

적응형 웹 서핑 지원을 위한 에이전트 시스템

국 형 준[†]

요 약

본 연구는 웹 서핑 지원을 위한 적응형 사용자 에이전트의 설계를 위해 사용자 데이터 수집, 데이터 처리를 통한 사용자 프로파일 구축 및 개선, 그리고 사용자 프로파일의 적용을 통한 적응 등 세 가지 이슈를 집중 연구하였다. 그 결과 웹 상에서 작동하는 적응형 사용자 에이전트를 위한 기능 정의 및 주요 구성 요소들을 설계하고 세부 모형을 구현하였다. 내부적으로는 두 개의 독립된 에이전트의 협동 체제에 의해 작업 목표를 성취한다. 이들은 각각 IIA(Interactive Interface Agent) 및 UPA(User Profiling Agent)이다. 사용자 인터페이스를 관장하는 IIA는 사용자에게 현재 웹 문서의 대강을 파악하고 나아가서 검색 질의어를 선택할 수 있게 하는 키워드 색인(Keyword Index)과, 계층 구조 방식의 사용자 검색 과정을 나타내는 제안 링크(Suggest Link)를 제공함으로써 사용자 친숙한 인터페이스 환경을 제시한다. UPA는 사용자에게 관한 정적 정보와 브라우징 행위에서 나타나는 동적 정보를 사용자 프로파일에 반영한다. 특히, 사용자 관심을 반영하는 관심 벡터(Interest Vector)의 개념을 정립하고 근접도(similarity) 평가에 의해 이들을 갱신하고 추가함으로써 사용자 관심을 동적으로 프로파일링하는 체계를 제시하였다.

An Agent System for Supporting Adaptive Web Surfing

Hyung Joon Kook[†]

ABSTRACT

The goal of this research has been to develop an adaptive user agent for web surfing. To achieve this goal, the research has concentrated on three issues : collection of user data, construction and improvement of user profile, and adaptation by applying the user profile. The main outcome from the research is a prototype system that provides the functional definition and componential design scheme for an adaptive user agent for the web environment. Internally, the system achieves its operational goal from the cooperation of two independent agents. They are IIA (Interactive Interface Agent) and UPA (User Profiling Agent). As a tool for providing a user-friendly interface environment, the IIA employs the Keyword Index, which is a list of index terms of a webpage as well as a keyword menu for subsequent queries, and the Suggest Link, which is a hierarchical list of URLs showing the past browsing procedure of the user. The UPA reflects in the User Profile, both the static and the dynamic information obtained from the user's browsing behavior. In particular, a user's interests are represented in the form of Interest Vectors which, based on the similarity of the vectors, is subject to update and creation, thus dynamically profiling the user's ever-shifting interests.

키워드 : 적응형 웹 서핑(Adaptive Web Surfing), 대화식 인터페이스 에이전트(Interactive Interface Agent), 사용자 프로파일 에이전트(User Profile Agent)

1. 서 론

최근 웹을 포함한 하이퍼미디어를 대상으로 한 사용자 적응형 인터페이스에 관한 관심이 집중되고 있다. 하이퍼미디어(hypermedia)란 텍스트, 멀티미디어 및 하이퍼링크(hyperlink)와 같은 다양한 정보로 구성된 정보 매체를 말한다[8, 12]. 적응형 하이퍼미디어(adaptive hypermedia)에 관한 연구는 크게 온라인 정보시스템[4, 9], 정보검색[2], 그리고 교육 등 응용 분야에서[1, 3] 활발히 진행되고 있다.

특히 인터넷상의 웹 정보들은 대표적인 온라인 하이퍼미디어라 할 수 있으며 본 연구의 주 대상이다. 최근 인터넷의 발전과 함께 웹 상의 다양한 정보를 사용자 적응 방식에 의

해 접근할 수 있도록 체계화하는 시스템에 관한 연구가 활발해 지고 있다. 적응형 하이퍼미디어 시스템의 적용 분야 가운데 가장 사용 효과 면에서 기대가 되는 분야가 검색 및 서핑 지원이라고 할 수 있다. 선진 기술국에서는 이를 새로운 연구 과제로 설정하여 많은 연구가 있어 왔으며, 국내의 경우에도 인터넷 인구의 증가와 정보화에 대한 수요에 부응하여 이에 관한 관심과 연구가 증대되고 있다.

본 논문은 사용자 적응형의 웹 서핑 지원을 위한 에이전트 시스템에 관한 연구의 결과를 본 연구에서 구축한 시스템의 설계 내용과 함께 제시한다. 논문의 구성은 다음과 같다. 먼저 2장에서는 본 연구의 접근 방식을 관련 연구와 함께 소개한다. 3장에서는 본 연구에서 구현한 시스템 Auto-Naver의 설계 원리와 구현 내용을 제시한다. 이어 4장에서 본 연구에 수반된 몇 가지 연구 이슈에 관한 상이한 접근 방식에 관한 고찰과, 5장에서 보완점과 향후 연구 방안을 논

* 이 논문은 1997년도 한국과학재단의 특정기초연구과제 (과제 번호 : 97-0100-0801-3) 연구비에 의해 지원되었음.

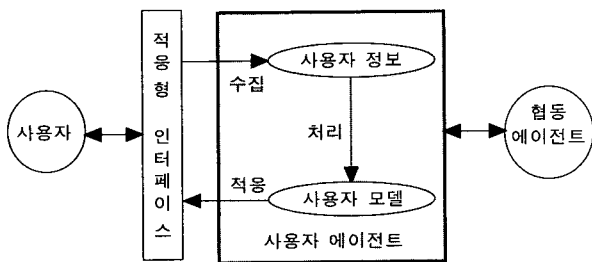
† 종신회원 : 세종대학교 컴퓨터공학과 교수

논문접수 : 2001년 11월 5일, 심사완료 : 2002년 8월 16일

한 후, 6장에서 결론을 제시한다.

2. 관련 연구 및 접근 방식

적응형 웹 서핑을 위한 에이전트의 성격과 기능은 여러 다른 관점에서 연구되고 정의될 수 있다. 본 연구에서는 이를 (그림 1)에 보인 바와 같이 세 단계로 나누어 접근한다. 사용자 데이터 수집 단계, 데이터 처리를 통한 사용자 모델 구축 및 개선 단계, 그리고 사용자 모델의 적용을 통한 적용 단계가 그것이다. 아래에 각 부문에서의 주요 연구 이슈, 관련 연구 및 본 연구가 적용한 접근 방식을 소개한다.



(그림 1) 적응형 웹 서핑 에이전트 개념도

2.1 적용 대상

적용 대상은 사용자가 보유하는 어떤 요소(feature)를 대상으로 시스템(사용자 에이전트)이 적용할 것인가 하는 문제이다. 이 요소들은 궁극적으로 사용자 모델을 구성하는 요소가 되며 시스템은 사용자 모델에 표현된 각 요소의 현재 값(수치, 기호, 또는 구조적 값)에 의존하여 개별 사용자에 대한 사용자 에이전트의 행위를 결정한다. 적용 대상으로는 주로 사용자의 지식[1, 4, 13], 목표[7, 10], 배경 및 경험 [1, 4] 등이 사용될 수 있다.

본 연구에서는 사용자 프로파일 에이전트라는 개념을 도입하여 사용자에게 관한 정보를 정적 데이터와 동적 데이터의 두 가지로 구분하여 저장한다. 정적 데이터는 사용자의 나이, 직업, 학력, 학습 주제, 관심 분야 등을 기호 정보로 표현한 것이다. 동적 데이터는 사용자 브라우징 행위를 추적, 분석한 데이터로서 관심 키워드들의 목록인 관심 벡터(Interest Vectors) 형태로 표현된다. 관심 벡터는 사용자 프로파일의 주 내용으로서(키워드, 가중치) 쌍으로 구성된다.

기존 시스템이 검색 세션의 중간에 개입하여 사용자 관심의 변화를 직접적인 피드백 방식에 의해 수집하는데 반해 본 연구에서는 사용자의 브라우징이 계속됨에 따라 시스템의 직접 개입 없이 벡터의 내용을 갱신하거나 새로운 벡터를 추가함으로써 특정 분야에 대한 사용자의 관심이 증진되는 양상과 새로운 분야로 관심이 이동하는 양상을 자연스러운 방식으로 표현하게 된다.

2.2 적용 수단

적용 수단은 시스템의 어떤 요소(feature)를 적용에 사용

하는가 하는 문제이다[11, 12]. 웹과 같은 하이퍼미디어 인터페이스에서 사용 가능한 요소는 크게 텍스트, 멀티미디어 항목과 같은 내용(contents) 요소와 링크의 존재, 링크의 순서, 링크 주석, 맵 구조 등 항해(navigation) 요소로 구분하여 생각할 수 있다[1, 4, 9].

효과적인 적용 수단으로서 본 연구에서는 이 가운데 링크 순서와 주석을 통해 웹 항해를 지원한다. 웹 항해 지원은 사용자에게 의해 입력된 키워드를 기반으로한 검색을 거쳐 얻어진 웹 문서 링크들 가운데 사용자 프로파일과 연관성이 높은 것들을 추출 제시하는 것을 말한다. 제시 목록은 사용자 관심과 관련도가 높은 순서로 제시됨으로써 사용자의 선택을 일차적으로 도울 수 있다. 또한 사용자는 순서화된 목록 가운데 자유롭게 링크를 선택할 수 있음은 물론, 깊이와 너비 우선이 혼합된 임의의 순서에 의해 링크 방문을 계속할 수도 있다. 이와 함께 사용자의 필요에 의해, 관심있는 웹 문서의 주요 어휘(term)들을 추출, 화면에 표시함으로써 사용자 하여금 현재 문서의 대강을 인지하도록 하는 동시에, 키워드 색인으로부터 관심 키워드(들)를 손쉽게 선택하여 다음 검색의 질의어로서 사용될 수 있도록 한다.

본 연구 모델이 제공하는 키워드 색인은 순서화 및 계층화된 웹 문서 목록과 함께 다양한 항해 수단을 제공함으로써 질의 입력, 그리고 이에 따라 검색된 웹 문서 목록에 주로 의지하는 기존 방식에 비해 시스템의 사용성과 사용자 적응성을 높이는데 일조한다.

2.3 사용자 모델링

사용자 모델링은 적용의 대상이 될 사용자 정보를 어떻게 획득하는가 하는 문제이다. 사용자 모델에 의하지 않는 적용이란 기대하기 어려우며, 이의 구축을 위해서는 크게 사용자의 개입이 없는 자동적인 기법과 에이전트-사용자 간의 협동에 의한 기법으로 나누어 생각할 수 있다[1, 4-6, 8].

본 연구에서의 사용자 모델링은 별도의 사용자 프로파일 에이전트에 의해 수행된다. 사용자 프로파일 에이전트는 초기 사용자에게 대한 데이터 수집을 위해서는 대화 화면을 통해, 이후에는 사용자의 브라우징 행위를 백그라운드에서 지속적으로 모니터링한 데이터를 사용하여 사용자 프로파일(User Profile)을 갱신해 나간다. 모니터링 대상이 되는 사용자 행위는 대화식 인터페이스 에이전트가 수집, 전달해 주는 동적 브라우징 행위를 말하는데, 여기에는 특정 웹 문서에 대한 사용자의 '관심있음(Mark As Cool)' 표시나 사용자의 검색 질의어 제공이 포함된다. 한편 북마크, 저장, 인쇄 등 웹 브라우저 자체의 명령 버튼을 이용한 관심 표시는 본 연구의 장기적 모델에는 포함되어 있으나 MS Explorer, Netscape Navigator 등 특정 브라우저에 의존적인 구현을 필요로 하는 관계상 현 시점에서는 본 연구의 구현 범위에 포함되지 않았다.

사용자 프로파일 구축에 있어서 기존 시스템은 사용자 관심을 웹 문서 단위로 파악(관심, 비관심)하여 이를 사용자 프로파일에 직접 반영하는 방식을 취함으로써 처음 방

문하는 웹 문서에 대한 적응성이 낮은 한계를 가진다. 본 연구는 이와 대조적으로 관심 웹 문서의 키워드 목록을 기존의 관심 키워드 목록과 통합하고 이를 기초로 하여 관심 웹 문서를 판별하는 방식을 취함으로써 처음 방문하거나 새로운 웹 문서에 대한 적응성이 높다는 장점을 가진다.

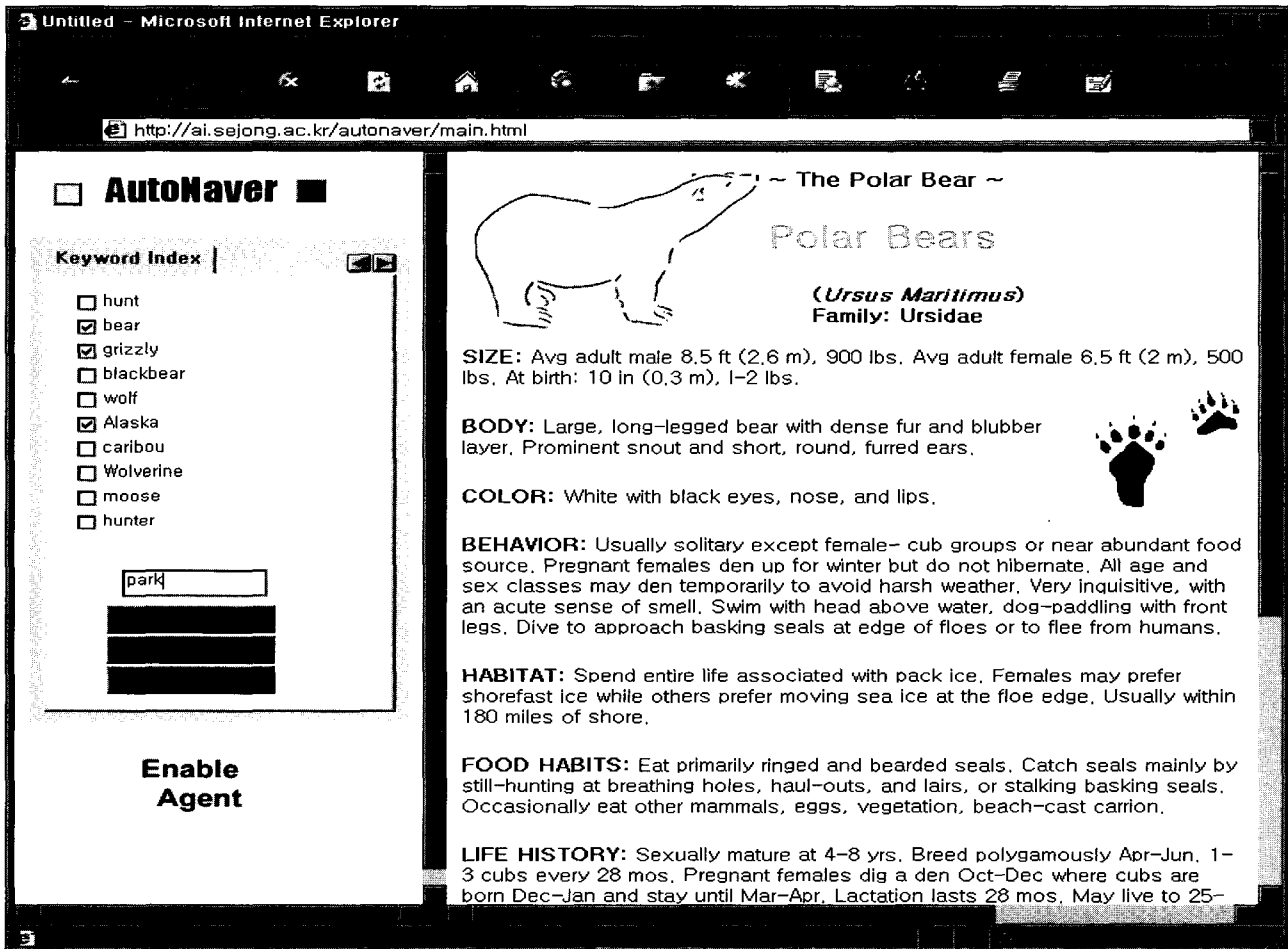
3. AutoNaver 설계 원리 및 구현 내용

이 장에서는 앞서의 접근 방식에 의해 본 연구가 구축한 시스템(AutoNaver)의 내용을 설명한다. 사용자 적응에 관련된 모든 작업은 두 개의 독립된 세부 에이전트의 협력으로 수행된다. 이들은 각각 IIA(Interactive Interface Agent) 및 UPA(User Profiling Agent)이다. 따라서 AutoNaver는 두 개의 세부 에이전트로 구성된 적응형 사용자 에이전트 시스템이라 할 수 있으며, AutoNaver 서버에 접속하면 이 가운데 IIA는 서버로부터 다운로드되어 클라이언트에서 작동하는 에이전트로, UPA는 서버에서 작동하는 에이전트로서 상호 연계 수행하여 사용자 적응형의 웹 브라우저를 가능하게 한다. 이어지는 절에서는 본 연구의 핵심적인 내용을 구성하는 이들 에이전트 각각에 대한 설계 및 구현 내용을 제시한다.

3.1 대화식 인터페이스 에이전트(IIA)

IIA(Interactive Interface Agent)는 AutoNaver 시스템의 중심에 위치하는 에이전트이다. 그 이유는 IIA가 사용자 인터페이스 및 다음 장에서 설명할 사용자 프로파일 에이전트(UPA), 그리고 외부의 협동 에이전트(Facilitator)와의 통신 채널을 유지하며 작동하기 때문이다. 이 가운데 협동 에이전트는 본 연구와는 별도의 연구로 수행된 에이전트 시스템으로서, 본 연구의 시스템과 연계하여 작동하도록 설계 구현되었다. 다음은 각 채널에 관한 수행 내용이다.

먼저, IIA는 적응형 사용자 시스템 내의 세부 에이전트로서 시스템 내외의 다른 세부 에이전트들과 필요시에 통신을 수행한다. 먼저, 사용자의 '탐색(Search)' 요구나 '관심 있음(Mark-As-Cool)' 표시 행위는 적응형 사용자 에이전트 시스템의 서버에서 작동하는 UPA에게 메시지로 전달되어 사용자 정보를 갱신하는데 쓰인다. 여기서 탐색 요구는 키워드들을 수반하는데 이들은 현재 웹 문서의 키워드 목록 중 선택된 키워드들 및 키워드 입력창에 직접 입력된 키워드들을 말한다. IIA의 탐색 요구는 궁극적으로 외부의 협동 에이전트에게 전달되어 수행된다. 수행의 결과는 URL 정보의 형태로 전달되는데 구체적으로는 URL 목록과 각 URL



(그림 2) AutoNaver의 인터페이스 화면

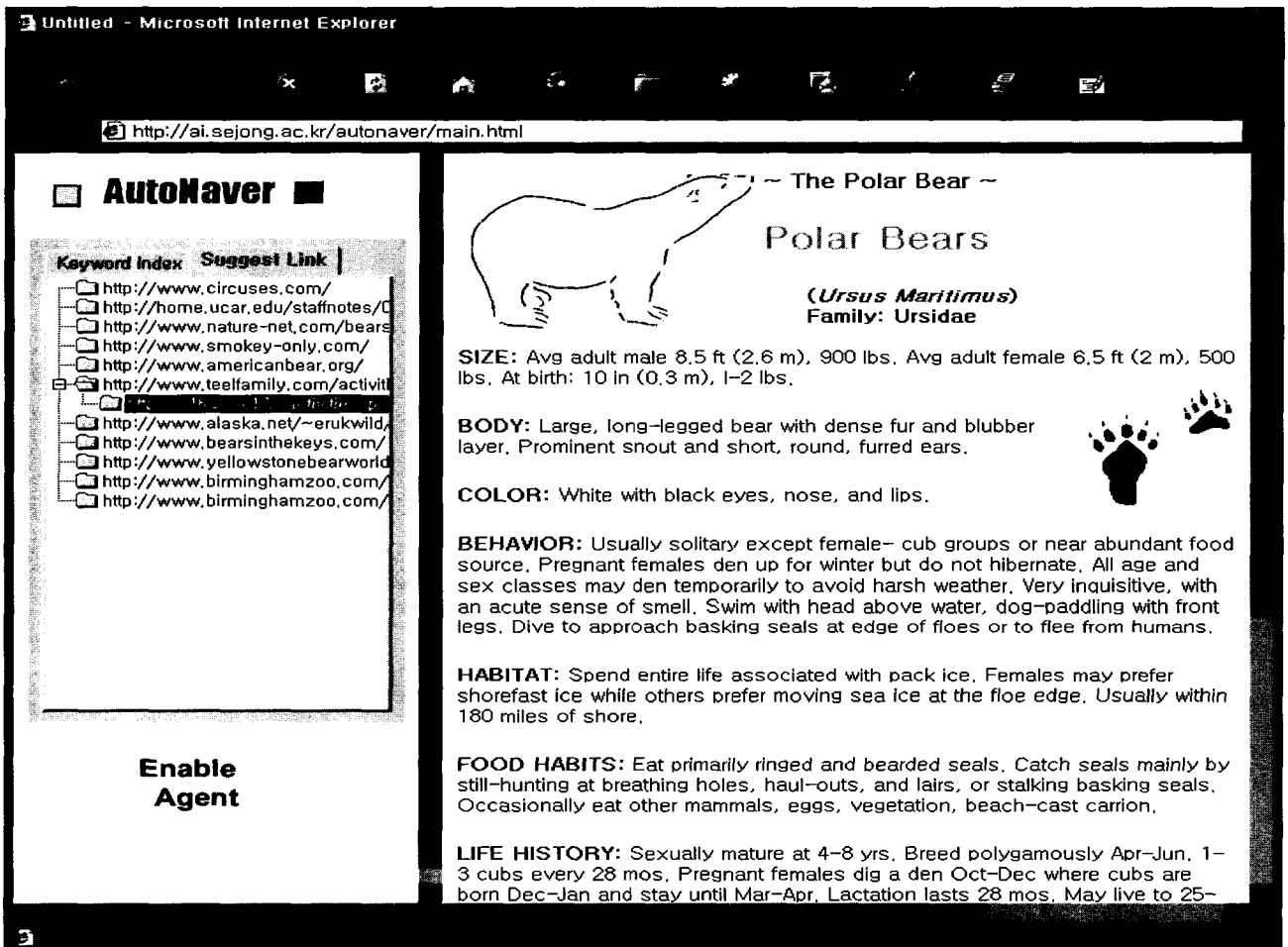
웹 문서의 어휘 벡터(Term Vector)들로 구성된다. IIA는 이 URL 목록 가운데 사용자 관심과 부합하는 일부 목록만을 추출, 사용자에게 제안한다(Suggest Link).

다음, IIA는 사용자 인터페이스를 관장하는 역할을 수행한다. (그림 2)는 IIA는 사용자의 직접적인 명령을 대기하고 수행하기 위한 사용자 인터페이스 화면을 보인다. 인터페이스 화면은 크게 좌우의 두 영역으로 나누어지는데, 오른쪽은 현재 웹 문서(Current Webpage) 출력 영역이며, 왼쪽은 키워드 색인(Keyword Index) 혹은 검색된 URL 목록(Suggest Link)의 출력을 위한 탭 영역이다. 아래에 각 영역별 기능을 중심으로 IIA의 설계 내용을 구체적으로 기술한다.

- 현재 웹 문서(Current Webpage) 출력 영역: 일반 웹 브라우저에서와 동일한 브라우징 화면으로, 현재 사용자가 브라우징하고 있는 웹 문서를 표시한다. 사용자가 스크롤 바를 이용하여 문서를 읽거나, 문서내의 링크를 선택함으로써 현재 웹 문서를 갱신할 수 있는 점등은 일반 웹 브라우저에서와 같다. 여기에는 사용자가 직접 선택한 웹 문서가 출력된다. 현재 웹 문서 영역에 표시되는 웹 문서는 항상 IIA에 의해 문서 여과(filter)를 거쳐서

표시되게 되는데, 이는 문서 내의 하이퍼링크에 대한 사용자의 선택 행위를 IIA가 모니터링하기 위해서이다.

키워드 색인(Keyword Index) 탭 영역: 키워드 색인 탭은 하단에 'Mark As Cool', 'Extract Keywords', 'Search' 등 세 개의 버튼을 제공한다. 이 중 'Mark as Cool'은 현재 문서에 대한 사용자 흥미 여부를 시스템에게 알려주는 역할을 한다. 'Extract Keywords'는 현재 웹 문서 내에 포함된 주요 어휘들의 목록을 추출하고자 할 때 사용된다. 이 버튼이 선택되면 IIA는 현재 문서에 포함된 어휘(term)들을 왼쪽의 키워드 색인 탭 영역에 어휘의 가중치(weight) 순으로 나열한다. 이와 같은 키워드 목록은 현재 웹 문서에서 사용되고 있는 주요 어휘들을 일목요연한 형태로 제시해 줌으로서 사용자가 현재 웹 문서의 성격을 신속히 판단할 수 있도록 도와준다는 이점을 가진다. 여러 페이지에 걸친 복잡한 내용을 포함하는 웹 문서의 주요 내용은 일반 텍스트, 그림 등과 섞여 한눈에 식별하기 어려운 때가 많다. 이런 경우에 가중치 순으로 나열된 어휘 목록은 사용자에게 편리한 보조 정보가 될 수 있다는 것이다. 본 설계의 또 하나의 특징은 어휘 목록이 현재 웹 문서내의 어휘들을 단순 나열하는



(그림 3) 제안 링크(Suggest Link) 탭이 보여진 화면

데 그치지 않고 현재 웹 문서에 기초한 심층적인 추가 검색을 지원한다는 점이다. 즉, 체크박스 및 입력창을 사용, 추가 검색 질의어를 편리하게 생성할 수 있게 한다. 제안 링크(Suggest Link) 탭 영역: 이 영역에는 키워드 색인 탭 영역에서 탐색 요청에 대한 응답으로서 검색된 웹 문서들의 URL 가운데 사용자 관심과 가까운 것들만 추출되어 제시된다(그림 3 참조). 사용자가 이 가운데 한 URL을 선택하면 해당 웹 문서가 인터페이스 화면의 현재 웹 문서 영역에 불러 오게 된다. 한편, 사용자의 브라우징 경로를 보이기 위한 수단으로서 제안 링크의 URL들은 단일 계층 구조의 목록이 아닌 트리 형태의 복합 계층 구조의 목록으로 제시된다. 즉, 깊이 우선과 너비 우선을 임의로 병행하는 방식의 사용자 브라우징 행위를 제안 링크에 트리 형태로 도식화되는 것을 말한다. 계층 구조상의 노드는 + 또는 - 버튼에 의해 확장 및 축소 가능하다. 이러한 경로 제시는 일반 브라우저에서 뒤로 가기 또는 앞으로 가기 방식으로 제공하고 있는 직선적 방문 경로 표시에 비해, 사용자의 무질서하기 쉬운 브라우징 경로를 보다 체계적이고도 입체적인 형태로 제시하는 이점이 있다.

3.2 사용자 프로파일 에이전트(UPA)

사용자 프로파일(User Profile)은 브라우징이 진행됨에 따라 동적으로 변화하는 사용자 관심을 반영하는 자료 구조이다. 본 연구에서는 사용자 관심을 크게 두 가지 유형의 정보로 나누어 고려한다. 이들은 각각 정적 정보(static information)와 동적 정보(dynamic information)이다. 먼저, 정적 정보는 사용자에 관한 기초 신상 정보로서 나이, 학력, 전공, 직업, 학습 주제, 관심 분야 등과 같이 현재의 로그인 세션을 통해 변하지 않는 정보를 의미한다. 이와 같은 정보들은 UPA(User Profiling Agent)에 의해 로그인 세션 초기에 사용자와의 직접 대화를 통해 수집된다.

다음, 동적 정보는 현재 로그인 세션 동안 IIA에 의해 수집, 전달된 사용자의 브라우징 행위에 관한 정보를 말한다. 여기에는 특정 문서에 대한 관심 표시(Mark As Cool)나 검색 질의어 제시 등 명시적 관심을 표출하는 행위가 포함된다. 이와 같은 정보들은 앞서의 정적 정보와는 달리, 브라우징의 실시간에 동적으로 분석되어야 한다는 특성을 가진다. 따라서 UPA는 이들을 백그라운드에서 수행하는 양식으로 처리하여 그 결과를 사용자 프로파일을 갱신하는데 사용한다. 이어지는 절에서는 UPA에 관련한 세부 내용을 설명한다.

3.2.1 사용자 프로파일(User Profile)

UPA는 사용자의 명시적 관심이 표시된 현재 웹 문서에 대해 해당 웹 문서의 어휘 벡터를 추출한다. 어휘 벡터는 (어휘, 가중치) 쌍의 집합으로 해당 웹 문서가 가지는 주요 어휘를 빈도 수에 의거하여 나열한 자료 구조로서 정보에

이전트에 의해 수집 저장된 데이터이다. 복수의 검색 질의어인 경우 백분율로 균일하게 분배된 수를 각 검색 질의어의 가중치로서 부여함으로써 이에 상응하는 어휘 벡터를 생성한다. 최초 사용자의 경우 첫 로그인 세션에서 수집된 관심 키워드들을 역시 백분율로 분배하여 최초의 어휘 벡터를 생성한다. 이 같은 방식에 의해 얻어진 어휘 벡터는 사용자 프로파일의 내용을 초기화 및 갱신하는데 사용된다.

사용자 프로파일은 사용자별 텍스트파일로서 사용자의 관심 분야를 나타내는 정보가 복수개의 관심 벡터의 형태로 저장된다. 한 사용자는 초기에는 최초에 입력한 관심 분야 키워드들로부터 생성되는 단 하나의 관심 벡터만을 가지게 되지만 검색이 거듭되면서 벡터 수는 늘어나게 된다. 아래의 예에서 보듯이 관심 벡터는 고유 번호, 갱신 회수, 키워드 개수, 그리고 (Keyword, Weight) 쌍 집합으로 구성된다.

```
((Inerest_Vector 13)
 (Number_of_Updates 7)
 (Number_of_Keywords 5)
 ((bear 46)(alaska 25)(arctic 12)(food 11)(life 6))
 )
```

사용자 프로파일은 AutoNaver 시스템의 서버에 사용자별로 유지 관리된다. 그렇게 함으로써 사용자 접근과 관리의 용이성을 추구함과 동시에 향후 협동 에이전트 등 외부 에이전트가 필요시 그 내용을 참조할 수 있게 했다. 한편, 로컬 파일 접근의 필요성과 작업 효율을 고려하여 UPA는 동일 서버에서 작동하는 서블릿으로 구현되었다.

3.2.2 사용자 관심 처리

사용자 프로파일의 갱신은 다음과 같이 이루어진다. 먼저, 특정 웹 문서가 관심 웹 문서임이 판명된 경우 해당 웹 문서의 어휘 벡터를 추출한다. UPA는 이것을 사용자 프로파일에 저장된 관심 벡터 가운데 가장 근접도(similarity)가 높은 벡터와 합병(merge)하는 방식으로 기존의 관심 벡터를 갱신함으로써 기존 관심사를 더욱 구체화하는 효과를 얻는다. 근접한 벡터가 없는 경우, UPA는 현재의 어휘 벡터로부터 새로운 관심 벡터를 생성하여 사용자 프로파일에 추가함으로써 새로운 관심 분야에 대해 기록한다. 한편, 검색 질의어로서 주어진 키워드들에 대해서, UPA는 이들을 균일한 가중치로 나누어 (Term, Weight)쌍의 집합으로 변환한다. 그리고는 앞서와 마찬가지로 기존의 사용자 관심 벡터와 합병하거나 새로이 생성하는데 사용한다. 합병이나 생성의 마지막 과정에서는 언제나 유효한 가중치 값을 가지는 키워드들만으로 관심 벡터를 단축함으로써 관심 벡터가 불필요하게 비대해지는 것을 방지한다. 마지막으로, 이렇게 하여 합병 또는 생성되는 관심 벡터는 현재 관심 벡터로 지정되어 차후의 적응형 항해의 기초 자료로 쓰이게 된다. 이어지는 절에서는 위 처리 과정 가운데 근접도 계산

방식과 관심 벡터로부터 유효 키워드 집단을 선택하는 방식에 관해 각각 상세히 설명한다.

3.2.3 근접도 계산

근접도(similarity)는 특정 어휘 벡터가 사용자의 특정한 관심, 즉 관심 벡터에 얼마나 근접한지를 나타내는 척도이다. 근접도는 다음과 같은 수식에 의해 계산된다.

$$\text{Similarity} = \frac{\sum(\text{weight of } t \text{ in } \{T_{iv} \cap T_{tv}\})}{\sum(\text{weight of } t \text{ in } \{T_{iv} \cup T_{tv}\} - \{T_{iv} \cap T_{tv}\})}$$

where T_{iv} = terms in InterestVector whose weights are over $W_{iv\text{-mean}}$

T_{tv} = terms in TermVector whose weights are over $W_{tv\text{-mean}}$

$W_{iv\text{-mean}}$ = mean weight of the terms in InterestVector

$W_{tv\text{-mean}}$ = mean weight of the terms in TermVector

즉, 근접도는 벡터 별로 평균 가중치를 상회하는 키워드들만을 추출하고, 두 벡터에 모두 존재하는 공통 키워드들의 가중치 합에서 한 쪽에만 존재하는 비공통 키워드들의 가중치 합을 빼서 구한다.

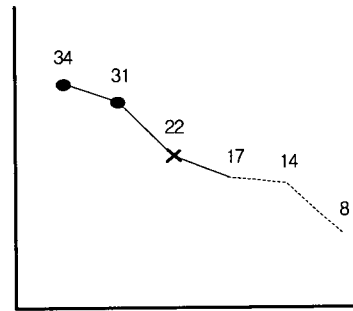
3.2.4 유효 키워드 집단의 선택

관심 벡터를 갱신 또는 생성하는 마지막 단계에서 벡터 내의 키워드들 가운데 가중치가 상대적으로 높은 것들만 추려내고 그 외의 것들은 버려야 한다. 이렇게 함으로써 어휘 벡터와 병합할 때마다 관심 벡터가 무한정 커지는 것을 방지하면서 비교적 사용자 관심도가 높은 핵심 키워드들 상대적으로 사용자 관심이 낮은 키워드들로부터 격리해낼 수 있기 때문이다. 이때 어려운 점은 어떻게 하면 관심도가 상대적으로 낮은 키워드들을 제거함으로써 사용자 관심의 집중성을 반영하는 동시에, 적정 개수 이상의 키워드를 유지함으로써 사용자 관심의 복합성을 반영하는가 하는 문제이다.

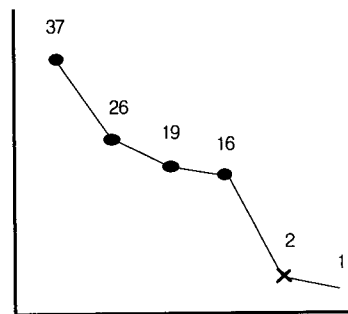
이 작업을 위해 본 연구는 다분히 경험적인(heuristic) 방식을 사용한다. 구체적으로, 벡터 내의 키워드들을 가중치가 높은 것부터 낮은 것으로 열거하고, 가중치가 처음으로 급격히 떨어지는 점을 찾아 떨어진 점에 해당하는 키워드를 포함하여 그 이후의 키워드를 모두 제외시킨다. 이를 바꿔 말하면 xy평면에 키워드들을 가중치의 내림차순으로 플롯하고 불록에서 시작하여 처음으로 오목해지는 지점을 찾아, 불록한 구간의 키워드들만 유효한 것으로 취하는 것이다.

위 원칙을 적용하는데 있어서 몇 가지 특수한 경우에 대한 고려가 필요하다. 첫째, 플롯이 오목해지지 않는 경우, 즉, 계속 불록한 경우엔 모든 점들을 취한다. 둘째, 처음부터 오목으로 시작할 경우 위 원칙을 그대로 적용하면 단 한 개의 키워드만을 취하게 되어 사용자 관심을 풍부히 반영할 수 없게 된다. 따라서 이때는 초기의 오목한 부분을 모두 취하고 다음에 나타나는 불록부터 앞서의 원칙을 적

용한다. 셋째, 플롯이 처음부터 끝까지 계속 오목한 경우, 즉, 불록한 곳이 없는 경우엔 모든 점들을 취한다. 네째, 불록도 오목도 아닌 직선구간의 경우 직선의 형태(불록 또는 오목)를 이어가는 것으로 간주하여 처리한다. 마지막으로, 처음부터 직선으로 내려가는 경우 직선 구간을 모두 취하고, 직선에서 불록 또는 오목해질 경우 여기부터는 앞서 말한 전략을 적용한다.



(a) 불록 → 오목 ...



(b) 오목 → 불록 ...

(그림 4) 플롯 형태에 따른 두 가지 예

(그림 4)는 실제 벡터를 사용한 두 가지 예에 대한 플롯을 보인다. 플롯에서 각 점은 키워드를, 점 위의 수는 가중치를 나타내는데, 이 중 ●점들은 유효한 것으로 선택되는 키워드들을, ×점들은 자신을 포함 그 이후가 버려지는 키워드들을 나타낸다. 먼저 (a)는 불록으로 시작하여 오목을 거쳐, 다시 불록으로 진행된다. 이런 경우는 첫 불록이 끝나는 부분, 즉 가중치의 차가 3→9로 증가(불록)하다가, 9→5로 감소(오목)하는데서 마지막 증가 구간의 왼쪽 점까지 유효하게 된다. 즉 플롯에서 두 번째 키워드까지 유효하게 된다. 다음 (b)는 가중치 값이 줄어들다가 늘어나는 형식, 즉 오목에서 불록으로, 그리고 다시 오목으로 가는 것을 볼 수 있다. 이런 경우는 첫 번째 불록의 마지막 구간의 왼쪽 점까지만 유효하게 된다. 따라서 네 번째 키워드까지만 유효하다.

4. 고 찰

이 장에서는 본 적응형 사용자 에이전트 시스템이 가지

는 몇 가지 이슈에 관해 본 연구가 채택한 접근 방식을 다른 대안과 함께 비교 고찰하기로 한다.

첫째, 근접성 평가에 관한 것이다. 관심 벡터와 어휘 벡터의 근접성을 평가한다는 것은 다분히 주관적인 과제라 할 수 있다. 따라서 본 연구가 제시한 방식 외에도 여러 가지 대안이 있을 수 있다. 예를 들어 평균을 상회하는 키워드에 만 기초해서 평가하기 보다 전체 키워드를 대상으로, 공통 키워드들의 가중치 합으로부터 비공통 키워드들의 가중치 합을 빼는 방식도 가능할 것이다. 그러나 이 방법은 한쪽 벡터에서는 아주 작지만 다른 쪽 벡터에서는 큰 값의 가중치를 가지는 공통 키워드가 존재할 경우 근접도를 왜곡할 우려가 있다. 다른 대안으로는 키워드의 개수를 계산에 반영하는 것인데 여기엔 두 가지 방식이 있을 수 있다. 먼저 공통 키워드의 평균 가중치로부터 비공통 키워드의 평균 가중치를 빼는 방식이다. 이렇게 하면 공통 키워드가 적을수록, 그리고 비공통 키워드가 많을수록 근접도는 높게 계산된다는 부자연스러움이 있게 된다. 이와는 대조적으로 키워드 개수만큼 가중하는 방법도 가능한데 이때는 낮은 가중치의 키워드들까지 근접도 계산에 상당한 영향을 끼친다는 문제가 있게 된다.

따라서, 다양한 대안 가운데 최선의 선택은 향후 많은 실험과 평가를 통하여 계속 연구되어야 할 과제이긴 하지만, 현재의 근접도 계산 방식은 인위적이고 부자연스러운 결과를 최소화하면서 사용자의 관심에 초점을 맞추는 적절한 방식이라고 할 수 있다.

둘째, 유효 키워드 선택에 관한 것이다. 유효 키워드를 선택한다는 것은 주관적이고 경험적인 측면이 있으므로 현재의 유효 키워드 선택 방식이 반드시 최선이라고 할 수는 없겠다. 본 연구에서는 현재 방식을 포함한 여러 가지의 대안을 생각해 보았다. 이를 정리하면 다음과 같다.

- (대안 1) 어휘 벡터와 관심 벡터의 평균 개수의 키워드 선택
- (대안 2) 관심 벡터의 갱신 회수를 반영(예를 들어 갱신한 회수만큼 개수의 키워드를 추가 선택)
- (대안 3) 어휘 벡터와 관심 벡터간의 공통 키워드는 무조건 선택
- (대안 4) 키워드들의 누적 가중치의 합이 100이 될 때까지 높은 것들을 모두 선택
- (대안 5) 가중치가 가장 급격히 떨어지는 곳을 찾아 그 전까지의 키워드들을 모두 선택
- (대안 6) 가중치가 처음으로 급격히 떨어지는 곳을 찾아 그 전까지의 키워드들을 모두 선택

(대안 1)은 어휘 벡터와 관심 벡터 내 키워드 개수의 산술 평균을 취함으로써 일견 공평 무사한 방법으로 보이지만 그다지 사용자의 관심 없는 키워드도 유효 키워드로서 선택되어 질 수가 있다. (대안 2)는 자주 갱신된 관심 벡터일수록 신뢰도가 높다는 점에서 다수의 키워드를 보존하자

는 취지이나, 벡터의 길이를 적절하게 제어할 필요가 있다는 문제는 여전히 남는다. (대안 3)은 낮은 가중치의 공통 키워드들이 지속적으로 선택될 경우 벡터의 신뢰도가 떨어지게 되고 길이 조정의 문제가 역시 남게 된다. (대안 4)는 연산의 효율 증대를 기대할 수 있는 방법이지만 하나, 위의 (대안 1~3)들과 마찬가지로 산술적인 측면에만 의존함으로써, 직관과는 거리가 먼 결과를 생성할 우려가 많게 되어 결국 벡터의 신뢰도가 떨어진다는 문제가 있다. (대안 5)는 현재의 구현 방식과 견주어 볼 수 있는 방식으로, 벡터의 키워드를 상대적으로 중요한 것과 덜 중요한 두 그룹으로 분할하여 전자의 그룹을 취한다는 취지이다. 현재 구현된 방식보다 벡터의 신뢰도를 더욱 높일 수 있는 방법일 수도 있지만 최대 감소 구간을 찾기 위해 모든 구간을 검사해야 하므로 계산 시간이 많이 요구된다. 이에 비해 (대안 6)은 현재 본 연구가 채택한 방안으로, 처음 나타나는 급격한 감소에서 분할을 수행하기 때문에 그 이후의 가중치 변화에 대한 검사를 위한 시간을 절약한다. 따라서 계산 효율과 신뢰성을 동시에 추구하는 장점을 가진다고 할 수 있다.

5. 향후 연구

이 장에서는 현재 미비하거나 개선의 여지가 있는 몇 부문에 대해 본 연구가 고려중인 향후 연구 방안에 대해 논의하기로 한다.

첫째, 다양한 동적 정보에 관한 처리이다. 현재 시스템은 'Mark As Cool'과 같은 명시적인 사용자 행위에 관해 처리하고 있다. 더욱 세밀하게 조율된 사용자 프로파일을 구축하기 위해서는 이같은 명시적 정보외에도 특정 웹 문서에 대해 스크롤 바 이동 등을 수반하는 시간 지체나 두 번 이상 같은 웹 문서를 방문하는 등의 암묵적인 행위도 해당 웹 문서에 대한 사용자 관심으로 반영할 필요가 있다.

둘째, 시스템과의 직접적인 상호 작용을 통한 가이드에 의한 정보 검색 방식은 때로는 사용자의 인지적 부담을 가중시킴으로써 검색의 흥미와 효과를 저하시킬 수 있다. 이런 관점에서, 본 시스템에 있어서 키워드 색인 및 제안 링크를 통한 항해는 사용자 관심도를 직접 반영하는 검색을 상당히 지원하지만, 학습자가 저연령이거나, 시스템 사용에 익숙치 않거나 또는 최선보다는 편리한 것을 선호하는 사용자에게는 다소 불편한 방식일 수도 있다. 이런 관점에서 본 연구는 직접적 상호 작용을 최소화할 수 있는 보다 자동화된 항해 지원 방안에도 대해서도 연구하고 있다.

마지막으로, 본 시스템 설계 모델에 대한 평가는 연구 과정에서 개발의 타당성 검증 수준에서만 이루어졌으나, 본격적인 차원에서 성능 평가는 위 개선점들을 포함한 연구와 구현이 완료되어 실용 가능한 수준의 시스템이 구축된 시점에서 보다 객관적인 평가 도구와 검증 절차를 통해 수행되어야 할 것이다.

6. 결 론

본 연구는 웹 항해 지원을 위한 적응형 사용자 에이전트를 구축하기 위한 기능 정의 및 주요 구성 요소들을 설계하고 세부 모형을 구현하였다. 현재의 설계와 구현상의 미비점을 향후 평가와 보완을 통하여 개선해 나간다면 실용 가능한 수준의 적응형 사용자 에이전트 시스템으로 발전할 수 있을 것으로 본다. 본 시스템의 작업 목표는 두 부문의 세부 에이전트의 연계아래 수행되는데 각 부문에서 성취한 결과들을 요약하면 다음과 같다.

IIA(Interactive Interface Agent) 부문에서 사용자에게 현재 웹 문서의 대강을 파악하고 나아가서 검색 질의어를 선택할 수 있게 하는 키워드 색인과, 계층 구조 방식의 사용자 검색 과정을 나타내는 제안 링크를 제공함으로써 사용자 친숙한 인터페이스 환경을 제시하였다.

UPA(User Profiling Agent) 부문에서는 사용자에게 관한 정적 정보와 브라우징 행위에서 나타나는 동적 정보를 사용자 프로파일에 반영할 수 있는 방안을 제시하였다. 특히, 사용자 관심을 반영하는 관심 벡터의 개념을 정립하고 근접도 평가에 의해 이들을 갱신하고 추가함으로써 사용자 관심을 동적으로 프로파일하는 체계를 제시하였다.

참 고 문 헌

[1] Beaumont, I., "User Modeling in the Interactive Anatomy Tutoring System ANATOM-TUTOR," *UMUAI* 4(1), pp. 21-45, 1994.
 [2] Boy, G. A., "On-Line User Model Acquisition in Hypertext Documentation," *IJCAI '91 Workshop on Agent Modeling for Intelligent Interaction*, pp.34-42, 1991.
 [3] Brusilovsky, P., Schwarz, E. and Weber, G., "ELM-ART : An Intelligent Tutoring System on World Wide Web," *ITS '96*, 1996.
 [4] Chin, D., "Intelligent Interfaces as Agents," In Sullivan, J. W. and Tyler, S. W., *Intelligent User Interfaces*, Academic

Press, pp.177-206, 1991.

[5] Cook, R. and Kay, J., "The Justified User Model : A Viewable, Explained User Model," *UM '94*, pp.145-150, 1994.
 [6] Dieterich, H., Malinowski, U., Kuhme, T. and Schneider-Hufschmidt, M., "The State of Art in Adaptive User Interfaces," In Schneider-Hufschmidt, M., Kuhme, T. and Malinowski, U. (eds.), *Adaptive User Interfaces : Principles and Practice*, North Holland, pp.269-283, 1993.
 [7] Grunst, G., "Adaptive Hypermedia for Support Systems," In Schneider-Hufschmidt, M., Kuhme, T. and Malinowski, U. (eds.), *Adaptive User Interfaces : Principles and Practice*, North Holland, pp.269-283, 1993.
 [8] Hayes-Roth, B., "An Architecture for Adaptive Intelligent Systems," *AI* 72(1/2), pp.329-365, 1995.
 [9] Maybury, M. T. (ed), *Intelligent Multimedia Interfaces*, MIT Press, 1993.
 [10] Micarelli, A. and Sciarone, F., "A Case-Based Toolbox for Guided Hypermedia Navigation," *UM '96*, pp.129-136, 1996.
 [11] Moore, J. D. and Swartout, W. R., "Pointing : A Way toward Explanation Dialogue," *AAAI '89*, pp.457-464, 1989.
 [12] Perez, T., Gutierrez, J. and Lopsteguy, P., "An Adaptive Hypermedia System," *AI-ED '95*, pp.351-358, 1995.
 [13] Wenger, E., *Artificial Intelligence and Tutoring Systems*, Morgan Kaufmann, 1987.



국형준

e-mail : kook@sejong.ac.kr

1979년 서울대학교 공과대학 졸업(학사)

1983년 Univ. of South Carolina
(Computer Science 석사)

1989년 Univ. of Texas at Austin(Computer Science 박사, Post-Doc)

1989년~현재 세종대학교 컴퓨터공학과 부교수

관심분야 : 인공 지능, 전문가 시스템, 지능형 교수시스템, 지능형 에이전트