색이 가능한 반면, 지나치게 많은 정보의 양으로 인해 양질의 정보를 접근하는데는 어려움이 있으며, 또한 이를 극복하기 위하여 많이 이용되는 질의응답시스템인 경우는 양질의 정보를 접할 수 있다는 장점이 있으나 사용자가 요구한 질문에 대한 정보를 얻기까지는 상당한 시간이 요구된다는 단점을 가지고 있다[2].

* 큐빅아이(주) 연구원

Assistant Research Engineer, Cubici Inc.

** 제주대학교 대학원

Graduated School, Cheju Nat'l Univ

*** 제주대학교 통신·컴퓨터공학부, 첨단기술연구소

Faculty of Telecommunication & Computer Eng., Research Institute of Advanced Technology, Cheju Nat'l Univ.

따라서, 본 논문에서는 기존의 정보검색 시스템과 질의응답시스템의 단점을 극복하고 사용자에게 양질의 정보를 제공하기 위한 목적으로 히스토리 정보를 활용하면서 사용자 지향의 검색 시스템을 설계하고 구현한다. 이를 위해 기존의 다양한 응용분야에서 활용되고 있는 상용 에이전트 환경[3.5-8]하에 사전에 입력된 사용자 정보와 질의에 대한 적절한 키워드를 생성[4]하여 검색 결과와 사용자의 관심항목을 비교 분석한 후 보다 근접한 정보를 사용자에게 제공토록 한다. 또한 검색되었던 정보들은 또 다른 검색 시 검색 결과로 추가 제공될 수 있도록 히스토리 데이터베이스화한다. 그럼으로써 검색 빈도수가 증가될수록 누적되는 히스토리 정보와 현재의 검색결과를 사용자에게 제공함이 가능했으며, 정보의 질도 보다 사용자에게 근접하게 검색되었고, 일반 상용검색 시스템의 중복 검색 결과도 제거할 수 있었다.

본 논문은 다음과 같이 구성된다. 2장에서는 에이전트 환경하의 사용자 지향 검색 시스템을 설계하며, 3장에서는 구현 및 구현 결과 그리고 4장에서 결론 및 향후 연구방향을 제시한다.

II. 시스템 설계

2.1. 시스템 구조

제안하는 시스템을 구성하는 세부 모듈의 기능과 전체 시스템의 동작과정을 설계한다. 시스템의 전체 구조는 Fig. 1과 같다.

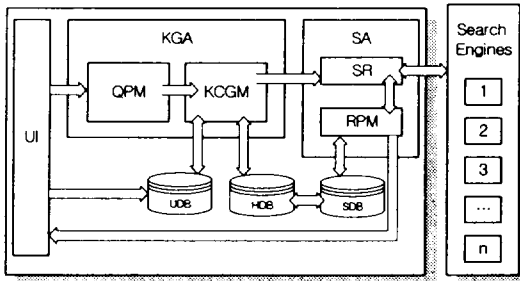


Fig. 1. System Architecture.

2.1.1. 사용자 인터페이스(User Interface)

사용자 정보를 수집하기 위한 인터페이스이다. 수집되는 정보들은 사용자의 아이디, 나이, 직업, 3개의 관심분야정보와 사용자의 질의내용이며 이 정보들을 가지고 UDB를 구성한다.

2.1.2. KGA(Keyword Generation Agent)

사용자가 질의한 내용에 대하여 자연어처리를 수행하는 질의 처리 모듈과 사용자에게 적합한 검색 키워드 및 카테고리를 생성하는 모듈로 구성된다.

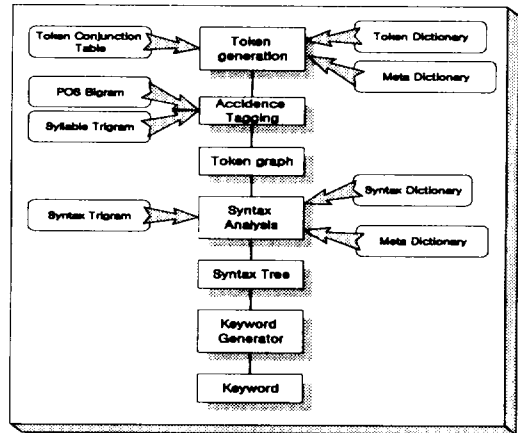


Fig. 2. Architecture of QPM.

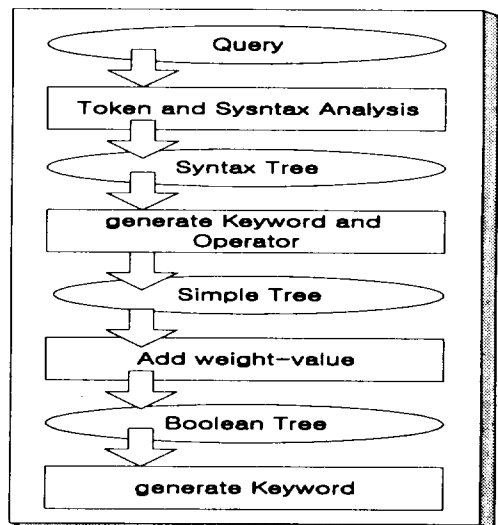


Fig. 3. The flow of QPM.

▶ QPM(Question Processing Module)

이미 개발된 모듈을 이용하며, 그 처리 과정 구조는 Fig. 2와 같으며 알고리즘은 Fig. 3과 같다. 즉, QPM 모듈을 거치면 검색 시 사용하게 될 키워드의 예비 키워드 리스트가 생성되게 되며, 이 예비 키워드 리스트들을 사용하여 KGM이 실제 사용할 검색 키워드를 생성하게 된다.

▶ KCGM(Keyword and Category Generation Module)

KCGM은 SA가 사용할 검색 키워드와 카테고리를 생성한다. 검색 키워드는 QPM에서 생성한 예비 키워드 리스트를 바탕으로 생성되고, 카테고리는 UDB 데이터베이스에서 사용자 관심항목을 분석하여 생성된다. 이 때 UDB 데이터베이스내에서 카테고리가 검색되지 않으면 사전에 검색되었던 정보들을 저장하고 있는 HDB에서 검색하여 생성하게 된다. 카테고리가 UDB에 없는 경우는 사전에 사용자 입력정보에서 입력이 누락될 수도 있기 때문이며, 이러한 경우에는 생성된 키워드에 기반하여 HDB에서 카테고리를 가져오도록 한다. 반면 카테고리가 UDB에 있는 경우에는 UDB에서 가져오게 된다. 이 과정의 알고리즘은 Fig. 4와 같다.

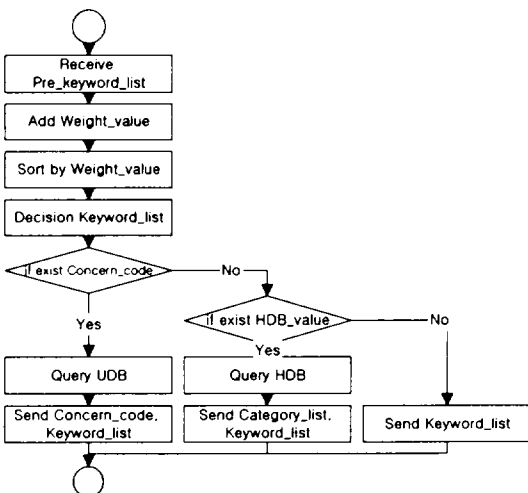


Fig. 4. Algorithm of KGM module.

2.1.3. SA(Search Agent)

KGA에서 생성된 키워드와 카테고리를 사용하여 상

용검색엔진들을 통해 정보를 검색하고 그 결과를 사용자에게 제시하는 서버릿 기반의 에이전트로 지능형 웹 로봇의 메타검색에 기반하여 문서를 수집하고 분석하여 데이터베이스(SDB)화하는 SR(Search Robot)과, 데이터베이스에 저장된 정보를 사용자에게 제시하기 위한 결과 처리 모듈로 구성된다.

▶ SR(Search Robot)

KGA가 넘겨준 키워드와 카테고리를 가지고 상용 검색엔진들 통해 정보를 수집하고 카테고리별로 정보를 분류하여 SDB를 구성한다. 분류되는 정보들은 각 검색엔진들이 가지고 있는 사이트들의 주소, 제목, 분류, 내용, 기타 정보들이 된다. Fig. 5는 SR의 처리과정이다.

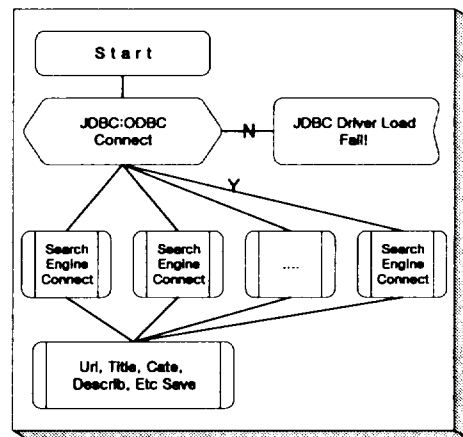


Fig. 5. Processing Step of SR.

▶ RPM(Result Processing Module)

SDB에 저장된 정보들에 대하여 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. 3개의 카테고리를 비교하는 이유는 관심분야의 정보가 UI에서 3개로 정했기 때문이며, 응용에 따라 항목 조정은 가능하다. 그리고

HDB에 결과를 저장하는 이유는 현재의 시점과 과거에 검색되었던 정보들을 추가적으로 제공함과 동시에 KGA에 카테고리 정보를 제공하기 위함이다. Fig. 6은 RPM의 처리과정을 보여준다.

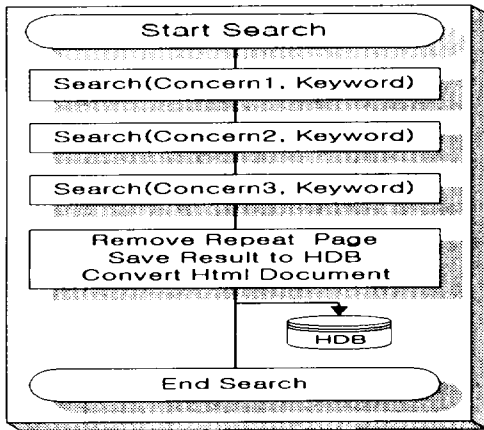


Fig. 6. Processing step of RPM.

2.2. 데이터베이스 설계

사용자의 정보와 질의, 그리고 질의를 통해 연결된 결과와 검색결과를 저장하기 위해서 사용자 데이터베이스(UDB)와 히스토리 데이터베이스(HDB) 그리고 검색결과 데이터베이스(SDB)를 사용하게 되며 그 구조는 Fig. 7, Fig. 8, Fig. 9에서 보여준다.

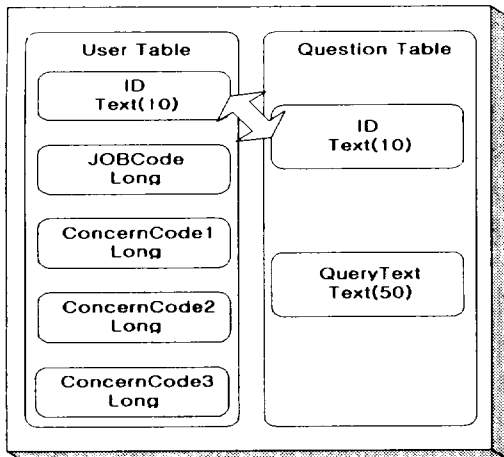


Fig. 7. Structure of user database.

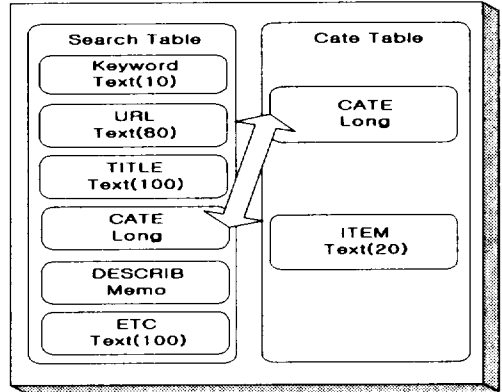


Fig. 8. Structure of history database.

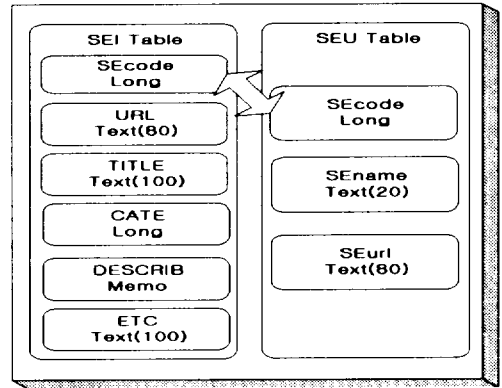


Fig. 9. Structure of search database.

Table 1. Job code

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

Table 2. Concerning code

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(06)	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)	

위의 Table 1과 Table 2는 사용자의 관심분야와 직업등록을 위한 직업 코드표와 관심분야 코드표이다.

III. 시스템 구현 결과

본 장에서는 위에서 제안한 모델을 구현하기 위한 환경과 결과를 기술한다.

3.1. 시스템 구현 환경

- ▶ OS: Microsoft Windows 2000 Server
- ▶ Programming Language : SUN's JDK 1.3, JSP, HTML

▶ Servlet Platform : SUN's JSDK 2.1

▶ Database : Oracle 8i

3.2. 시스템 모듈 구현

3.2.1. 사용자 인터페이스

Fig. 10은 사용자의 정보와 질의내용을 입력받기 위한 사용자 인터페이스 화면이다.

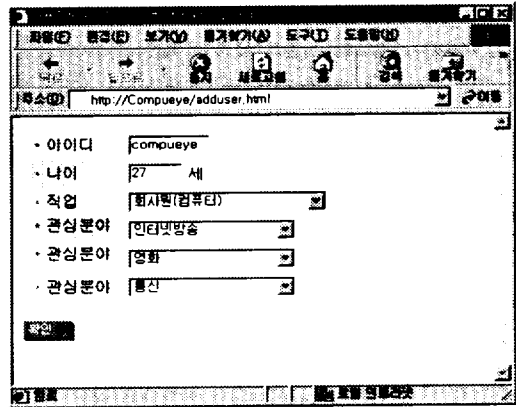


Fig. 10. User interface A.

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.

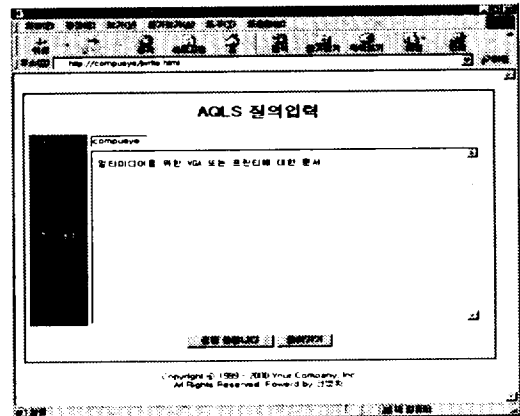


Fig. 11. User interface B.

Fig. 11는 사용자의 질의를 입력받기 위한 화면으로, 사용자의 아이디와 질의내용을 입력받도록 하였다. 입력된 아이디는 'compueye'이고, 질의내용은 '멀티미디어를 위한 VGA 또는 프린터에 대한 문서'이다.

3.2.2. 사용자 질의에 대한 키워드 생성

사용자가 입력한 문장을 통해 검색 시 사용할 키워드를 생성하기 위해 자연어처리를 거쳐 가능한 예비 키워드를 다음과 같이 생성해 내었다.

 자연어 질의 : 멀티미디어를 위한 VGA 또는 프린터에 대한 문서

```
(0)3.956e-87 /complete/end---s<문장시작>()  

(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<~>  

(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<~>  

(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>()
```

 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'가 키워드로 생성되었고, 검색 시 중점키워드로 사용되었다.

3.2.3. 검색결과화면

Fig. 12는 사용자 질의가 등록되고, 질의에 대해 기존의 질의응답방식으로 응답이 가능한 기능과 자동으로 질의에 대한 검색어를 사용하여 검색한 결과를 보여주는 화면이다.

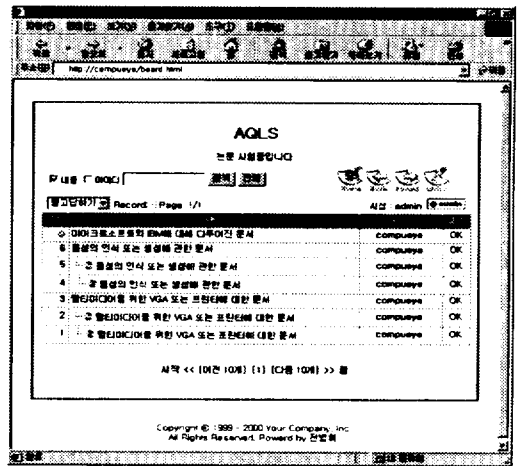


Fig. 12. User interface C.

결과의 내용은 기존의 질의응답방식과 비슷한 화면을 구성하고 있으며, 다른 기능은 질의내용에 대해 자동으로 연결한 페이지의 정보를 담고 있다는 것이다.

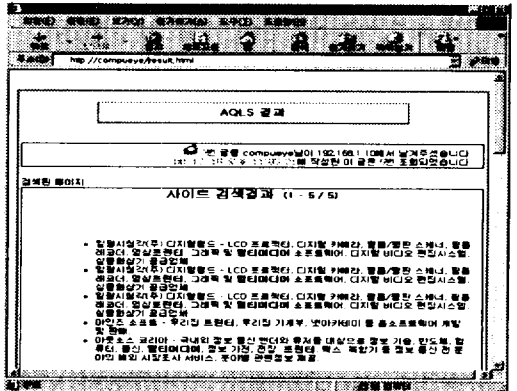


Fig. 13. Search result.

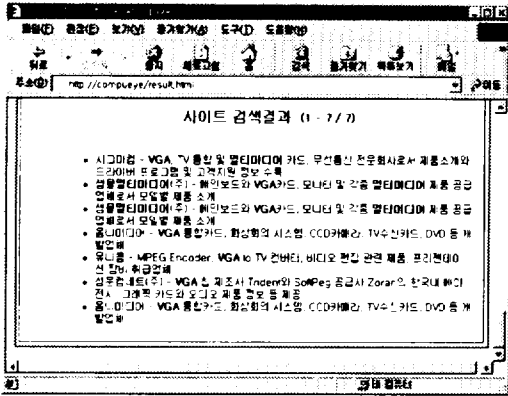


Fig. 14. Search result.

Fig. 13과 Fig. 14는 시스템의 검색 결과를 보여주고 있다.

IV. 결론

기존의 질의응답시스템(Q&A)이 가지는 불필요한 검색 결과와 중복 검색의 단점을 해결하기 위하여, 사용자의 정보에 입각한 검색을 통해 정보를 제공할 수 있는 본 시스템의 구현으로 사용자 관심항목에 관련된 보다 근접한 정보들을 사용자에게 제공할 수 있음을 알 수 있었다. 또한 관심항목이 비슷할수록 히스토리 정보에서 사전에 검색되었던 정보를 풍부하게 제공받을 수 있는 가능성이 많아지게 되며, 히스토리 정보는 검색 빈도수가 많을수록 제공할 수 있는 정보가 계속 누적되게 되어 사용자는 보다 쉽게 원하는 정보들을 제공받는 것이 가능하게 되었다.

하지만 본 시스템은 사전에 사용자의 정보를 유지해야 한다는 단점과 많은 사용자의 정보들을 효율적으로 관리하기 위한 방법들에 대한 연구들이 병행되어야 하며 검색 시 사용하는 키워드 생성 시 한글 자연언어 처리부분과 에이전트 시스템의 표준화에 대한 연구들이 이루어져야 한다고 본다.

참고문헌

- 1) 전용휘, 1996, "인터넷 정보검색," WWW-KR.
- 2) 김병학, 이광형, 조충호, 1996, "지능형 웹 브라우저 에이전트, HCI.
- 3) Stan Franklin, Art Graesser, 1996, "Is it an Agent, or Just a Program?", A Taxonomy for Autonomous Agents. ATAL pp.21-35
- 4) 정영미, 1990, "우리말 정보자료를 처리하는 지능형 정보검색시스템의 설계", 연세대.
- 5) Charles J. Petrie, 1996, "What is An Agent?", ATAL pp.41-43
- 6) Thomas Magedanz, Kurt Rothermel, Sven Krause, 1996, "Intelligent Agents: An Emerging Technology for Next Generation Telecommunications?", INFOCOM pp.464-472
- 7) M. Wooldridge, N. Jennings, 1995, "Intelligent Agents", Theory and practice. Knowledge Engineering Review, Vol.10, No.2.
- 8) Pattie Maes, 1994, "Agents that Reduce Work and Information Overload", CACM Vol.37, No.7, pp.30-40.