

퍼지 웹 정보 관리 에이전트

김병학*, 조승호**, 이광형*

* 한국과학기술원 전산학과

** 고려대학교 전산학과

요 약

현재의 인터넷에는 방대한 양의 정보가 있다. 그리고 이 정보는 시간이 지남에 따라 계속 증가하고 있는 추세에 있다. 이러한 인터넷 환경에서 사용자가 자신의 정보를 관리하고 변화에 대처하기 위해서는 보다 편리하고 지능을 가진 에이전트의 기능이 필요하다고 할 수 있다. 본 논문은 웹 브라우저에서 사용되는 Bookmark, History 등의 정보 관리 기능을 확장하여 에이전트에 의해서 자동으로 관리되는 정보 관리 시스템을 제안 하였다. 이 시스템은 사람이 하는 것과 비슷한 정보 관리 기능을 위하여 퍼지 이론을 도입하여 사용자가 정보에 대한 중요도를 계산하는 방식으로 정보에 대한 중요도를 계산한다. 또한 사용자의 편의를 위하여 사용자에게 중요한 정보라고 판단되는 페이지에 대해서 자동으로 변경 여부를 검사하고 이를 캐싱할 수 있게 하였다.

1. 소개

현재의 인터넷은 정보의 바다라고 불리울 만큼 많은 정보를 가지고 있다. 시간이 갈 수록 이 정보의 양은 기하 급수적으로 증가하고 있다. 현재의 웹 브라우저 환경에서 이 큰 정보의 바다에서 자신이 원하는 정보를 찾기란 굉장히 어려운 일이다. 또한 자신이

찾은 정보에 대한 관리를 하는 일도 굉장히 힘들어 지고 있다.

본 논문에서는 사용자가 원하는 정보에 대한 찾기와 또한 자신이 얻은 정보에 대한 관리를 효율적으로 할 수 있게 도와주는 사용자 에이전트에 대한 모델을 제시하고 이를 구현하여 보다 편한 정보 사용자 환경을 제시하고자 한다.

실제의 세상은 퍼지하다고 볼 수 있다. 전문가들의 판단이나 지식은 정확한 한계가 분명하지 않는 경우가 많으며 실제로 퍼지 이론은 실세상의 지식과 판단에 적용되어 많은 응용에서 유용하게 사용 되어지고 있다.

소프트웨어 에이전트는 지능과 스스로의 수행 능력을 지니면서, 실세상에 대한 모델을 바탕으로 하고 있기 때문에 많은 분야에 적용되어 지고 있다. 에이전트도 기존에 존재하는 지식과 판단 규칙에 의존하고 있다고 할 수 있으며, 이러한 측면에서 퍼지 이론의 도입은 좋은 에이전트의 지능 모듈로 적용될 수 있다.

본 논문에서는 현재의 사용자들이 일상적으로 관리하는 정보에 대한 관리 모델을 바탕으로 퍼지 이론을 적용하여 이를 지식 베이스로 구축한다. 이 지식베이스는 사용자의 행동에 의해 스스로 재 구성된다. 퍼지 에이전트는 사용자의 행동을 관찰하고 이를 입력으로 이용하여 사용자가 정보를

관리하듯 중요한 정보와 중요하지 않는 정보를 구분하여 저장한다. 또한 중요하게 판단된 정보에 대해서는 사용자의 입력이 없는 기간에 이를 가져와서 변경 사항의 유무를 점검한다. 이 과정을 통해서 웹 브라우저에 캐싱이 되며, 이를 통해서 사용자는 원하는 페이지를 더 빨리 볼 수 있다. 본 연구에서는 웹에서 사용되는 Bookmark와 History의 관리를 담당하며, 사용자의 관심에 대한 페이지를 미리 가져와서 문서의 내용과 사용자의 관심과의 유사성을 계산하고, 이를 캐시에 저장되게 하여서 사용자가 인터넷을 사용하는데에 편리하고 빠른 사용을 할 수 있게 한다.

2장에서는 정보 관리 에이전트를 위한 요구 사항을 살펴보고, 3장에서는 모델을 제시하고 4장에서는 이의 구현에 대해서 이야기한 후 5장에서 결론을 맺는다.

2. 요구 사항 및 모델

본 논문에서 고려하는 정보 도구를 사용하는 사용자의 특성은 다음과 같이 요약 된다.

- 사용자는 자신의 관심 분야를 갖는다.
- 사용자의 관심 분야는 때로 갑자기 생성되거나 사라지기도 한다.
- 사용자가 자주가게 되는 페이지는 정보의 양과 질, 그리고 페이지 디자인이나 기타의 서비스에 의한 취향에 의해서 주로 결정된다.
- 사용자는 자신의 관심 분야에 대한 페이지를 많이 보는 경향이 있기 때문에 이런 페이지에 대해서는 에이전트가 지속적인 캐싱과 더불어 페이지의 정보의 확장 여부를 점검해야 한다.

사용자는 정보의 양과 질, 그리고 자신의 관심 분야, 이 세가지 것에 의해 자신의 정보의 유용성을 따진다고 볼 수 있다.

정보의 양이라는 것은 파일의 크기라고도 볼 수 있으며, 질이라는 것은 페이지 디자인의 취향과 자신의 관심 분야와의 관련성등에 의해서 결정되어 진다고 볼 수 있다. 또한 사용자의 관심은 갑자기 새로운 분야에 대해서 생길 수 있기 때문에 어느날 갑자기 새로운 페이지에 대한 접근을 많이 하게 되는 경우도 발생한다.

본 논문에서 이러한 사용자의 특성을 고려하여 다음과 같은 요구사항을 고려하였다.

사용자는 자신의 관심분야를 입력할 수 있어야 한다. 곧 사용자는 에이전트와의 대화를 통하여 자신의 관심사항을 이야기 할 수 있어야 한다. 에이전트는 이를 바탕으로 사용자가 원하는 페이지를 가져올 때마다 이를 점검하여, 자주 가져온 페이지라던지 정보가 사용자의 관심과 아주 유사하다던지를 판단해서 중요 정도를 계산하게 된다.

정보의 양을 점검하기 위해서는 방문한 URL의 정보의 크기 곧 파일의 크기를 검사한다. 또한 사용자가 머무른 시간을 점검하여 적당한 시간, 읽어 본다면지의 활동이 일어날 수 있는 시간을 미리 점검할 수 있어야 한다.

정보의 질에 있어서는 사용자의 관심 분야와 페이지와의 유사도를 계산하고 사용자가 방문한 횟수나 방문 주기등을 고려하여 정보의 질을 판단한다. 사용자마다의 특성이 다르기 때문에 페이지 디자인등에 관련된 것은 판단에 사용하지 않는다. 대신 사용자의 취향은 방문 횟수나 주기에 의해서 결정되어 진다.

이를 위해서 사용자가 한 행동에 대한 로깅이 가능해야 한다. 곧 사용자가 어떤 페이지를 방문한 횟수나 방문한 주기 등은 기억되어야 하며, 방문한 곳의 관심분야와의 관련성도 조사 되어야 한다. 또한 파일의

크기와 방문해서 머무른 시간등을 점검할 수 있어야 한다.

웹 에이전트는 사용자의 관심 분야에 대한 자동적인 가져오기와 이를 통한 퍼지 지식 베이스의 변경이 필요하다. 이렇게 함으로써 사용자는 캐시에 저장된 정보를 봄으로써 빠른 접근이 가능하고 에이전트가 관리하는 지식베이스에도 빠른 정보 변경이 가능하다.

마지막으로 에이전트의 가장 중요한 특성중의 하나인 Autonomous의 특성을

만족하는 에이전트가 되게 하기 위해서 본 논문에서는 사용자의 웹 브라우저는 항상 동작한다고 가정한다.

3. 모델

본 논문에서 고려하는 모델은 기존의 대부분의 웹 에이전트라고 불리는 탐색 엔진이 아닌 사용자측면에서의 에이전트이다. 그림은 아래와 같다.

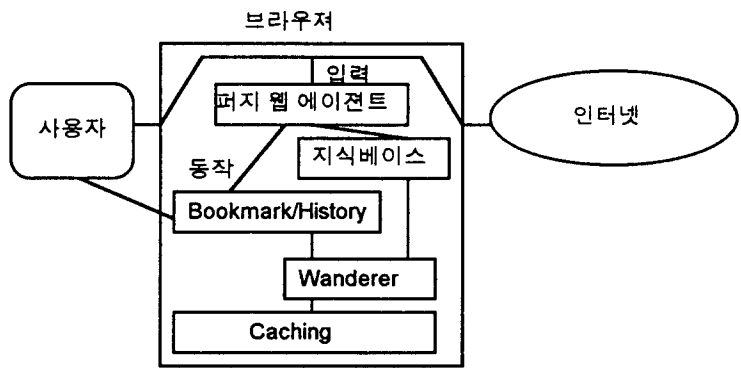


그림1. 웹 에이전트 모델

전체적으로 크게 두부분으로 구분되는데 하나는 사용자의 입력, 곧 웹 공간을 돌아다니는 사용자의 입력에 의한 부분이고 다른 하나는 에이전트가 자동으로 사용자의 관심 분야를 가져와서 이를 캐싱하고 또한 지식베이스에 정보를 변경시키는 부분이다.

4. 구현

우선 사용자의 입력을 받아서 이를 처리하는 퍼지 정보 관리 에이전트는 다음과 같은 구조를 갖는다.

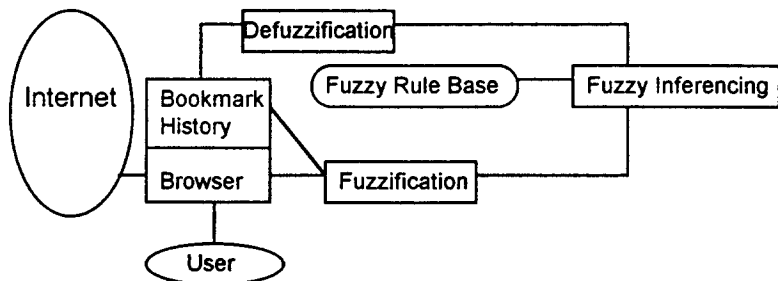


그림 2. 퍼지 웹 에이전트

사용자의 웹 사용은 곧 바로 퍼지 입력으로 작용한다. 여기에서의 입력은 URL과 기존의

저장된 방문 횟수, 방문 주기, 최고 방문 주기, 관심 분야와의 유사도, 정보의 양등의 기준

정보와 사용자의 입력이다. 이때 입력은 현재 사용자가 선택하여 도착한 URL에 대한 데이터이다.

이렇게 입력된 값들은 우선 퍼지화과정을 거치고 이후 퍼지 규칙에 의해서 결과값을 낸다. 그 이후 비퍼지화 과정을 거쳐서 Bookmakr와 History가 수정된다.

자동으로 웹 공간을 돌아다니는 에이전트는 사용자의 입력이 오래동안 없을 경우에 동작이 실행된다. 이 에이전트는 사용자의 Bookmark에 등록된 중요한 URL에 대해서 방문을 하게되며, 만약 페이지가 Update되어 있으면, 이를 가져온다. 그리고 이 정보는 캐시에 기록된다. 그리고 새로이 가져온 페이지에 대해서 웹 에이전트는 사용자와의 관심 정도를 계산하게 된다.

4.1 입력 변수

퍼지 웹 에이전트에 입력으로 들어가는 변수는 사용자의 관심분야를 알고 이를 이용하여 사용자의 정보 이용 행동을 분석하여 사용자의 행동에 따른 중요한 페이지와 안중요한 페이지간의 분리등의 일을 자동으로 해주어야 한다. 본 연구에서는 퍼지 이론을 이용하여 사용자의 행동에 대한 저장과 이의 분석을 통한 중요한 정도의 계산을 수행한다.

퍼지변수는 사용자가 선택한 페이지가 바뀔때마다 새로이 입력된다. 이 변수들은 대부분 최고치에 대한 상대값을 가지는데, 이를 위해서 전처리 과정이 필요하다.

다음은 본 논문에서 고려하는 중요 변수이다.

- 방문 횟수 (VC: Visiting Count)
- 방문 주기 (VF : Visiting Frequency) : VC / (today - starting_date)

- 최고 방문 주기(MVF : Maximum VF per Week) 1주일 단위
- 최근 1주일간의 방문 주기 : (LVF : Last Visiting Frequency)
- 방문해서 머무른 시간 : (VT : Visiting Time):(총 방문 시간 / 총 방문 횟수)
- 정보의 양 (파일 크기) : (CL : Content Length)
- 페이지와 관심분야와의 유사도 (CO : Corelation between Page and Concern)

여기에서 방문 주기를 계산하기 위해 필요한 start_date는 고정 변수로 웹 브라우저를 처음 사용하여 퍼지 에이전트가 최초로 작동을 시작한 날을 의미한다. 최고 방문 주기와 최근 1주일간의 방문 주기도 또한 고정 변수로 고정되어 있다. 만약 이 값을 초과하는 새로운 정보가 생기는 경우, 에이전트는 이를 새로운 고정 변수로 정한다.

그리고 페이지와 관심 분야의 유사도를 위해서 다음과 같은 처리를 한다. 사용자는 초기에 자신의 관심 분야를 웹 에이전트에게 알리게 되는데 관심 분야의 유사도는 사용자가 입력한 관심 분야에 대한 주제어가 문서에 포함되어 있으면, 포함되어 있는 위치에 의해 중요도를 정하여 놓는다. 이 방법은 실제로 언어적인 동의어나 기타의 시소러스를 이용하지 않고 빠르게 처리한다는 장점을 가진다고 할 수 있지만 정보의 유사도를 정확히 표현하기에는 모자라는 면이 있다고 할 수 있으나 여기에서는 한사람의 사용자를 위한 에이전트의 구현에 목적이 있으므로 이를 이런 형태로 표현하였다.

사용자가 웹 브라우저를 통해서 새로운 페이지를 가져오게 되면, 이 새로운 URL은 위한 같은 변수를 가지며, 이를 바탕으로 사용자가 방문한 페이지의 중요도를 새롭게 계산하게 된다.

4.2 퍼지화

위의 각 변수들은 이제 퍼지 언어 변수로 변환 되는데 여기에서는 아래와 같은 5가지의 언어 변수를 갖는다.

VS : Very Small , S : Small, M : Mediul,
L : Large, VL : Very Large

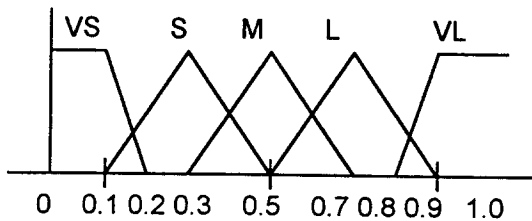


그림 3. 퍼지 멤버 함수

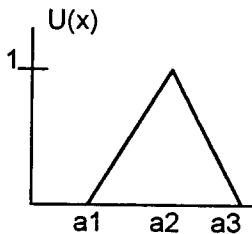


그림 4. 각 언어변수에 대한 정의

이 변수들로 입력값들을 변환 시키기 위해서 다음과 같은 함수를 이용한다.

$$\begin{aligned}
 U(x) &= 0, & x < a_1 \\
 &= (x - a_1) / (a_2 - a_1), & a_1 \leq x \leq a_2 \\
 &= (x - a_2) / (a_3 - a_2), & a_2 \leq x \leq a_3 \\
 &= 0, & x > a_3
 \end{aligned}$$

4.3 퍼지 규칙

퍼지 규칙은 크게 두부분으로 구분된다. 하나는 사용자의 접근에 대한 부분으로 방문 횟수, 최대 방문 주기등에 관련된 부분으로 사용자의 URL에 대한 취향을 나타내는 부분이다. 여기에서는 다음과 같은 퍼지 규칙이 적용된다.

IF VC = A and VF = B and MVF = C and LVF = D
then V = E

최고 625개의 규칙의 정의가 가능하지만 현재 64개의 규칙을 이용하고 있다. 다음은 Rule중의 하나를 표시한다.

IF VC=L and VF=L and MVF=VL and LVF=S
then V1=L

여기에 적용되는 몇 개의 Heuristic을 살펴보면 현재의 시점에서 가장 중요한 정보는 LVF이다. 최근에 얼마나 방문했는가 가장 중요하고, 다음으로 중요한 정보는 VC와 VF가 된다. 최근에 많이 방문하지 않았더라도 꾸준히 많이 방문한 곳이면 중요한 정보로 취급되어야 한다. 마지막으로 MVF는 어느 시기에 굉장히 중요한 정보였으므로 이것도 중요 정도를 어느 정도 인정해 주어야 한다.

본 논문에서는 이 규칙을 적용하여 추론하는 방법으로 Mamdani의 Max-Min연산을 하였다.

다음은 정보의 양과 질에 대한 부분이다. 정보의 양은 파일의 크기에 의해서 판단하며, 정보의 질은 사용자의 관심과 페이지간의 유사도에서 판단한다. 그리고 양적인 질적인 면에서의 취향은 방문해서 머무른 시간으로 간주한다. 여기에서 사용되는 규칙은 다음과 같은 형식이 된다.

IF VT=A and CL=B and CO=C then V2=D

여기에는 총 50개의 규칙이 적용된다.

그리고 실제적으로 적용되는 규칙을 보면

IF VT=L and CL=L and CO=VL then V2=VL이 된다.

여기에 적용되는 기준은 VT가 아주 크면 실제로 방문해서 정보를 읽어 다기 보다는 다른 일로 인해서 사용을 안했다고 보는것이 타당하다. 곧 VT는 어느 정도의 시간이상을 머무러야 하며, 내용도 또한 너무 짧으면 결국 정보의 양이 많지 않기 때문에 중요하다고 볼 수 없다. 마지막으로 관심

분야와의 유사도는 관심있는 페이지인 경우 주제가 반드시 나타날 것이다. 또한 나타나는 위치도 HTML 안에서 중요도가 다를 수 있다. 이를 이용하여 중요한 부분에 나타난 주제는 중요한 정보일 가능성이 높다고 판단한다.

이렇게 나온 두가지의 퍼지값 V1과 V2는 다시 한번의 퍼지 규칙을 적용하여 최종의 결과를 계산하는데 사용된다. 여기에는 25개의 규칙이 있고, 이때 규칙은 다음과 같다.

IF V1=VL and V2 = L then V=VL

여기에서는 V1이 사용자의 취향에 대한 데이터이고 V2는 정보에 대한 데이터로 판단하고 V1을 위주로 고려한다. 곧 V1이 중요 고려 대상이며, V2는 여기에 부가 정보로서의 역할을 한다.

4.4 비 퍼지화

여기에서 나온 두 개의 퍼지 값은 다시 한번 퍼지 연산을 통해 하나의 퍼지 값으로 변환시키도 마지막으로 이를 Defuzzify 연산을 하여서 최종 값을 얻는다. 본 논문에서는 Center of Area 방법을 사용하여 비퍼지화를 한다. 마지막으로 비퍼지화 연산을 거친 데이터는 정렬 과정을 거쳐서 중요한 정보의 순서를 매기게 된다.

$$U_0 = \frac{\sum_{j=1}^n U(u_j) * U_j}{U(U_j)}$$

여기에서 나온 결과는 웹 브라우저의 Bookmark에 저장되게 되는데, 이때 너무 많은 URL이 한꺼번에 Bookmark에 등록되지 않게 하기 위해서 적당한 크기로 순서를 매긴 값에 따라 등록여부를 결정하게 된다.

4.5 Wanderer

웹 공간을 자동으로 navigation하는 것으로 사용자의 사용이 10분이상 안들어오는 경우, Wanderer가 동작한다. Wanderer는 Bookmark에 있는 중요한 정보를 순서적으로 가져온다. 이때 정보가 변경되었을 경우, 이 페이지를 새로 가져온다. 이때 새로 가져온 페이지에 대해서는 사용자와의 관심 정도를 계산하게 된다. 현재 관심 분야와의 유사도는 단순히 사용자의 관심에 대한 주제가 HTML 페이지 안 어디에 존재하는가에 의해서 결정된다. 타이틀이나 헤더에 정의되어 있으면 중요한 데이터라고 판단하며, 강조된 부분에 있으면 어느 정도의 중요성을 지닌다. 하지만 문서의 중간에 있게 되면 중간 정도의 중요도를 지니며, 한번도 나타나지 않으면 중요하지 않다고 판단한다. 태그에 따른 중요도 판단은 다음과 같다.

- <TITLE> , <H1> : VL
- <H2~6> : L
- 등의 강조 부분 : M
- 텍스트 중간 : S
- 없으면 : VS

여기에서 결정된 변수는 바로 퍼지 지식 베이스의 페이지와 관련 분야와의 유사도에 적용된다.

5. 결론

본 논문에서는 퍼지 이론을 이용하여 현재까지는 사용자에게 의해서 이루어졌던 정보 관리를 자동으로 하는 시스템을 개발하였다. 본 시스템의 특징은 다음과 같다.

- 사용자와 에이전트간의 대화를 통하여 사용자의 관심을 에이전트에게 알게한다.
- 이를 바탕으로 사용자가 가져온

페이지에 대하여 취향 정도를 높인다.

- 사용자가 방문한 횟수, 주기등을 바탕으로 사용자가 많이 방문하는가, 아니면 자주 방문하는가들에 대한 정보를 모으며, 이를 바탕으로 사용자의 취향 정도를 계산하여 실제로 사용자의 손쉬운 접근이 가능한 Bookmark로의 삽입 여부를 결정한다.
- 사용자에게 중요한 페이지에 대해서는 변경 여부를 계속 확인하여 변경되어 있는 경우 이를 자동으로 지식베이스에 저장한다.

본 시스템은 사용자가 직접 관리하던 정보 관리를 웹 에이전트로 대체 할 수 있는 시스템을 개발하였다. 이 시스템에서는 정보의 양과 질, 사용자의 취향에 따라 정보의 중요도를 계산하고 이를 바탕으로 사용자에게 중요한가 아닌가를 판단한다. 사용자들이 일일이 이를 관리하면서 가지게 되는 규칙들을 이용함으로써 보다 효율적으로 그리고 자동으로 이 일을 할 수 있게 하였다.

이 일을 사용자들이 직접하게 된다면, 항상 똑같은 관점에서 관리하지 못하고 시간과 상황에 의해 서로 다른 판단을 하게 될 수도 있으며, 정해진 규정된 방법을 사용하지 않고 퍼지 이론을 적용함으로써 사용자들이 가지는 정보 분류에 대한 애매 모호한 성질을 구체화할 수 있었다.

본 시스템은 아직까지 규칙의 조율과 성능향상, 그리고 사용자들의 관심과의 URL과의 유사도를 측정하기 위한 새로운 알고리즘의 개발이 필요하다.

참고문헌

[1]Michael L. Mauldin, John R. R. Leavitt, Web Agent Related Research at the Center for Machine

Translation, <http://fuzine.mt.cs.cmu.edu/mlm/signid r94.html>, Aug 1994.

[2] Brian Pinkerton, Finding What People Want: Experiences with the WebCrawler, Proceedings of the First International World-Wide Web Conference, 1994.

[3] DeBra, P. and Post, R., Information Retrieval in the World-Wide Web: Making Client-based searching feasible, Proceedings of the First International World-Wide Web Conference, 1994.

[4]Michael Wooldridge and Nicholas R. Jennings, "Intelligent Agents : Theory and Practice," Knowledge Engineering Reivew, Oct, 1994.

[5] C.Mic Bowman, Peter B. Danzig, Darren R. Hardy, U. Manber and Michael F. Schwartz, The Harvest Information Discovery and Access System, Proceedings of the Second International World Wide Web Conference, Oct 1994.

[6] Bard de Boer, Artificial Autonomous Agents and Learning Systems, Lecture Note from free univ. of Brussels, Nov 1994.

[7] Colin G. Harrison, David M. Chess, Aaron Kershenbaum, Mobile Agents : Are they a good idea ?, IBM Watson Research Center, Mar 1995.

[8] T. Berners Lee, R. Calliau, L. Lutonen, H. Frystyk Nielsen and A. Secret, "The World Wide Web," CACM, Aug 1994.

[9] Michael H. Coen, Sodabot : A Software Agent Environment and Construction System, MIT AI TR 1493, Jun 1994.

[10] Iain D. Craig, A Perspective on Multi-Agent Systems, Lecture notes in First Bulgarian Summerschool on Cognitive Science, Feb 1995.

[11] Fim Finin, Jay Weber, Jaames McGuire, Suart Shapiro, Chris Beck, Gio Wiederhold, Michael Genesereth, Tichard Frizsom, Donald McKay, Draft Specification of the KQML Agent Communication Language, Jun 1993.

[12] Taeha Park and Kilnam Chon, " Collaborative indexing over networked information resources by distributed agents," Distib. Syst. Engng 1, 1994.

[13] Christin Guifoyle and Ellie Warner, Intelligent Agents : the New Revolution in Software,

- OVUM, 1994.
- [14] White, J.E ,Mobile Agents in Software Agents, Bradshaw, J.(ed), The MIT Press , 1996.
- [15] White, J.E, Telescript Technology : An Introduction to the Language, General Matic Inc. Oct, 1995.
- [16] OSF, Hypermedia Browsing Technology for Wide Area Information Environments, OSF RI Proposal submitted to BAA 93-42, 1994.
- [17] Michael R. Genesereth,Dave Gunning Rick Hull Larry Roger KING Bob Neches Gio Wiederhold, Reference Architecture : Intelligent Integration of Information Program," <http://logic.stanford.edu/architecture/reference.html>.
- [18] Michael R. Genesereth and Richard E. Fikes,, Knowledge Interchange Format Version 3.0 Reference Manual, <http://logic.stanford.edu/kif/Hypertext/kif-manual.html>. 1994.
- [19] Jose C. Brustoloni, Autonomous Agents : Characterization and Requirements, CMU-CS-91-204, Nov 1991.
- [20] ACM,"Special Issues on Intelligent Agents," CACM, Jul 1994.
- [21] Ellen Voorhees, Agent Collaboration as a Resource Discovery Technique, CIKM-94, 1994.
- [22] Nina M. Berry and Soundar R.T. Kumara, Intelligent Software Agents for Non-Traditional Information Sharing Environment, CIKM-94, 1994.
- [23] Leslie L. Daigle & Peter Deutsch, Agents for Internet Information Clients, CIKM-95, 1995.
- [24] Donald F. Geddis, Michael R. Genesereth, Arthur M. Keller & Narinder P. Singh , Infomaster: A Virtual Information System , CIKM-95, 1995.
- [25] Ram Kumar , Internet Information Resource Discovery Tools: Current and Future Trends , CIKM-95, 1995.
- [26] Donald McKay, Jon Pastor, Rbon McEntire and Tim Finin, An architecture for information agents, Proceedings of the Third International Conference on Artificial Intelligence Planning Systems (ARPI Supplement), (AIPS-96), AAAI Press, May 1996.