

동물 영역 지식 기반의 지능형 정보 에이전트*

A Knowledge-Based Intelligent Information Agent for Animal Domain

이 용 현** 오 정 옥*** 변 영 태****
(Yong-Hyun Rhi) (Junk-Ok Oh) (Young-Tae Byun)

요약 네트워크의 기술 발달로 웹상의 정보 제공자가 증가함에 따라 정보 사용자가 필요한 정보를 신속하고 정확하게 획득하기는 것이 더욱 어려워졌다. 이를 위해서 키워드 정합 방식의 검색 엔진이 많이 개발 보급되고 있으나 여전히 많은 부담이 사용자에게 주어지고 있는 상황이다. 이러한 문제를 해결하기 위해서 본 논문에서는 특정 영역인 동물 분야에 대한 지식베이스를 기반으로 사용자의 의도에 보다 적합하고 해당 영역에 적절한 형태로 사용자 질의를 가공하고, 대용량의 다양한 정보로부터 사용자가 필요로 하는 정보를 제공하는 일을 해주는 지능적인 정보 검색 대리자, 정보 에이전트(HIIA-1a : HongIk Information Agent)를 제안한다. HIIA-1a는 온톨로지 형태에 접근한 동물 관련 지식베이스를 가지고 있으며, 이를 기반으로 사용자 또는 다른 에이전트 시스템의 정보 요청에 대해 필요한 정보를 제공할 뿐만 아니라, 관련 웹 문서 정보도 제공한다. 효율적인 웹 문서의 제공을 위하여 방대한 양의 웹 문서를 대상으로 동물 영역에 관련된 문서를 저장·색인하는 웹DB를 가지고 있다. 또한 사용자의 의도를 좀더 명확하게 표현할 수 있도록 유연한 연산자로서의 질의 확장을 하였으며, 축적된 처리 결과와 사용자의 피드백 정보를 통해 학습을 하게 된다. 본 논문에서는 이와 같은 요소들을 포함하는 HIIA-1a를 구현하고, 실험을 통해 시스템의 효율성을 보인다.

Abstract Information providers on WWW have been rapidly increasing, and they provide a vast amount of information in various fields. Because of this reason, it becomes hard for users to get the information they want. Although there are several search engines that help users with the keyword matching methods, it is not easy to find suitable keywords. In order to solve these problems with a specific domain, we propose an intelligent information agent(HIIA-1a : HongIk Information Agent) that converts user's queries to forms including related domain words in order to represent user's intention as much as it can and provides the necessary information of the domain to users. HIIA-1a has an ontological knowledge base of animal domain, supplies necessary information for queries from users and other agents, and provides relevant web page information. One of system components is a WebDB which indexes web pages relevant to the animal domain. The system also supplies new operators by which users can represent their thought more clearly, and has a learning mechanism using accumulated results and user feedback to behave more intelligently. We implement the system and show the effectiveness of the information agent by presenting experiment results in this paper.

1. 서론

컴퓨터의 하드웨어와 소프트웨어, 그리고 통신 및 네트워크의 급속한 기술 발달로 지리적 제한의 한계를 넘어 인터넷을 이용하는 사용자가 급증하게 되었다. 이와 같이 웹(World Wide Web) 상의 정보 제공

* 본 연구는 과학 재단 (97-0100-0801-3 제 2세부과제)의 지원을 받았다.

** 홍익대학교 전자공학과 박사 과정

*** 로커스주식회사 연구원

**** 홍익대학교 전자계산학과 부교수

자가 폭발적으로 증가함에 따라 필요한 정보를 웹에서 찾아내야 하는 정보 사용자는 망망대해의 정보 바다 앞에서 있는 상황이 되었고 필요한 정보를 신속하고 정확하게 획득하는 것이 더욱 어려워졌다. 이를 위해서 키워드 정합방식의 검색 엔진이 많이 개발 보급되고 있으나 여전히 많은 부담이 사용자에게 주어지고 있는 상황이다. 따라서 특정한 분야에 대한 지식베이스를 기반으로 질의를 사용자의 의도에 적합하면서 특정 분야에 적절한 형태로 가공시켜 대용량의 다양한 정보로부터 사용자가 필요로 하는 정보를 제공하는 일을 해주는 지능적인 정보 검색 대리자, 정보 에이전트(Information Agent)가 필요하게 되었고 사용자의 요구에 보다 적합한 정보를 제공하기 위해서 여러 개의 에이전트가 서로 협력하여 작업을 처리하는 다중 에이전트 개념을 도입하게 되었다(1-3). 즉, 서로 다른 분야에 대해서 전문적인 지식을 갖는 여러 정보에이전트들이 자신들의 지식베이스를 기반으로 서로 협력해서 문제를 해결하고 사용자에게 필요한 정보를 제공할 수 있게 된다.

본 연구는 이러한 필요에 의해서 동물 관련 지식베이스를 기반으로 사용자 또는 다른 에이전트 시스템의 정보 요청을 처리하여 필요한 정보를 제공할 뿐만 아니라 특정 분야에 적절한 형태의 사용자 질의 확장을 통해서 효율적인 웹 문서를 제공하는 정보 에이전트(HongIk Information Agent : 이후 HIIA-1a라 명명)를 제안한다.

HIIA-1a는 동물에 대해서 전문적인 지식을 보유하는 정보에이전트이다. 하지만 전문가 시스템이나 테이터베이스 시스템과는 달리 다른 정보에이전트와의 통신을 통해서 서로의 정보를 공유하게 된다. 또한, 서로간의 정보를 이해하기 위해서는 특정 영역에 대해서 개념화된 온톨로지가 필요하기 때문에 HIIA-1a의 영역 지식베이스는 온톨로지 형태에 접근한 지식베이스로 구축되었다.

온톨로지(ontology)란 개념화의 명세(specification of conceptualization)이다(4). 에이전트가 특정 영역에 관한 질의를 하거나 통신하고자 할 때, 먼저 그 영역에 대한 지식들의 개념화가 요구되어지는데 이를 온톨로지라고 한다. 즉, 다중 에이전트 시스템에서 에이전트끼리 서로 대화하기를 원할 때 서로 사용하는 기호나 표현을 동일한 의미로 이해하기 위해서는 특정 영역에 대한 공통적 어휘 기반을 구축해야한다. 따라서 정보 에이전트가 다른 에이전트와 통신을 할 때 서로의 메시지를 완전하게 이해하기 위해서 에이

전트들은 온톨로지를 공유할 필요가 있고 이러한 이유에서 에이전트는 온톨로지의 형태로 지식베이스를 구축할 필요가 있다(5-9).

질의의 확장에 있어서는 일반 검색엔진이나 문헌 정보 검색 엔진에서는 시소러스나 지식베이스를 기반으로 단순히 입의 용어로 입력된 질의를 동의어, 상·하위 개념 혹은 관련어(related term)로 확장함으로써 문서 검색에 사용할 수 있는 색인어(index term)의 형태로 변환한다(10-14). HIIA-1a에서는 동물에 관련된 지식베이스를 기반으로 사용자 질의를 처리해서 사용자가 알고 싶어하는 정보를 제공할 뿐만 아니라 좀 더 다양한 정보의 제공을 위해서 질의를 속성, 동의어, 상·하위 개념으로 확장시킴으로써 사용자 의도를 보다 구체화시켜 웹 문서를 제공하게 된다. 또한 사용자의 질의에 대해 사용자의 의도를 명확하게 표현하는데는 단순 불리언 연산자만으로 우리가 따르기 때문에 사용자의 의도를 보다 명확하게 표현할 수 있도록 한정사, 형용사 등을 연산자로 확장하였다.

HIIA-1a는 에이전트이고, 에이전트란 인간처럼 자치적이고 지능적으로 동작하는 시스템이다(19). 항상 똑같은 행동을 반복하기보다는 축적된 경험을 갖고 행동이나 지식상의 오류를 수정하고 시스템의 성능 향상을 위해 학습을 하게된다. 본 시스템에서는 두 가지 형태의 학습이 이루어진다. 하나는 주어진 질의가 처리된 정보를 축적하고 이를 바탕으로 영역 지식의 확장을 도모하고, 다른 하나는 주어진 결과에 따른 사용자의 반응을 통해 학습에 이용하게 된다.

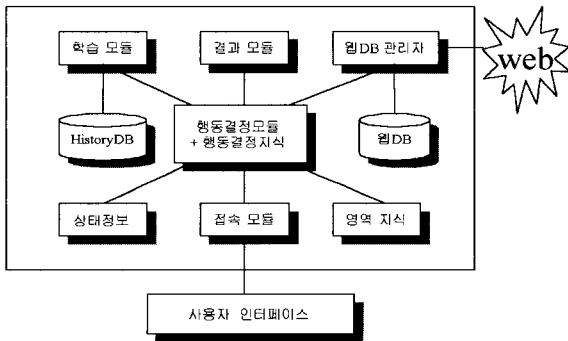
사용자에게 다양한 정보를 제공하기 위하여 질의에 관련된 웹 문서의 정보도 제공하고 있다. 웹 문서의 제공에 있어서 현재 웹이 다양하고 방대한 양의 문서들을 대상으로 하고 있기 때문에 본 시스템에서는 동물 영역에 적합한 문서만을 대상으로 저장·관리할 수 있는 데이터 베이스(이하 웹DB라고 칭함)를 구축하여 사용하고 있다. 이런 데이터 베이스의 구축은 크게 색인과 검색으로 나누어진다. 색인은 정합과정에서 사용할 수 있도록 문서에서 필요한 정보를 추출하는 과정이다. 색인 과정에서의 기존 연구로는 웹 문서를 표현하는 방법으로 salton의 원리(15), 문서를 적절한 형태로 변환하여 정합 과정에서 효과적으로 사용할 수 있도록 하는 방법으로 역 색인구조(16)나 요약파일(17) 등이 연구되고 있고, 문서에서 유용한 단어나 문장들을 추려내기 위해 정보 검색 분야에서 많은 연구가 진행되고 있다. 또한 추출된 색인어들에 대한 문서상의 중요도에 관한 대표적인 연구로

는 각 색인어들에 대한 여러 가지 가중치 요소에 대해 연구되고 있다(15). 검색은 질의에 대한 정보를 제공하는 과정이다. 일반적으로 웹을 기반으로 하는 검색 엔진들은 단어의 유·무에 따른 문서의 검색을 처리하고 있다. 또한 관련 문서들의 순위를 정하기 위해서는 질의와 문서간의 유사도 값을 계산하여 순위를 결정하는 알고리즘들이 사용되는데, 그 중 대표적인 검색 모델로는 Vector space model(16)과 Probabilistic model, Fuzzy set model(17), Extended boolean model(18) 등에 대한 연구가 이루어지고 있다.

본 논문은 2장에서 HIIA-1a의 전반적인 구조와 각 구성 모듈의 목적·동작에 대해서 언급한다. 3장에서는 동물 영역 지식베이스의 구조와 사용자의 의도를 명확하게 하기 위해서 확장된 연산자들에 대해서 확장 의도와 사용 방법에 대해서 보이고, 4장에서는 HIIA-1a에서 이루어지고 있는 학습에 대해 언급한다. 그리고 5장에서는 웹 관련 정보 제공처로의 웹 DB에 대해서 알아보고, 마지막으로 결론 및 향후과제로 끝을 맺는다.

2. HIIA-1a의 구조

(그림 1)은 HIIA-1a의 시스템 구조이다. HIIA-1a는 기본적으로 사용자 인터페이스를 통해 주어진 질의에 대한 정보를 제공하게 된다. 또한, HIIA-1a는 상태정보, 행동결정모듈과 행동결정지식, 영역 지식, 웹DB관리자, 결과 모듈, 학습 모듈, 접속 모듈로 구성되고 웹DB관리자의 웹DB와 상태정보의 History DB와 같이 두 개의 DB를 가지고 있다. 이들은 서로 유기적인 관계를 유지하여 통합적으로 작동되어진다. 각 모듈의 용도와 동작 방법에 대해서 좀 더 구체적으로 다룬다.



(그림 1) HIIA-1a의 시스템 구조

2.1 사용자 인터페이스

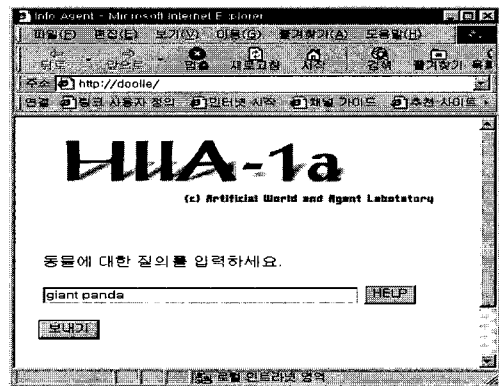
사용자 인터페이스는 (그림 2)와 같이 웹 브라우저를 통해서 제공된다. 사용자가 웹 브라우저를 통해서 질의를 입력하면 그에 대한 정보와 웹 문서의 URL의 리스트가 제공된다. 또한 URL 리스트의 제공에서는 각각의 문서에 대한 사용자의 만족 정도를 표현하기 위해서 사용자는 “만족”, “그럭저럭”, “불만족” 중 하나를 표시한다. 이 사용자 피드백 정보는 HIIA-1a에게 전송되어 학습을 위한 정보로 활용되어진다. 사용자 인터페이스는 HIIA-1a와 독립되어 HIIA-1a의 접속 모듈과 TCP/IP를 기반으로 메시지를 주고받는다.

2.2 상태정보

상태정보는 현재 에이전트의 상태를 나타내는 정보와 시스템이 처리한 결과들의 정보를 저장하는 HistoryDB로 되어 있다.

(1) 현재상태정보

HIIA-1a의 현재 상태를 나타내는 정보는 질의 큐와 에이전트 상태정보가 있다. 입력된 질의는 바로 질의 큐에 넣어지며 질의 큐가 비어 있지 않으면 질의 큐가 빌 때까지 질의 큐에서 질의를 꺼내서 처리한다. 에이전트 상태정보는 에이전트가 현재 무슨 작업 중인지를 나타내며 현재로서는 3개의 상태(IDLE, ANSWERING, LEARNING) 중 하나의 상태값을 가진다. 질의를 처리중이면 ANSWERING이 되고 아무 동작도 하지 않고 있으면 IDLE 상태가 된다. 이러한 정보는 행동결정모듈이 어떠한 행동결정지식을 선택할지의 조건이 된다.



(그림 2) 사용자 인터페이스

(2) HistoryDB

HIIA-1a에 의해서 처리된 모든 결과가 저장되는 곳으로 HIIA-1a의 축적된 경험이 된다. 시스템이 처리한 모든 결과는 HistoryDB의 두 개의 테이블에 저장되어 있다. 하나는 영역지식에 의해서 처리된 결과를 저장하는 Result 테이블로 이 테이블에는 일련번호(id)와 사용자의 원래의 질의(Query), 확장된 질의(Param)를 표시하는 3개의 필드로 구성되어 있다. 다른 하나는 사용자에게 제공된 URL 리스트와 사용자 피드백 정보를 저장한 WebIR 테이블로 이 테이블에는 일련번호(Qid), 사용자의 원래의 질의(Query), 질의를 처리한 시간(Time), 검색된 문서들의 문서번호(PidList)들, 사용자로부터 받은 피드백 정보(FeedbackList)를 저장하고 학습여부(Learning Status)를 표시하는 총 6개의 필드로 구성되어 있다.

2.3 행동결정모듈과 행동 결정 지식

행동결정지식은 사람의 지식과 같은 역할을, 행동결정모듈은 HIIA-1a의 두뇌와 같은 역할로서 환경의 변화에 반응하여 행동결정지식을 기반으로 적합한 행동을 추론하게 된다. 행동결정모듈은 행동결정지식과 상태 정보를 사용하여 시스템이 수행해야 하는 행위를 추론하는 모듈이다. 즉 HIIA-1a의 현재 상태에 부합되는 조건부분을 갖고 있는 행동결정지식이 있는지를 알아보고 해당되는 행동결정지식의 행동부분을 수행하게 된다. 예를 들면 상태정보에 의해서 질의 큐가 비어있지 않다는 정보를 얻으면 행동결정모듈은 조건부분이 만족하는 행동결정지식의 행동부분을 수행하여 질의 큐가 빌 때까지 질의를 처리한다. 행동결정모듈은 CLIPS(25)의 추론 엔진을 사용하고 있다. 행동결정지식은 HIIA-1a에서 가장 중요한 부분으로서 시스템이 상황에 따라서 어떻게 행동해야 하는지를 조건-행동규칙(condition-action rule)의 형태로 표현한 정보이다. 마치 사람이 상황에 따른 행동규칙을 갖고 있듯이 정보에이전트는 발생한 상황에 대해서 행동결정지식을 기반으로 상황에 대처할 행동을 결정하게 된다. 행동결정지식은 시스템의 일반적인 동작원리에서부터 특별한 상황에 대한 대처방안 등이 있다. 이런 내용이 다양하고 풍부할수록 에이전트가 자율적이고 지능적으로 동작할 수 있다.

2.4 영역 지식

HIIA-1a는 동물에 관련된 영역 지식을 가지고 있으며, 이러한 영역 지식은 다음과 같은 부분에서 사

용되어 진다. 영역 지식에 대해서는 3장에서 구체적으로 설명한다.

- 웹DB에서 웹 문서의 내용이 동물 영역에 부합되는지를 판단하기 위한 지식 정보
- 사용자 혹은 다른 에이전트의 정보 요청에 응답하기 위한 지식 정보
- 사용자의 질의에 대한 개념을 구체화시키고 확장시키기 위한 지식 정보

2.5 웹DB관리기와 웹DB

웹DB관리기(29)는 주어진 질의로부터 추론을 통해 확장된 질의를 행동결정모듈에서 넘겨주면 이미 구축된 웹DB(30)에서 해당 문서를 검색하여 순위 결정 알고리즘을 거친 문서 리스트를 넘겨준다. 웹DB관리기는 행동결정모듈에 의해서 질의처리과정에 호출되고 웹DB로 3개의 질의를 파라미터로 넘겨받는다. 첫 번째 질의는 주어진 질의를 시스템 처리를 위해 변형한 확장 질의이고, 두 번째 질의는 시스템에서 질의를 처리한 결과이고, 세 번째 질의는 첫 번째 질의와 두 번째 질의에 존재하는 객체의 계층구조상의 상·하위 개념이다. 웹 DB에서는 첫 번째 질의를 통해서 웹 문서를 검색하고 두 번째와 세 번째 질의를 갖고 검색된 웹 문서의 유사도(similarity)를 계산하게 된다. 웹DB는 웹 문서에서 특정 정보만을 추출하여 구축한 데이터 베이스로 정보의 검색과 정보의 제공에 사용된다. 웹DB는 검색을 효율적으로 수행하기 위해서는 특정 용어를 가지고 있는 모든 문서를 빠르게 찾아내는 기능이 필요한데, 이를 위해서 역 색인(inverted index) 구조를 사용한다. 웹DB에서는 총 7개의 table을 가지고 있고, 각 table은 고유의 기능을 가지고 웹 정보의 색인과 검색에 참여한다. 웹DB의 세부적인 내용은 5장에서 설명한다.

2.6 학습모듈

본 시스템에서는 두 가지 학습 유형을 처리한다. 하나는 영역 지식에 대한 학습으로 HistoryDB의 축적된 결과를 대상으로 영역 지식의 계층 구조상에서 기존에 존재하는 객체들의 동의어 확장과 새로운 속성의 추가 혹은 속성값을 확장하게 된다. 즉, 동일한 질의에 대해서 다수의 사용자가 공통적으로 만족한 페이지를 추출하고 추출된 페이지에서 확률적으로 공통된 단어들을 추려내어서 현 영역 지식의 정보에 포함되어 있지 않으면 그 단어를 영역 지식 전문가에게 제공함으로써 해서 영역 지식의 확장을 피하게 된다.

또 다른 하나는 퍼지 멤버쉽 함수(Fuzzy Membership Function)의 조정에 대한 학습으로 본 시스템에서는 형용사를 사용한 질의를 허용하고 이러한 처리를 위해서 퍼지멤버쉽 함수를 사용한다. 사용자의 피드백 정보를 기반으로 사용자들의 성향을 알아보고 퍼지멤버쉽 함수를 조정해서 사용자의 성향을 학습하게 된다.

2.7 접속모듈

HIIA-1a의 모든 입력과 출력이 이루어지는 곳으로 사용자의 질의 입력과 결과 전송, 다른 에이전트들과의 정보 교환, 사용자 피드백 정보의 입력 등의 사용자 인터페이스와 다른 정보 에이전트와의 교신이 전부 접속 모듈을 통해서 처리된다. 접속 모듈은 에이전트의 특성 중에 하나인 사고성을 갖기 위해서 필요한 요소이다. 동시에 다중 접속이 가능하며 WWW의 기반 프로토콜인 TCP/IP를 기반으로 하고 있으며 소켓을 이용하여 구현되었다.

2.8 결과모듈

사용자에게 제공될 영역지식을 통해서 처리된 결과들과 웹DB관리에 의해서 제공된 문서의 리스트를 결합시켜서 HTML문서를 구성하게 된다. 이 문서를 웹 브라우저를 통해 사용자에게 제공한다.

3. 지식의 표현과 질의 확장

3.1 지식베이스 구조

본 시스템의 지식베이스는 DIC과 BKB로 구성되어 있다. DIC은 동의어와 시스템어에 관한 정보로 구축되었고 BKB는 동물에 대한 전문적인 지식으로서 동물의 계층 구조와 각 동물에 대한 속성과 속성값 등을 갖고 있다.

3.1.1 DIC

DIC은 일반 용어들 사이의 동의어 관계와 일반 용어에 관련된 시스템어 정보를 갖고 있다. 시스템어는 한 객체(object)를 의미하는 여러 용어들 중에서 학명을 객체로 표현하기 위해서 사용하는 용어이다. 즉, 한 동물을 의미하는 용어는 여러 개가 존재하기 때문에 여러 용어들 중에서 어떠한 단어가 입력되어도 동일한 객체로 처리하기 위해서 시스템어라는 것을 정의한다. 이러한 시스템어의 사용은 여러 가지 용어를 하나로 표준화하는 의미와 처리상의 편의를 위해서이다. DIC은 이러한 지식을 이용해서 질의의 일반

용어를 시스템어로 바꿔서 BKB에게 건네주고, BKB의 처리 결과에서 시스템어를 일반 용어로 바꿔주는 BKB의 front-end 역할을 한다. BKB의 지식은 이런 시스템어에 의해 구축되었다. 또한 한글에 대한 처리도 가능하다. 질의큐에서 가져온 질의가 영역 지식으로 보내지기 전에 전처리 과정을 거친다. 이 과정에서 한글로 입력된 질의에 대해서는 한영사전을 이용하여 영어로 바꾸어지게 된다. 예를 들어 질의가 '북극곰'으로 들어오면 'polar bear'로 바꾸어 추론을 진행하게 되는 것이다.

3.1.2 BKB

분류 생물학에서는 동물들을 특성별로 분류해 놓은 계(문>강>목>과>속)종의 7단계의 계층 구조가 있다. 각 계층구조는 좀더 세부적으로 나뉘거나 묶여서 확장하기도 한다[20-24]. 본 시스템의 동물에 대한 지식은 분류 동물학의 계층 구조를 기반으로 구축되어 총 1,114개의 계층정보를 갖고 있으며 포유류 데이터가 110개, 양서류 데이터가 15개, 파충류 데이터가 30개, 어류 데이터가 25개, 조류 데이터가 50개, 그래서 총 230개의 데이터가 존재한다. BKB의 지식은 COOL (CLIPS Object Oriented Language)로 표현되었다[25]. BKB는 두 가지 종류의 지식을 갖고 있다. 하나는 동물에 관한 전문적인 지식으로 계층 구조와 동물들에 해당하는 속성(값)들이고, 다른 하나는 동물에 대한 직접적인 정보는 아니지만 동물과 관련 있는 지식으로 단위와 지리 등에 대한 지식이다. 이러한 지식들을 사용하여 단위의 변환과 속성값 제한 등의 작업을 수행해 줄 수 있다. 각 지식 정보는 다음과 같다.

(1) 객체 정보

BKB의 지식은 객체 지향 개념을 사용하여 분류 동물학의 계층 구조를 기반으로 구축되었다. 객체 지향 개념은 객체의 추상화(encapsulation), 속성 상속(inheritance) 등의 개념을 제공하기 때문에 본 시스템에서는 BKB의 모든 지식 객체를 클래스의 인스턴스(instance)로 모델화한다. 계층 구조상의 다른 객체들은 자신의 특성에 맞는 속성들을 갖고 상위 객체들의 속성을 상속받게 된다. 즉, 객체는 상위 객체로부터 속성뿐만 아니라 속성값도 상속받을 수 있다. 이러한 속성값의 상속성을 이용하면 객체들을 분류 동물학상의 계층 구조 외의 특성별로 그룹화(grouping)할 수 있다. 숫자값을 갖는 속성들은 각각 기본 단위를 가지고 정보로 제공될 때는 적합한 단위로 변환된다.

(2) 속성 정보

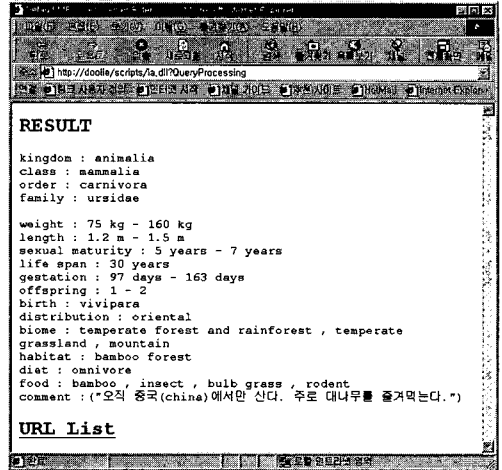
HIIA-1a에서는 객체의 정보 외에 그 객체에 대한 속성(값)에 해당되는 것들에 대해서도 정의하였다. 속성 정보는 객체의 속성들에 대한 정보로서 속성 특유의 특징에 대한 정보를 갖는다. 데이터베이스와 기존의 지식베이스에서는 사용하는 용어에 대해서 목적으로 사용한다. 이런 목적인 정의를 이해하기 위해서는 사람들이나 혹은 다른 시스템은 정의에 대한 배경 지식을 갖고 있어야 한다. 따라서 이러한 의미의 모호성을 배제하기 위해서 온톨로지가 필요하고 이것이 다른 시스템과 통신해야하는 다중 에이전트 시스템에서는 잘 정의된 온톨로지가 필요한 이유이다. 방대하고 전문적인 지식을 필요로 함으로 완전한 온톨로지 형태를 갖춘 지식베이스는 아니지만 나름대로 본체론적인 관점에서 객체들을 해석하려고 노력했으며 이것이 객체의 계층 정보 외에 속성과 속성값에 대한 지식을 구축하게 된 이유이다.

(3) 지리 정보

지리 정보는 각 객체별로 별도의 속성을 갖고 있지는 않지만 각 객체들이 계층구조를 이루고 있다. 따라서 객체 정보에서 사용되는 속성값으로 바다 이름, 동물군계 이름, 동물지리 이름이 각각 세계의 지리를 특징하게 분류해 놓은 것일 뿐만 아니라 세계라는 계층에 속한다는 것을 알 수 있게 된다.

(4) 단위 정보

HIIA-1a에는 단위를 표시하기 위한 일환으로 길이(length), 질량(mass), 시간(time)에 대한 3개의 차원(dimension)이 존재하고, 각 차원은 기본 단위로 m, kg, day를 갖는다. 모든 단위들은 <dimension, base, definition>의 3가지로 표현되어진다. 여기서 dimension은 단위의 차원이고, base는 기본 단위이다. 또한 definition은 기본 단위와의 차이를 표시할 수이다. 예로 km의 표시는 dimension은 길이이고, base는 미터(m)이고, 기본 단위 m를 기준으로 km를 표현하기 위해서는 기본단위에 1000을 곱해야 하기 때문에 definition은 1000을 갖는다. 이 때 base는 반드시 기본 단위일 필요가 없고 기존에 존재하는 단위이면 되고, 기본 단위가 될 때까지 계속 변환을 해주게 된다. 시스템어의 개념을 여기에도 도입하여 각 단위를 나타낼 수 있는 경우의 단어들을 일반어로 놓고 시스템에서 기본 단위로 사용하는 단어들도 시스템어로 놓는다.



(그림 3) 'giant panda' 질의 결과

3.2 확장된 연산자를 이용한 질의의 확장

질의의 처리에 있어서 단순 불리언 연산자만으로 사용자가 자신의 의도를 명확하게 표현하기에는 어려움이 따른다. 예를 들면 육지에서 서식하는 포유류들 중에서 가장 큰 동물을 알고 싶다는 사용자의 의도를 AND 혹은 OR만으로 표현하기 어렵고 'land AND largest AND mammal'와 같이 표현하여도 단순히 단어의 열거에 불과하기 때문에 그 의미는 해석하기 어렵다. 따라서 본 시스템에서는 불리언 연산자 외에 사용자 의도를 보다 명확하게 표현할 수 있도록 연산자를 확장하였다. 이와 같이 연산자를 포함한 질의는 연산자의 특성에 따라 자동적으로 질의를 확장하여 웹DB관리자에게 질의를 넘긴다. 웹DB에게 넘겨지는 질의는 3개의 파라미터로 되어 있다.

(1) 연산자가 없는 질의문

질의 자체가 하나의 객체를 의미하고 사용자는 해당 객체에 관련된 모든 지식의 제공을 원한다고 간주한다. 따라서 해당 객체의 계층 구조상의 상위·하위 개념으로 한 단계씩 확장하고 해당 객체에 대해서 알고 싶은 것들을 나타내기 위해서 속성의 이름만을 질의에 추가한다. 사용자는 주어진 질의에 대한 객체의 계층 정보, 속성 정보 등을 전부 제공받게 된다. [그림 3]에서 'giant panda'라는 한 객체로만 구성된 질의의 결과를 보여주고 있다. 질의는 다음과 같이 확장되어진다.

query1 : bamboo panda OR ailuropoda melanoleuca
OR great panda OR giant panda,

query2 : length, weight, size, habitat, biome,
distribution, geographical range,
geographical distribution, food, diet,
offspring, litter, cub, sexual maturity, life
span, birth, feature, characteristic

query3 : ursidae, bear

(2) OF 연산자 : “〈속성〉 OF 〈객체〉”

〈객체〉의 〈속성〉값을 알고 싶을 때 사용하는 경우로 사용자에게는 속성값이 제공되지만 문서를 검색하기 위해 속성값없이 속성 이름만을 질의에 추가한다. 질의는 〈객체〉에 대한 상·하위 개념과 동의어 개념으로 확장된다. 'baby OF bamboo panda'라는 질의에 대해 다음과 같이 질의가 확장되어진다.

query1 : baby AND bamboo panda

query2 : <null>

query3 : ailuropoda melanoleuca, great panda, giant
panda, ursidae, bear, offspring, litter, baby,
cub

(3) AND 연산자 : “〈객체1〉 AND 〈객체2〉”

모두에 대한 정보를 원하는 질의 구조로 주어진 질의와 두 객체의 공통요소를 추론해서 질의를 확장한다. 공통요소는 속성에만 국한되지 않고 상·하위 개념에서의 공통점도 찾아본다. 사용자에게는 두 객체의 공통 요소를 알려준다. 'bamboo panda AND polar bear'라는 질의에 대해 다음과 같이 질의가 확장되어진다.

query1 : bamboo panda AND polar bear

query2 : ursidae, diet, birth

query3 : ailuropoda melanoleuca, great panda, giant
panda, thalarctos maritimus, bear

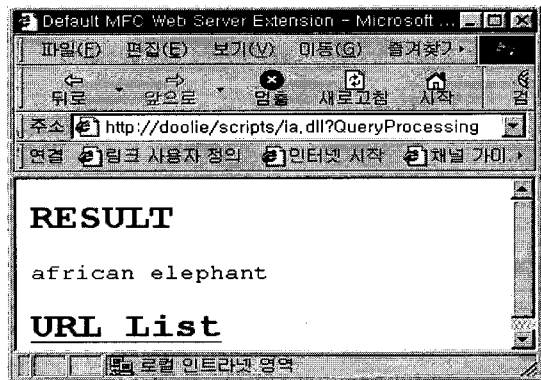
(4) OR 연산자 : “〈객체1〉 OR 〈객체2〉”

둘 중 하나라도 알고 싶은 경우에 사용되는 질의 구조이다. 따라서 질의는 〈객체1〉과 〈객체2〉의 속성 이름, 상·하위 개념과 동의어로 확장된다. 이 때 확장은 중복을 허용하지 않는다. 즉, 사용자에게는 각각의 객체에 대한 정보를 연산자가 없는 질의의 경우처럼 넘긴다. 'bamboo panda OR polar bear'라는 질의에 대해 다음과 같이 질의가 확장되어진다.

query1 : bamboo panda OR polar bear OR ailuropoda
melanoleuca OR great panda OR giant
panda OR thalarctos maritimus

query2 : length, weight, size, habitat, biome,
distribution, food, diet, geographical range,
geographical distribution, offspring, litter,
cub, sexual maturity, life span, birth, feature,
characteristic

query3 : ursidae, bear



(그림 4) "MAX size animal AMONG mammal IN land" 처리 결과

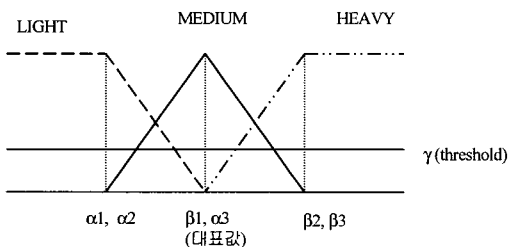
(5) 한정사

단순한 불리언 연산만으로는 표현하기 힘든 개념을 처리하기 위해서 MAX, MIN, LT, LE, GT, GE와 같은 한정사를 지원한다. 이런 한정사들은 추론에 따른 결과와 그 결과의 상·하위 개념으로 확장되어 질의를 확장한다. MAX, MIN값은 지식베이스의 데이터를 기반으로 〈객체집합〉, 〈속성〉, 〈지역〉에 대해서

모든 가능한 경우를 조합해서 얻을 수 있는 값을 모두 미리 구해 놓는다. LT, LE, GT, GE와 같은 한정사의 경우 조건이 되는 <속성값>이 질의에 포함되어 추론이 가능하게 된다. 이 경우 <단위>, <속성> 등을 같이 명시할 수 있도록 하였고, 입력된 값이 없는 경우 각각 시스템의 기본값으로 처리한다. [그림 4]는 한정사를 이용한 질의인 "MAX size animal AMONG mammal IN land"를 처리한 결과이다.

(6) 형용사

한정사를 사용해서 질의를 만들기 위해서는 규칙을 따라야 하기 때문에 익숙하지 못한 사용자에게는 규칙을 익혀야만 하는 불편이 따른다. 따라서 사용자에게 보다 친숙한 형태의 연산자를 제공할 수 있도록 형용사를 연산자로서 사용한다. 가령 무게에 관해서 무거운(heavy), 보통의(medium), 가벼운(light)의 형용사를 사용해서 질의를 할 수 있다. 모든 속성은 형용사 집합을 가지고 있고 이는 각각 3개의 범위로 표현한다. 형용사를 포함한 사용자 질의는 처리 결과로 나온 객체들에 대한 문서를 제공하게 된다. 형용사는 사용하는 사람의 주관이 개입되기 때문에 그 의미가 명확하지 못하다. 따라서 개념의 모호성을 다룰 수 있도록 퍼지멤버십 함수를 사용하였다. [그림 5]는 무게에 대한 퍼지멤버십 함수이다. 각 범위는 α, β 를 가지고 있고, HEAVY 함수에 대한 퍼지함수값이 0이 되는 점을 α_3 라고 하고 퍼지함수값이 1이 되는 점을 β_3 라고 한다. 마찬가지로 LIGHT 함수의 경우 α_1, β_1 을 가지며 MEDIUM 함수의 경우 α_2, β_2 를 가지고 있다. 이 값의 초기 설정은 데이터의 특성상 대표값을 중앙값으로 정하고 [그림 5]에서와 같이 α_1, α_2 는 제1사분위수, α_3, β_1 은 대표값, 그리고 β_2, β_3 은 제3사분위수가 된다. 사용자의 질의에 대해서 각 함수의 값이 임계값($\gamma = 0.1$)을 넘는 객체들이 사용자에게 제공된다. 퍼지멤버십 함수



[그림 5] 퍼지멤버십 함수들(HEAVY, MEDIUM, LIGHT)

는 MIN, MAX와 마찬가지로 계층구조상의 모든 가능한 객체집합에 대해서 자동적으로 생성이 된다.

4. 학습

에이전트 시스템이 기존의 시스템과 다른 특성 중에 하나는 지능이다. 지능은 계획을 세우고 학습을 통해서 새로운 지식을 스스로 터득하는 성질로서 이러한 지능을 바탕으로 효과적으로 동작할 수 있게 된다. 따라서 HIIA-1a가 지능적으로 동작하기 위해서 2가지 학습을 한다. 하나는 영역지식의 확장이고 다른 하나는 사용자의 반응에 대한 학습으로 퍼지멤버십 함수의 조정이다.

4.1 영역 지식의 확장

영역 지식의 확장을 위한 보조적인 기능을 수행하기 위해 본 시스템은 다음과 같은 작업을 한다.

- ① HistoryDB에서 학습하지 않은 질의를 검색한다.
- ② 각각의 질의에 대해서 사용자가 공통적으로 만족하는 페이지들을 추출한다.
- ③ 추출된 페이지에서 정해진 비율(50%)에 대해 공통적으로 나오는 단어들을 추려낸다.
- ④ 검색된 단어들 중에서 문서를 검색할 때 사용한 단어를 제거한다.
- ⑤ 아직 학습하지 않은 질의들에 대해서 ② ~ ④를 반복한다.

위와 같은 과정을 거쳐서 나온 단어들은 동물에 관한 전문가에게 제공해서 필요한 단어들과 영역지식에서의 적절한 위치를 알려줄 것을 요청함으로써 영역 지식을 확장하게 된다. 자율적이고 능동적인 학습은 아니지만 에이전트가 스스로 판단하여 영역 지식을 확장하기에는 오류의 위험성을 배제할 수 없으므로 동물 영역 전문가에게 조언을 통한 형태의 학습이 이루어진다.

영역 지식의 확장을 위해서 <표 1>은 각각의 질의에 대해서 추려낸 단어들이다. 표를 살펴보면 많은 단어는 아니지만 속성값 혹은 동의어 등으로 확장시킬 수 있는 단어들이 나오는 것을 알 수가 있다. 표에서 밑줄 그어진 단어들이 확장의 가능성이 보이는 것이다. 가령 giant panda의 경우 지리정보(china), 색깔정보(black, white), 그리고 상태정보(endanger)를 얻을 수 있다. HIIA-1a의 지식베이스에는 동물지리와 동물군계를 사용해서 분포를 표현하

기 때문에 china에 대한 정보가 없다. 하지만 giant panda에 대한 만족한 문서에 대해서 공통적으로 china가 존재한다면, 보다 향상된 검색 효과와 사용자에게 필요한 정보를 제공하기 위해서 영역 전문가의 조언에 의해 이 정보를 지식베이스에 확장시킨다. 그리고 특징적인 색상정보를 넣기 위해서 새로운 속성을 확장해야 한다.

질 의	추려낸 단어들
<i>giant panda</i>	animal average <i>black</i> body cat <i>china</i> day eat endanger exist fe forest live mountain nation natural people plant protect remain specy threaten unlike <i>white</i> wild world zoo
<i>african elephant</i>	africa african animal bark bath bull <i>calf</i> call conservation cow day drink eat enemy family femal female fish forest fruit ft grass herd <i>ivory</i> join kg kill land largest live mal male mammal mil sahara savanna service shoulder <i>social</i> south spend threaten trade trunk <i>tusk</i> water weigh wildlife

(표 1) 영역지식의 확장을 위해서 추려낸 단어들

4.2 퍼지멤버십 함수의 조정

HIIA-1a에서는 형용사를 연산자로 사용하기 위해서 퍼지멤버십 함수를 사용한다. 가령 사용자가 '큰 포유류(heavy mammal)'이라고 질의를 하게 되면 시스템은 HEAVY 멤버십 함수의 값이 임계값을 넘는 객체들(고래, 코끼리, 곰 등)을 제공하게 된다. 만약 사용자가 제공된 문서들에서 '고래'에 대한 문서에 '만족'이라는 표시를 한다면 그것은 사용자의 '크다(heavy)'는 개념이 시스템의 '크다'는 개념과 부분적 일치한 것을 알 수 있다. 하지만 사용자가 '곰'에 대해서 '불만족'이라고 한다면 그것은 사용자의 개념과 시스템의 개념이 부분적으로 불일치하다는 것을 알 수 있게 된다. 따라서 다수의 사용자에 대해서 동일한 질의에 대해서 학습을 계속적으로 반복함으로써 대다수의 사용자를 만족시킬 수 있는 개념을 시스템이 갖게 된다. 퍼지멤버십 함수를 조정하기 위해서 본 시스템은 다음과 같은 작업을 한다.

시스템은 학습을 통해 사용자에게 제공한 결과로부터 관심 동물과 비관심 동물에 대한 피드백 정보를

갖고 HEAVY 함수를 조정하게 된다. 이런 조정에 대한 규칙은 (표 2)와 같다. HEAVY 함수의 α_3 , β_3 값은 학습이 이루어질 때마다 관심을 표현한 정도에 따라 일정 크기만큼 범위가 증·감되어 진다. 일정 길이(INTERVAL)는 초기에 계층 구조상의 데이터 집합에서 값들 사이의 간격이 가장 작은 값을 2등분한 값이 된다.

- ① HistoryDB에서 학습하지 않은 형용사를 사용하는 질의를 검색한다.
- ② 각 질의에 대해 다수의 사용자가 공통적으로 만족하는 문서들을 추출한다.
- ③ 추출된 문서들에서 질의 처리 결과로 제공된 객체들이 존재하는지를 검사한다.
- ④ 이러한 검사를 통해서 관심 동물과 비관심 동물을 구분한다.
- ⑤ 질의 처리 결과로 제공된 모든 객체들의 문서들 상의 존재여부 검사 결과를 갖고 일정한 규칙에 의해서 퍼지멤버십 함수를 조정한다.
- ⑥ 아직 학습하지 않은 질의들에 대해서 ② ~ ⑤를 반복한다.

결과로 제시한 동물에 대한 사용자 관심도	α_3	β_3
100%	INTERVAL만큼 범위 증가	INTERVAL만큼 범위 증가
75% 이상	학습 없음	학습 없음
75% ~ 50%	INTERVAL만큼 범위 감소	학습 없음
50% ~ 30%	학습 없음	INTERVAL만큼 범위 감소
30% 이하	INTERVAL만큼 범위 감소	INTERVAL만큼 범위 감소

(표 2) 퍼지멤버십 함수의 조정 규칙

(표 3)은 퍼지멤버십 함수에 대한 학습의 결과이다. 원 질의로 주어진 "heavy bear"에 대해서 결과로 "brown bear, polar bear"를 제공하게 된다. 제공된 결과에 대해서 사용자가 모두 만족한다고 의사를 표시한 경우 HIIA-1a는 사용자의 반응을 학습해서 퍼지멤버십 함수 조정 규칙에 의해서 값을 조정하게 된다. 조정된 결과에 의해 α_3 , β_3 의 값이 조정되어 있

다. 이 값은 같은 질의에 대해 또 다른 사용자의 피드백 정보를 대상으로 계속 학습되게 된다. 두 번의 학습을 거쳐 무거운 곰에 대한 질의에 대해 사용자 피드백 정보가 모두 만족할 경우에 결과로 "giant panda"가 추가됨을 알 수 있다. 초기 INTERVAL 값은 20(kg)을 갖는데, 이는 초기에 곰(ursidae)의 경우 가장 폭이 좁은 값이 40(kg)이기 때문에 이 값을 이등분한 값 20으로 고정되어 있다.

	질의 'heavy bear'에 대한 결과 값	α_3	β_3
학습 전	brown bear, polar bear	117.5	400.0
첫 번째 학습 후	brown bear, polar bear	97.5	380.0
두 번째 학습 후	brown bear, polar bear, giant panda	77.5	360.0

(표 3) 퍼지멤버십 함수에 대한 학습의 결과

5. 웹DB

웹DB는 웹을 통해 모아진 정보를 효율적으로 색인 및 관리를 수행하고, 효과적인 검색을 처리하기 위한 것으로 웹의 정보들 중 특정 영역에 해당하는 문서의 정보를 저장하여 특정 영역 정보의 제공처로서의 기능을 가지고, 정보 이용자에게 효과적인 정보 검색을 지원하기 위한 것이다. HIIA-1a로부터 질의를 주어 지면 질의에 관련된 문서 정보를 제공한다. WebDB 관리자는 좀 더 나은 정보를 제공하기 위하여 여러 가지의 기능들을 가지고 있다. 이런 기능들을 포함하여 서로 유기적으로 통합되어 작동되어진다.

5.1 적합 문서의 판정

웹에 산재되어 있는 정보들 중에서 특정 영역에 적합한 문서만을 저장하기 위해 특정 영역 지식 정보를 이용한 문서의 평가 작업을 수행한다. 이런 평가에 의해 특정 영역 문서를 색인 및 저장할 수 있도록 하는 자동 색인 기능을 가지고 있다. 문서의 적합성을 판단하기 위해서는 먼저 문서를 분석하여 색인어의 형태로 변환하는 과정이 필요하다. 웹 문서를 표현하기 위해 Salton의 원리를 적용하여 최종적으로 posting file을 구축하게 된다. 이 posting file은 색인어와 이 색인어의 출현빈도수(term frequency)로 구축되어지게 된다. 출현빈도수는 검색 과정에서 term weighting이 된다. 이 posting file은 문서의 적합성 여부의 판정과 웹DB로 저장을 할 때 사용하게 된다. 각 단어의 빈도수 w 는 다음 식 (1)을 적용한다.

$$w_i = 0.5 + 0.5 \cdot \frac{\log(tf_i + 1.0)}{\log(\max tf + 1.0)} \quad (1)$$

여기서 tf_i 는 i 번째 단어가 문서에 나타난 횟수이고 $\max tf$ 는 그 문서에서 출현 빈도가 가장 큰 단어의 빈도수이다. 이렇게 구축된 posting file은 평가 모듈에 보내어 관련성에 대한 평가치를 얻는다. 평가 함수는 그 문서의 posting file의 각 색인어에 대해 비교 대상이 되는 평가 단어의 집합을 비교해 가면서 문서의 적합성 평가 값인 RE(Relevance Evaluation)를 구해낸다. 이때, 평가 함수 식은 식 (2)와 식(3)과 같다. 평가 단어의 집합은 지식베이스를 통해 계층정보와 속성 정보로 나뉘어져 있다.

$$RE(D) = \left\{ \sum_{i=1}^n re(t_i) \right\} / n \quad (2)$$

$$re(t_i) = \begin{cases} \alpha * w_i, & \text{if } t_i \in DH \\ \beta * w_i, & \text{if } t_i \in DP \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

여기서 n 은 문서의 전체 색인어의 개수이고, DH 는 계층 구조 관련 색인어 집합에 속하는 색인어들의 집합이고 DP 는 속성 관련 색인어 집합에 속하는 색인어들의 집합이다. α, β 는 각 색인어 집합에 나타난 경우의 중요도 값이다. ($0 \leq \alpha \leq 1, \alpha > \beta, \beta \geq 0$) 모든 문서는 posting file을 만드는 작업과 이 평가 함수에 의해 자신의 평가 값인 RE를 가진다. 이 값은 해당 문서에 대해 DH, DP 가 가지고 있는 특정 도메인과 비교해 얼마나 도메인에 적합한 지를 판정해 주는 값이다. 문서는 RE(D)가 일정값 이상이면 적합 문서로 판정하게 된다. 다음 (표 4)는 HIIA-1a와 기존의 검색 엔진을 대상으로 문서의 적합함에 대해 실험한 결과의 비교이다. 검색 결과에서 상위 30개의 문서를 비교해보면 확장된 질의가 더욱 효과적인 웹 검색 결과를 보여주는 것을 확인할 수 있다. 다른 검색엔진의 질의는 검색엔진의 특성에 맞게 변형 후 검색을 실시하였다. 예를 들어 "MAX size animal AMONG mammal IN land"는 검색식으로 "+largest +land +mammal"으로 변형되어 처리하였다.

5.2 정보 변화 감시

많은 양의 웹 정보는 그 내용이 변화될 소지를 항상 안고 있으며, 이를 사용자가 일일이 검사하는 것은 너무나 힘든 일이다. 웹DB는 이런 정보의 변화에 민감하게 적응하도록 하여 사용자에게 최신의 정보를 제공할 수 있도록 하는 정보 감시 기능을 가지고 있다.

Query	HIIA-1a		Altavista		Excite	
	적합문서 개수	적합율	적합문서 개수	적합율	적합문서 개수	적합율
giant panda	28/30	93%	19/30	63%	15/30	50%
bamboo panda OR polar bear	29/30	97%	12/30	40%	11/30	27%
cub OF giant panda	21/30	70%	4/30	13%	9/30	30%
MAX size animal AMONG mammal IN land	25/30	83%	5/30	17%	9/30	30%

(표 4) 웹 문서 검색 결과

Smart-Spider[26]나 Netmind[27]와 같이 주기적으로 문서의 변화를 감지하여 저장되어 있는 웹 문서에 대해 항상 최신 정보의 유지가 가능하도록 하고 문서 변화에 대한 검사시간을 유연하게 제공한다. 웹 문서의 변화를 감지하여 항상 최신의 정보를 제공할 수 있는 기능을 하는 곳으로 WebDB와 연계하여 정보의 저장과 변화의 감시를 수행한다. 일정 기간이 지난 후 정보 감시기는 이 문서의 변화 여부를 검사하게 된다. 변화가 감지된 경우에는 이 문서의 재 방문을 수행하게 된다. 또한 정보 변화의 횟수를 파악하여 정보의 변화가 자주 일어나지 않는 문서의 경우 변화 감지를 검사하는 시간을 늘림으로서 유연성을 제공한다.

5.3 순위 결정

정보 이용자로부터 정보의 요구에 대해 관련 문서들간의 순위를 결정하게 된다. 순위 결정에는 각 문서의 유사도 값을 구하고 이 유사도 값을 통해 문서의 순위를 제공함으로써 해서 사용자에게 더 나은 정보의 제공을 수행할 수 있는 정보 검색 기능을 가진다. HIIA-1a로부터 확장된 질의는 3개의 파라미터로 건네 지게된다. 그 중 첫 번째 파라미터에 의해 각 단어가 나타나는 문서들을 웹DB에서 정합에 의한 탐색으로 찾아내게 된다. 이렇게 찾아낸 문서들은 질의에 관련된 문서의 집합으로 간주한다. 추출된 문서들을 대상으로 순위 결정을 위한 유사도 계산을 적용하게 된다. 이 때 유사도 계산에 참여하는 파라미터는 두 번째와 세 번째만을 사용하게 된다. 첫 번째 파라미터는 이미 정합 과정에 의해서 추출된 문서에 모두 나타나기 때문에 순위 결정 과정에서는 제외하게 된다. 이는 첫 번째 파라미터가 DF(document frequency)의 효과를 가지고 있기 때문이다. 최종적으로 주어질 질의와 추출 문서들간의 유사도 계산에 의해 각 문서의 중요도를 계산할 수 있고 이를 바탕으로 문서

의 순위를 결정하게 된다. 이 때 사용되는 순위 결정 방법은 Extended boolean model의 P-norm model을 이용하고 p의 값은 2로 처리하였다.

6. 결론 및 향후 과제

HIIA-1a는 Window95, Pentium PC 환경에서 Visual C++ 5.0과 CLIPS로 구현되었으며 사용자 인터페이스는 ISAPI를 사용해서 구현되었기 때문에 웹 브라우저를 통해서 접하게 된다.

HIIA-1a에서 지식베이스의 구축에 있어서 객체 정보와 속성 및 다른 정보들을 온톨로지에 최대한 접근한 형태로 구축함으로써 일반적인 시소러스나 지식베이스를 사용한 시스템보다는 사람의 추론과 유사하도록 진행되었으며, 또한 다른 정보 에이전트들간의 통신을 통해 정보를 공유할 수 있는 기반을 마련하였다. 지식 베이스가 이런 특정 영역에 적합한 구조적인 방식으로 구축됨으로써 사용자에게 필요한 정보를 제공할 뿐만 아니라 사용자의 의도에 보다 구체적인 추가 정보로 질의를 확장시킴으로써 문서 검색의 효율을 극대화 시켰다.

HIA-1a에서는 웹 문서를 효과적으로 검색할 수 있도록 동물 영역관련 웹DB를 구축하고 이를 기반으로 확장된 질의와 사용자 초기 질의를 일반 검색 엔진에서 테스트한 결과보다 HIIA-1a에서 검색한 결과가 효과적인 것을 확인할 수 있었다. 또한 사용자의 의도를 보다 정확하게 표현할 수 있도록 다양한 연산자를 제공함으로써 보다 효과적으로 정보를 얻을 수 있도록 하였다. 그리고 학습을 통해서 자율적이고 능동적으로 동작하게 되었다.

앞으로는 지식베이스를 완전한 온톨로지로 구축하여 인간의 추론에 최대한 접근할 수 있도록 하여 좀 더 지능적으로 동작할 수 있어야 하고, 좀 더 다양한 학습 형태를 제공해야 한다. 또한 여러 에이전트간의 정보의 공유를 원활하게 하고, 정보 교환을 위한 다중 에이전트의 처리가 가능하도록 연구되어야 한다.

참고 문헌

- [1] R. J. Bayardo et al., "InfoSleuth: Agent-Based Semantic Integration of Information in Open and Dynamic Environments", Technical Report, MCC, 1997
- [2] N. Jacobs, R. Shea, "the Role of Java in InfoSleuth: Agent-based Exploitation of

- Heterogeneous Information Resources", IntraNet96 Java Developers Conference, 1996
- [3] 이은석, "멀티에이전트 기술의 실세계 시스템으로의 응용", 정보과학회지 제 15권 제 3호, 1997
- [4] T. R. Gruber, G. R. Olsen, "An Ontology for Engineering Mathematics", Technical Report KSL-94-18 Knowledge Systems Laboratory, Stanford Univ., 1994
- [5] T. R. Gruber, "Translation Approach to Portable Ontology Specifications", Knowledge Acquisition, 5(2):199-220, 1993
- [6] M. R. Genesereth, R. E. Fikes, "Knowledge Interchange Format Version 3.0 Reference Manual", Logic Group Report Logic-92-1, 1991
- [7] T. R. Gruber, "The Role of Common Ontology in Achieving Sharable, Reusable Knowledge Bases", The Second International Conference on Principles of Knowledge Representation and Reasoning, 1991
- [8] T. R. Gruber, "Translation Approach to Portable Ontology Specifications", Knowledge Acquisition, 5(2):199-220, 1993
- [9] T. R. Gruber, "Toward Principles for the Design of Ontologies Used for Knowledge Sharing", Technical Report KSL-93-04, Knowledge Systems Laboratory, Stanford Univ., 1993
- [10] Y. Chlaramella, B. Defude, "A Prototype of an Intelligent System for Information Retrieval : IOTA", Information Processing & Management, Vol. 23, No. 4, pp. 285-303, 1987
- [11] B. P. McCune, R. M. Tong, J. S. Dean and D. G. Shapiro, "RUBRIC: A System for Rule-Based Information Retrieval", IEEE Transactions on Software Engineering, Vol. SE-11, No. 9, pp.940-945, 1985.
- [12] 신동욱, 임형목, 윤용운, 최기선, "도메인 독립 및 종속지식을 이용한 효율적 정보검색", 한국정보과학회논문지 제 21권 제 3호, 1994
- [13] 신정훈, 안윤애, 류근호, 박현주, "문헌검색을 위한 지식기반 질의 처리기 구현", 정보과학회논문지 제 3권 제 5호, 1997
- [14] 박영봉, 김민구, 이정태, "지식 기반의 정보 검색 시스템", 정보과학회논문지 제 21권 제 11호, 1994
- [15] G. Salton, C. Buckley, "Term-Weighting Approaches in Automatic Text Retrieval", Information Processing & Management, Vol. 24, No.5, pp.513-523, 1988.
- [16] B. Yuwono, D. L. Lee, "Search and Ranking Algorithms for Locating Resources on the WWW", Proceedings of the 12th International Conference on Data Engineering, pp164-171, 1996
- [17] D. Harman, "Ranking Algorithms", Information Retrieval Data Structure & Algorithms, pp 363 - 392, 1983
- [18] E. A. Fox, S. Betrabet, M. Koushik, "Extended Boolean Models", Information Retrieval Data Structure & Algorithms pp 393 - 406, 1983
- [19] M. Wooldridge, N.R. Jennings, "Intelligent Agents : Theory and Practice", Knowledge Engineering Review, 1994
- [20] <http://www.oit.itd.unich.edu/bio108/Animalishtml>
- [21] http://www.york.biosis.org/zrdocs/res_biol.html
- [22] 광대오 외 공저, "일반생물학", 형설출판사, 1994
- [23] "한국동식물도감 제 7권 동물편(포유류)", 문교부, 1969
- [24] J. W. Kimball, "김볼생물학 [제 6판]", 1996
- [25] "CLIPS reference manual", Volume I Basic Programming Guide, 1997
- [26] <http://www.smart-spider.com/rover.htm>
- [27] <http://www.netmind.com>
- [28] 오정욱, 변영태, "정보에이전트를 위한 지식 기반 (동물) 질의 처리 시스템", 한국정보과학회 추계학술발표논문집, 1998
- [29] 이용현, 변영태, 구연건, "웹 정보 검색 시스템의 설계", 한국정보처리학회 추계학술발표논문집, 1998
- [30] 이용현, 변영태, 구연건, "지식 정보를 이용한 순위 결정 방법", 한국정보과학회 추계학술발표논문집, 1998