

종합적 기능 포함한 다기능 에이전트 출현



현재 웹 상에는 많은 검색 엔진들이 존재한다. 최근에는 Yahoo나 Altavista와 같은 검색 엔진 이외에도 CD나 책 등 특정 도메인에 한정된 정보만을 찾아주는 전문 검색 엔진들이 늘어나고 있다. 쇼핑몰, 포털 서비스 제공 사이트 등 대부분의 사이트에서 기본적으로 정보 검색 기능을 가지고 있다. 이는 정보 검색이 인터넷에 접속하는 사람들의 가장 큰 관심사가 되는 작업임을 입증한다. 이 글에서는 정보 검색 분야에서 에이전트 관련 기술들이 어떻게 적용되고 있는지를 알아보기로 하겠다.

■김영민/ (주)쓰리소프트 연구소 연구원

정

보 검색 도구로서의 검색 엔진은 이렇듯 인터넷에서 늘어나는 정보 속에서 자신이 원하는 정보를 찾아내기 위해 사용되고 있다. 그럼, 이러한 정보 검색 사이트를 방문하는 사람들의 행동 양식을 살펴보자. 무언가를 찾고자 하는 사용자는 검색 엔진이 구축된 사이트에 들어가서 원하는 검색어를 입력하게 된다.

그러면 그 검색어에 매칭되는 결과 사이트들이 URL링크들의 리스트로 나타나고, 이 결과 사이트들의 링크를 사용하는 자신이 원하는 정보를 찾을 때까지 돌아다닌다(Navigate). 이런 반복된 과정을 통해서 사용자는 자신이 원하는 정보를 찾으려 할 것이다.

그러나, 연산자를 잘 사용하고, 여러 가지 검색 옵션을 사용할 줄 아는 사용자라 할 지라도(실제로 하나의 연산자 이상을 사용해서 검색을 하는 사용자는 전체 사용자의 10% 미만으로 극히 드물다), 나온 결과의 양이 많아지면 결과링크를 따라다니는데 어려움을 겪게 되고, 힘들게 따라간 링크가 자신이 원하는 분야와는 전혀 별개인 경우도 발견하게 되어 검색을 포기해 버리기도 한다. 인터넷을 돌아다니며 검색을 해 본 사용자라면 누구나 이런 경험을 해보았을 것이다.

이렇듯 웹 검색엔진을 통한 정보 검색 기능에 만족하던 사용자들이 점차 불편을 느끼게 되었다. 이렇게 된 원인은 무엇보다도 인터넷의 장점이자 단점이 되는 그 개방적인 성격에 있다. 누구나 정보를 손쉽게 보내고 받을 수 있음으로 해서 정보의 양은 상상을 초월하게 되었고, 질의 연산자를 이용한 단순한 검색 기능만으로는 사용자의 요구를 만족시킬 수 없게 되었다.

이에 최근의 사이트들은 정보를 세부 분류까지를 보여주어서 필요 없는 도메인에 대해서는 결과가 나타나지 않도록 하여 유효하지 못한 결과 정보의 양을 줄인다거나, 개인화된 정보를 제공한다는 표어를 내걸고 사이트를 방문하는 사람의 선호도에 따라 메뉴를 바꾸어서 보이게 한다거나, 개인에게 특화된 정보를 보여주어서 자신만을 위한 사이트, 자신의 홈페이지인양 보이게 하는 서비스를 하는데 주력하고 있다. 이런 사이트들은 또한 무료로 개인 홈페이지나 이메일 계정을 주어서 그런 서비스를 하고 있는데, Yahoo의 Directory Service나 MSN의 초기화면은 이런 서비스를 제공하는 예라 할 수 있다.

정보 검색 분야에서도 이러한 사용자의 요구에 부응하기 위해서 정보 검색을 위한 기본 기능이외에도 에이전트라 불리는 AI의

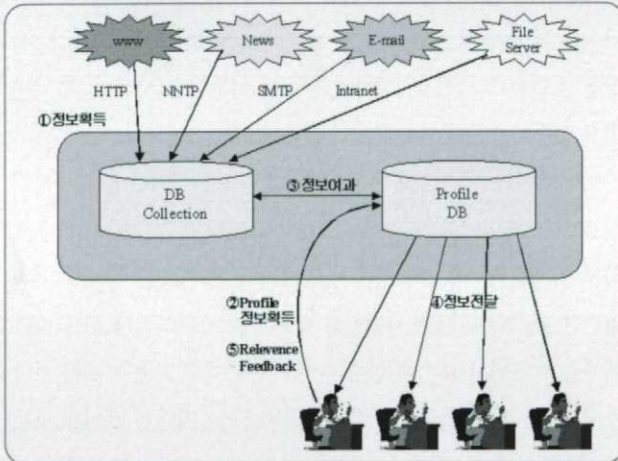
새로운 응용 기술들이 응용되고 있다. 이러한 변화는 최근의 클라이언트 요구사항의 변화를 보아도 쉽게 알 수 있다.

예전에는 검색 엔진의 기본 기능을 얘기할 때, 검색 엔진사이의 비교시에 boolean 검색이 되느냐, wildcard 검색이 가능한가?, 구검색이 되는가, 시소러스는 지원하는가 등의 검색 질의 연산자에 대한 것이 대다수 였는데 반해, 최근에는 검색 결과를 사용자에게 자동으로 알려준다던지, 검색 대상이 변화되었을 때 그 변화된 결과를 다시 사용자가 질의를 던져서 검색해 보지 않더라도 자동으로 변경사항을 알려주거나, 정보 획득 시에 검색을 위한 정보원이 자동으로 분류될 수 있는가에 대한 요구사항이 늘어나고 있다.

에이전트는 그 정의가 바라보는 관점에 따라서 무척이나 다양하며, 그에 따라 에이전트의 분류도 다양하다. 따라서, 이러한 프로그램이 에이전트나 아니냐를 어떤 특정한 정의를 통해 규정하기도 힘들어지는 것이다.

앞으로 이 글에서 전개해 갈 내용은 정보 검색분야에서의 에이전트 기술에 대한 것으로 한정해서 설명할 것이며, 에이전트 응용 분야는 많아서, 필자는 몇 가지의 예와 기술만을 설명해 갈 것이다. 설명해 나가는 중간 중간에는 그 분야에서 상용화되어 사용되는 제품에 대한 간단한 설명과 비교도 할 것이다. 그리고 결론에서는 필자 나름의 개선점과 나아가갈 방향에 대해 얘기해 보겠다.

그럼 일단 세부 내용에 들어가기 전에 아래의 <그림 1>을 보자.



<그림 1> 정보 검색에서의 에이전트 부가 기능

이는 정보 검색의 핵심인 검색 엔진의 기능에 에이전트의 기능들이 첨가된 개괄적인 그림이다. 이 그림을 기준으로 앞으로의 세부 사항을 기술해 나가겠다.

일단, 정보 검색을 하기 위해서는 정보의 원천에서 정보를 모아

오는 기능이 필요하다.

① 과정이 바로 이러한 정보 획득의 기능을 한다. 이 과정의 기능이 중심이 되어 제작된 에이전트를 흔히 Robot Agent라 부른다. 이렇게 획득된 정보는 적절한 저장 장소에 저장되게 된다 (저장 매체에 따라 달라지겠지만, DB에 저장되거나 검색 엔진에서는 collection이라 불리는 색인된 정보를 가지는 저장구조로 구축된다). 정보원에서 뿐이 아니라, 정보를 사용할 사용자에게서도 필요한 요소들을 뽑아온다. ②의 과정이 이러한 모습을 보인 것이다.

②의 과정을 통해 뽑아낸 정보가 '사용자 Profile'이라는 이름으로 저장된다. 이 프로파일에는 정의하기에 따라 여러 가지 요소가 정보로서 들어갈 수 있겠다. 예를 들어, 사용자의 나이, 직업, E-Mail 등이 그런 것이 될 수 있다. ③에서는 ①과 ②의 과정에서 축적된 정보를 매칭시키는 역할을 한다. ③을 매칭되고 여과된 정보는 ④ 과정을 통해 사용자에게 전달된다.

④ 과정을 통해 전달된 정보는 사용자의 선호도에 따라 피드백(feedback)정보를 받아서 다시 프로파일(profile)에 반영된다. 이를 나타내는 것이 ⑤번 과정이다. 이 그림은 필자가 기능을 위주로 그려본 것으로 실제로 이러한 제품이 구현되어 있는 것은 아니다.

<그림 1>을 다시 보자. 기존의 웹을 기반으로 하지 않는 단순한 정보 검색은 정보가 한 번 입력되고 나면 그 정보 대상 자체가 변하는 경우가 극히 드물었고, 관리자가 일일이 입력 폼을 통해 정보를 입력해주어야 했다. 그러나, 인터넷(Internet) 특히 웹(WWW)의 등장으로 인해서 정보의 양이 몇 명의 관리자가 입력하기에는 불가능한 수준으로 증가해 버렸고, 인터넷 상에서는 정보가 수시로 변화하므로 이것을 관리자가 알아내 변경하기란 불가능한 일이 되었다. 이에 이런 기능을 해줄 수 있는 프로그램이 필요한데, 이런 필요에 의해 만들어진 에이전트가 바로 로봇 에이전트이다.

로봇 에이전트

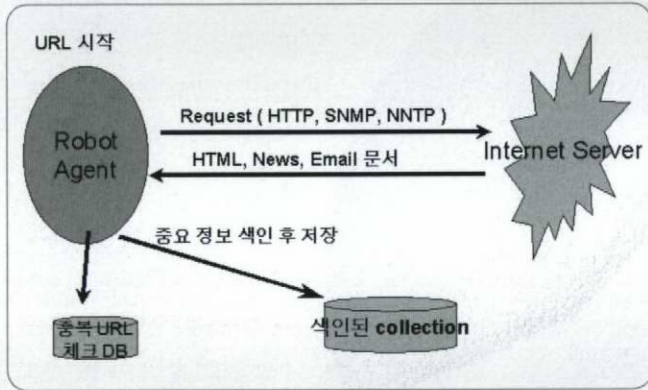
로봇 에이전트는 도메인을 웹이라고 한정하면, 웹 서버를 돌아다니며 각 사이트에 있는 정보를 수집하는 프로그램이다. 로봇 에이전트는 홈페이지의 내용을 분석하여, 그 안의 URL을 추출해 다른 URL로 연결시켜 주는 기능을 그 주요 기능으로 한다.

따라서, 로봇 에이전트를 실행시키면 자동으로 처음 URL과 링크된 다른 URL들을 차례로 따라가면서 돌아다니게 된다. 가령

웹 서버에 담긴 내용이 필요한 것이지만 링크된 문서가 많고, 그 문서 안에 수많은 이미지도 필요한 것이라고 할 때, 모니터 앞에 앉아 하나씩 저장할 수는 없다.

또 자신이 관리자로서 관리하는 홈페이지의 링크가 변화되었는지 변화되지 않았는지를 홈페이지의 링크를 하나하나 따라가면서 URL의 변경 사항을 체크할 수는 없다. 로봇 에이전트는 이와 같은 일을 웹 페이지를 돌아다니면서 자동으로 수행한다. 웹 상의 검색 엔진도 사람이 웹 사이트를 찾아다니면서 인덱스를 만드는 것이 아니라, 로봇 에이전트가 자동으로 돌아다니면서 정보를 수집하고 이렇게 모아 놓은 인덱스를 바탕으로 검색 엔진이 검색을 하게 되는 것이다.

다음 <그림 2>는 이런 로봇 에이전트의 기본 동작 구조를 나타내고 있다.



<그림 2> 로봇 에이전트의 기본 동작 구조

<그림 2>와 같은 방식이 기본적인 Robot Agent의 동작 유형이다. Internet Server(예를 들어, Web Server, News Server, Email Server 등)에 문서에 대해 프로토콜의 형식에 맞춰 Request를 작성해서 네트워크를 통해 보내면, Server는 이에 해당하는 문서를 Response로 보내온다. 이 때, 중복된 문서(Web Server에 Request를 했을 경우에는 중복된 URL)를 체크하고, 새로운 문서이면 파싱해서 링크된 URL들을 뽑아낸 후, 색인(Indexing)과정을 거쳐 Collection으로 구축한다. 이 과정을 반복하는 것이 Robot Agent의 주된 작업이다. 앞의 그림에서 알 수 있듯이 로봇 에이전트의 구현에는 네트워크 처리 기술, 프로토콜 처리 기술, 문서 Parsing기술이 필요하다. 이 때, 프로토콜과 문서의 파싱은 정보원이 어떤 형식이냐에 따라 달라진다.

로봇 에이전트를 이용함으로써 원하는 정보를 쉽고 빠르게 찾을 수 있다는 긍정적인 측면이 있는 반면, 네트워크나 상대방 웹

서버의 부하 가중 등 약기능이 발생하기도 한다. 따라서, 로봇 에이전트에 대한 인식은 물론, 올바른 제작이나 사용에도 관심을 가져야 하는데, 로봇 에이전트의 문제점을 개선하기 위한 표준으로 Robot Exclusion이라는 제안이 나오기도 했다. 그러나 이 표준은 단지 제안일 뿐이라서, 이를 무시하고 작성한 로봇 에이전트가 들어오는 것을 서버 관리자가 처리할 수 있는 방법은 실제로는 없는 실정이다.

여과 에이전트

두번째로 살펴볼 것은 정보 여과 에이전(Information Filtering Agent)이다.

정보 검색(Information Retrieval)과의 비교를 위해 Agent라는 말을 빼고 Information Filtering이라고 쓰겠으며, 이를 Profiling이라는 단어와 혼용해서도 사용할 것이다.

서두에서 설명했던 정보 검색의 기본 기능을 다시 한 번 살펴보면, 정보 검색과 프로파일링(Profiling)기반의 정보 여과와의 비교를 통해 정보 여과 에이전트의 기능에 대해 설명해 보고자 한다. 정보 검색은 일단 사용자의 요구에 대한 처리가 사용자의 검색어에만 의존한다.

즉, 모든 사용자를 동일시해서 각 개인을 구별할 수 있는 아무런 수단도 가지고 있지 않다. 이에 반해 사용자 Profile을 이용한 정보 여과는 사용자의 Profile을 가지고 있으므로 해서 적절한 시간에, 사용자에게 유효한(relevant) 정보를 전달할 수 있다. 정보 여과 에이전트는 이런 여과 과정외에 사용자에게 전달되는 방법과 피드백과정까지를 포함한다. 골프를 좋아하는 누군가가 '골프'라는 단어로 검색을 해보았다면, 기존의 '골프'라는 검색어만을 위주로 검색결과를 내놓는 것보다, 그가 어느 나라의 사람인가? 어떤 연령의 사람인가? 에 따라서, 골프 정보중 그 나라에 맞는 정보, 그 연령들이 좋아하는 정보들을 제공해 주며, 이를 자신이 원하는 시간에 제공 받을 수 있는 것이다.

점점 더 많은 사용자가 온라인 상에 접속됨으로 인해서, 사용자들은 어디서 어떻게 정보를 얻어야 할지를 모르게 된다. 기존의 정보 검색은 정보를 검색어를 통해 매칭되는 과정을 통해 이러한 일을 하고자 하는 방식이다.

앞의 예에서 보듯이 정보 검색은 유효하지 못한(irrelevant) 정보를 제거할 수 있는 기능이 떨어진다. 따라서, 정보 여과(Information Filtering)에서는 사용자의 피드백(Feedback)을 통해 유효한(relevant) 정보의 양은 늘리고, 유효하지 못한

(irrelevant) 정보의 비율을 줄이고 있다.

또한 정보 검색의 경우에는 그 순간 사용자가 입력한 검색어에 기반하는 짧은 기간(Short Term)의 목적을 달성하는데 반하여, 정보 여과(Information Filtering)란 오랜 시간(Long Term)을 통해 사용자와 상호 작용한다. 이렇게 함으로써 사용자의 선호도를 알 수 있고, 개개 사용자를 기억할 수 있게 된다.

정보 검색에서는 모아진 정보와 비교될 수 있는 것인 단순한 검색어였는데 반해, 앞의 <그림 1>에서 보듯이 정보 여과(Information Filtering)에서는 검색어 외에도 사용자의 각종 정보가 정보를 찾고, 찾아진 정보에서 필요한 정보를 걸러내는 데 사용됨으로써 좀더 나아진 정보를 사용자에게 제공할 수 있다.

이렇게 사용자 profile을 통해 걸러진 정보는 사용자에게 전달되고 피드백(feedback)을 통해 다시 profile을 변경하는데 사용된다. 정보 여과(Information Filtering)에서는 처음에는 사용자에게 적합한 정보가 제공되지 못할 수도 있으나, 시간이 지남에 따라, 점점 사용자의 선호에 맞도록 변화되어 갈 수 있는 것이다.

덧붙여 사용자 profile을 통한 정보 여과의 장점을 말하면, 자신과 성향이 비슷한 사용자의 그룹 프로파일을 통해 자신이 바로 원하던 않았지만 비슷한 경향의 다른 이들이 선택한 것들에 대한 정보를 공유할 수 있다. 이런 서비스를 제공하는 사이트로는 FireFly가 대표적인데, 웹 상의 CD를 판매하고, 성향이 비슷한 그룹의 사람들에게 그 그룹이 많이 구매한 CD의 리스트를 보여주고 있다.

정보 여과 에이전트(Information Filtering Agent)에 해당하는 상용화 사례를 구체적으로 두 가지만 살펴보고 앞에서 설명한 기능들이 어떻게 구현되었는지를 잠시 살펴보자. 많은 에이전트 관련 제품이 있겠지만, 여기서는 Tympani Development사의 NetAttache Pro라는 제품과 Autonomy사의 new Media에 대해 살펴보겠다.

NetAttache Pro의 주요 기능은 다음과 같다.

- 사용자의 로컬 디스크에 웹 정보를 저장 오프라인 검색(한가한 시간에

사용자가 원하는 페이지를 다운로드하고 저장, 웹 페이지의 중요한 정보가 변경 후 다시 검색하고 재차 방문할 필요가 없는 곳은 접속하지 않음)

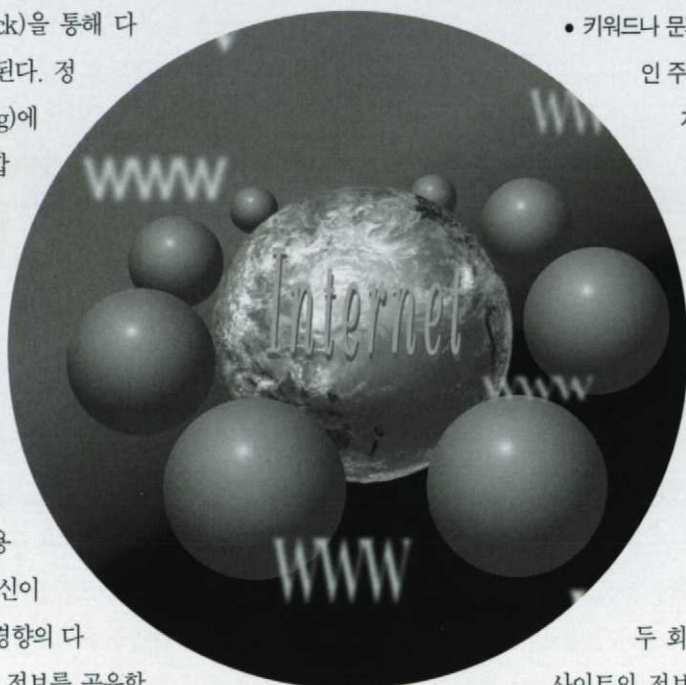
- 키워드 혹은 문장으로 검색 가능
- 다운로드 기능과 연계
- 웹 데이터베이스 파일 보관, 이들 파일의 전송 용이, 이동이 잦은 업무 환경에 효율

이에 비해 Autonomy의 New Media 제품은

- 사용자와 정보 제공자간의 대화기능의 확장
- 사용자가 정보를 탐색하는 대신, 에이전트가 사용자의 관심사를 파악하여 관련 웹 정보를 자동으로 접속

- 키워드나 문자열 수준이 아닌 웹 정보의 포괄적인 주제나 정보의 패턴을 이해, 개념적인 지식을 개발

- 사용자의 정보검색 행위를 통해 내재된 관심사를 관찰
- 사용자의 언어적 명령관습으로부터 trainnig, 관련 정보 자료를 미리 접속
- 인터넷이나 인트라넷 정보를 고용자나 웹사용자에게 개별화된 형태로 지능적으로 모니터링하고 전달



두 회사의 제품은 웹 사용자가 특정 사이트의 정보에 접속하기 위해 정기적으로 방문해야 하는 시간적 낭비를 개선하고, 중요한 정보에 접속하는 시간을 최소화 하기 위한 것을 모토로 하고 있는데, 둘의 차이점은 Tympany의 웹 에이전트는 오프라인 브라우저에 기초하고 있는 반면, Autonomy사의 new Media 제품군은 고수준의 지능을 갖은 온라인 정보 접속 시스템이라 할 수 있다. 그 기능을 보면 앞에서 설명한 사용자의 정보를 이용하여 개개인에게 특화된 정보를 제공하는 것을 그 기본 기능으로 하고 있음을 알 수 있다.

푸시 에이전트

마지막으로 푸시 에이전트(push technology agent)에 대해 언

급하고자 한다. 푸시 기술만을 가진 것을 에이전트라고 부를 수 있느냐고 반문하는 독자도 있겠지만, 최근의 에이전트에 대한 연구의 초기 형태는 바로 푸시 기술이 웹 상에서 그 관심이 증가되면서 시작되었다고 할 수 있으므로 에이전트 기술의 하나로 언급하고자 한다.

(그림 1)에서 4번 과정을 살펴보면 사용자에게 필요한 정보가 전달되는 과정을 보인 것이다. 푸시 에이전트란 이렇게 사용자에게 전달되는 과정을 자동화한다는데 초점을 맞춘 제품이라고 할 수 있겠다. 사용자가 일일이 자신이 필요로 하는 정보를 검색 엔진을 찾아갈 필요없이 사용자에게 자동으로 내려 보내 주는 것이다.

푸시 기술은 정보가 제공받고자 하는 사용자에서 미리 선택되어지고, 정해진 시간에 사용자에게 배달되는 방식이다. 이러한 정보의 푸시는 그 방향이 서버에서 클라이언트로 향하며, 푸시를 웹 상의 브로드캐스팅이라고 부르는 사람도 있으나, TV나 라디오와 같은 브로드캐스팅과는 차이를 보이는데, 푸시가 일어나는 시점에서 콘텐츠를 전달받는 사용자들이 불특정 다수가 아니라 한정되어 있다는 점에서 브로드캐스팅과는 성격을 조금 달리한다.

푸시 기술은 96년도 포인트 캐스트사가 인터넷에 공개한 뉴스 제공 서비스용 클라이언트인 '포인트 캐스트 브라우저'와 '스마트 스크린 세이버'로 처음 선보였다. 포인트 캐스트가 선보인 이 기술은 인터넷의 새로운 응용 가능성을 제시하게 되었다. 포인트 캐스트의 등장 이후, 유사한 기술을 사용하는 서비스가 무료 30여종이 등장했다. 현재는 조정기에 들어가서 일부는 마이크로 소프트웨어사의 인터넷 익스플로러에 통합되기 시작했으며, 자바를 이용한 푸시 기술을 선보인 마림바 등이 넷스케이프의 차세대 버전인 콘스탈레이션에 사용될 것이라 한다.

그러나, 여기까지 설명한 푸시 기술 자체는 지능적이지 못하다. 10%의 유익한 정보를 위해서 90%의 비유익한 정보를 함께 받아오는 것은 네트워크의 트래픽을 증가시킨다는 점과 다른 한편으로는 비유익한 정보들 중에서 유익한 정보를 찾아야 한다는 점에서 지극히 낭비적인 면이 없지 않다. 따라서, 최근에는 푸시 기술에 사용자가 필요로 하는 정보만을 선택할 수 있는 풀(Pull) 기능이 접목되고 있다.

푸시 기술이 보다 지능화 될 경우, 푸시 클라이언트 내부의 사용자 에이전트가 사용자의 사용추이, 접근 빈도 등을 분석해서 이 정보들을 근거로 사용자에게 보다 편리하게 정보를 제공할 수

있다. 즉, 푸시 기술은 앞의 출발은 사용자에게 정보 전달의 수단을 위해서 시작되었지만, 점차 앞에서 설명한 정보 여과(Information Filtering)의 기능이 첨가되어 에이전트의 모양을 갖추고 있다.

다가는 제품으로 발전

이제까지 정보 검색에서 쓰이는 에이전트의 큰 그림을, 그리고 각 기능들에 해당하는 것을 중심으로 설명해 보았다. 이 기능들은 정보 검색이라는 기본 기능을 보완하는 정보 획득, 정보 여과, 정보 전달, 피드백을 그 주요 기능으로 나누어 볼 수 있다. 최근의 제품들은 이런 기능들이 통합되어서 나타나는 형태를 띤다.

로봇 에이전트는 앞으로 표현만을 위한 언어였던 HTML 한계를 극복한 XML이 표준이 되고 있으므로 조만간 XML을 이용한 정보가 쏟아져 나온다면 기존의 정보 획득과정에서 좀 더 발전된 모습을 보일 수 있을 것이다. 그리고 정보 여과 에이전트 분야에서는 사용자의 Profile과 매칭되는 정보를 찾는 알고리즘으로 키워드 매칭 이외에 유전 알고리즘(Genetic Algorithm)과 같은 최적 해(Optimal Solution)를 찾는 AI의 알고리즘(Algorithm)들이 적용되고 있다.

이런 정보의 여과 과정에서 고려해보아야 할 문제가 하나 있는데, 그것은 개인의 프라이버시(Privacy)문제이다. 정보 여과를 효과적으로 하기 위해서는 좀 더 세세한 사용자의 정보가 요구되고 이것은 개인의 사생활 침해라는 심각한 문제를 야기할 수도 있기 때문이다.

따라서, 에이전트 시스템을 개발하고 개인의 프로파일을 이용할 경우에는 사용자에게 미리 경고를 하는 것을 잊지 말아야 한다. 개인의 정보를 다른 용도로 사용하지 않을 것임을 미리 주시켜 사용자에게 안정성을 미리 얘기할 필요가 있음은 물론이고, 사용자에게 자신의 정보를 보내지 않을 수 있는 장치 또한 제공해 주어야 한다.

에이전트는 정보 검색에 부가된 기능을 함으로써 여러 가지 편의를 제공해 줄 수 있을 것이다. 예를 들어, 문화나 공연에 관심을 갖는 사용자의 경우에 새로운 공연이나 콘서트가 열릴 때마다 사용자가 좋아하는 공연을 추천해 준다든지, 자주 보는 기사나 섹션들로 재구성해 줌으로써 사용자가 보다 편리하게 내용을 검색할 수 있게 될 것이다. 