광고학연구 : 제14권 5호.
The Korean Journal of Advertising, Vol.14, No.5 (Winter 2003).
pp.89~104.

인공지능 자동 분류체계에 따른 검색엔진의 특징과 사용자 만족도에 대한 연구

이 정 훈 · 동국대 멀티미디어과 석사과정
유 창 조 · 동국대 경영학부 광고학과 교수, 경영학박사

본 연구는 인공지능 자동 분류체계에 따라 새로이 개발된 정보인식 엔진의 특징, 핵심기술 및 기술적 과제를 소개하였고 이 새로운 정보인식 엔진을 도서 검색시스템에 적용하여 그에 따른 서비스 결과를 구현하였으며 마지막으로 이를 검색엔진 사용자에게 구현하여 기존 검색엔진과 새로운 검색엔진에 대한 사용자 만족도를 비교 및 검증하였다.
본 연구가 소개하는 새로운 정보인식엔진은 새로운 형태의 인식방식을 이용하여 인간의 도움이 없이 시스템이 스스로 언어를 학습하고 주제를 인식하여 자동으로 분류할 수 있도록 함으로써 정보구분의 효율성을 제고하였고, 새로운 인덱스를 생성하는 시간과 공간의 낭비를 제거하였으며 월등히 빠른 검색속도를 제공하고 있다. 한편, 이 새로운 시스템에 대한 사용자 만족도에 대한 조사결과도 구성에 대한 만족도와 편리성에 대한 만족도 측면에서 기존 기술시스템에 비하여 월등히 우수한 것으로 검증되었다.

1. 서 론

21세기 현대 사회에서 인터넷의 발달과 데이터베이스 구축에 따른 정보시스템발달은 엄청난 양의 정보를 축척, 생성 하게 되었다. 인터넷과 데이터 저장 시스템의 보급 및 발달 그리고 컴퓨터통신 기술의 발전은 전 세계를 디지털 정보화 시대에 접어들게 하였으며 개인의 생성 정보량은 극히 적지만 개인별 정보에 따른 연계성 및 전체 정보량은 기하급수적으로 커지고 있다. 이와 같은 엄청난 정보가 제공되는 환경에서 필요한 정보의 검색 및 활용은 사용자에게 새로운 과제가 되고 있다. 그에 따라 정보의 효율적 관리 및 사용을 위한 시스템을 구축하는 작업의 하나로 제시되고 있는 것이 정보의 자동분류 기술을 적용한 검색엔진의 개발이라고 볼 수 있다.

지금까지 제시되고 있는 자동분류 기술은 크게 실시간 자동분류와 오프라인 자동분류로 구분된다. 실시간 자동분류는 인터넷상의 정보를 실시간으로 수집하는 과정에서 동시에 많은 정보를 주어진 시간 내에 원하는 방식으로 분류해 주는 시스템을 말한다. 이에 비해 오프라인 자동분류는 문서나 정보를 의미 단계에서 분석해 관련성이 높은 정보들끼리 그룹을 만들어 주는 기법이다. 실시간 자동분류의 경우는 주로 전자상거래나 온라인 포털 서비스를 구축하고자 하는 경우에 오프라인

주제어: 인공지능 자동 분류체계, 정보구분, 인식방식,

자동분류는 중/대규모 기업에서 전자문서관리 시스템이나 지식관리 시스템을 구축하고자 할 경우 사전작업으로 많이 이용되고 있다.

2000년대의 검색엔진은 자연어 검색이나 자동분류 같은 기술이 중요한 부분이 되긴 하지만 단순히 한 가지 기술만으로 완성되는 형태는 지양되고 90년대 후반부터 인공지능 분야와의 결합이 추진되고 있고 이는 새로운 조류로 자리잡아가고 있다.

기존의 검색엔진은 'TITLE ? URL ? 요약 정보'의 패턴을 보였으나 정확한 요약 정보를 제공하기 위해서 자동요약 시스템이 필수적으로 요구되고 있으며 핵심 기술로는 검색기술내용자연어 검색지식베이스를 기반으로 질의어와 가장 유사한 질의어를 통계적인 기법이나 인공지능 이론을 통해서 결과를 제시하고 있으며 자동 분류 문서들 간의 키워드 가중치나 위치 정보 및 유사도를 기반으로 데이터를 구분해 목차를 생성시키고 있다. 그리고 사용자들의 정보나 문서에 추론엔진을 결합시켜서 새로운 정보를 생성하고 검색할 수 있도록 제공하며 다양한 시각화 기법을 통해서 체감 정확도를 높일 수 있도록 발전하고 있다.

이러한 환경에서 본 연구의 저자들은 모바일 하드웨어와 같은 축소화된 인터페이스에 적합한 새로운 검색엔진을 개발하였다. 사용검색엔진 내 다양한 프로그램 기법으로 문장의 앞뒤의 문맥을 인식해서 자동으로 분류하는 검색 알고리즘이 포함되어 검색엔진을 전부 "분류형 검색"이라고 명칭 할 수 없지만 검색결과의 도출 방안으로 분류 카테고리를 가지는 특징을 가져 이러한 형태를 가진 프로그램을 분류검색 또는 분류형 검색엔진이라고 본 연구에서는 칭하였다. 그리고 이번 연구를 위한 인공지능을 이용한 분류형 검색엔진을 정보인식엔진이라고 칭하였다.

따라서 본 연구의 과제는 새로운 검색엔진 기술의 핵심적인 내용을 소개하고 새로운 기술에 대한 사용자의 평가 조사결과를 소개하는 것이다. 본 연구는 사용자의 평가를 파악하기 위하여 기존의 검색엔진과 새로이 제작된 엔진에 대한 사용만족도에 관한 자료를 수집하여 그 결과를 비교하였다.

2. 검색엔진의 기술 발달과정과 새로운 기술의 개요

1) 검색엔진의 기술의 개요

검색엔진의 주요 기술은 현재 세대지능형 검색시스템과 자연어 검색, 자동분류시스템 그리고 최종적으로 검색결과 시각화하는 기술로 전개되고 있다. 그 중 이번 연구에서 생소한 자동 분류 시스템이라는 분야에 인공지능을 접목한 검색엔진을 개발 하였으며 이 새로운 기술의 장점은 형태소 사전 없이 인덱스 사전을 생성 시킨다는 점과 하드웨어 장비의 호환성이 우수 하며 대용량 DB 일수록 더욱 정확한 분류 결과를 도출해 낼 수 있다는 점이다.

검색엔진의 발전 방향에서 자동분류는 대용량 정보를 수집하여 동시에 많은 정보를 주어진 시간 내에 원하는 방식으로 분류해 주는 시스템을 말하며 자동분류의 경우는 주로 전자상거래나 온라인 포털 서비스를 구축하고자 하는 경우에 오프라인 자동분류는 중/대규모 기업에서 전자문

서관리 시스템이나 지식관리 시스템을 구축하고자 할 경우 사전작업으로 많이 이용되고 있다.

이번 연구에 사용 되는 엔진의 특징은 인공지능 검색 엔진 시스템을 이용한 자동 분류를 모바일 인터페이스 환경의 크기에 맞게 자동 분류화하여 PDA 인테페이스와 같이 검색 결과 화면을 클릭형 카테고리 형태로 보여줄 수 있다.

2) 엔진기술 분야 및 기존의 기술

정보인식엔진은 상세하게는 수집한 정보에 대해 신경회로망 기술을 도입한 새로운 형태의 인식방식을 이용하여 인간의 도움 없이도 스스로 언어를 학습하고 주제를 인식하여 데이터를 자동으로 분류할 수 있도록 함으로써 성장성과 환경변화에 적응력을 갖도록 하여 정보처리의 한계성을 극복할 수 있도록 정보인식엔진에 관한 것이다.

일반적으로 인터넷을 이용하는 사용자가 원하는 정보를 찾기 위해서 정보검색엔진 서버에 정보들을 입력시킨 상태에서 서버에 접속하여 원하는 정보를 요청하면 정보검색엔진을 통하여 해당 정보를 출력하도록 하였음은 이미 잘 알려진 사실이다. 정보검색엔진이 갖고 있는 여러 가지의 단점을 극복하기 위한 방안으로서 정보인식엔진을 발명한 것으로, 종래의 일반적인 정보처리를 위하여 정보검색, 인공지능, 자연언어처리, 패턴인식 등 4가지 기술 분야를 주로 이용하고 있으며, 이들 각 분야에 대해 살펴보면 다음과 같다.

첫째는 정보검색 분야로서 기존의 정보처리 과정은 수집된 데이터들에서 단어들을 색인화하여 데이터베이스에 색인된 단어를 저장하는 과정과, 상기의 저장된 색인어는 사용자가 검색어를 요청하였을 때 일치 여부를 확인하며 일치된 색인어의 데이터를 정리하여 나열해주도록 하였다. 그리고 상기의 정보검색기술은 단어들의 색인화 작업과 통계적 기법 등과 같은 자연어처리 기법과 인공지능 기법을 이용하여 데이터베이스를 형성하도록 하였으나, 수집된 데이터에서 색인어를 만들 때 색인어가 차지하는 비중을 정확히 알 수 없어서, 사용자가 검색어를 요청했을 때 제공되는 데이터들이 그 주제와는 관계없이 선택되어 서비스되었고, 이는 사용자와 제공된 데이터와의 관련성을 일일이 확인해야 한다는 단점이 있다.

둘째는 인공지능 분야로서 인공지능은 그 알고리즘에 따라 대표적인 4가지 종류로 정리할 수 있다. 퍼지 시스템, 신경회로망, 전문가 시스템, 유전자 알고리즘 등이 그것이며, 일반적인 인공지능 알고리즘을 이용하여 문제를 해결하는 경우도 있다. 이들 인공지능 중에서 신경회로망은 인간의 학습과정과 뇌의 신호전달과정을 모방함으로서 인공지능의 본질에 가장 충실한 알고리즘이라 할 수 있다.

이러한 신경회로망을 구현하기 위한 설계요소는 3가지로서 뉴런과 연결강도의 구조적 설계, 학습규칙 그리고 성능평가로 요약할 수 있다. 이들 중에서도 학습규칙은 신경회로망의 모든 부분을 결정하는 매우 중요한 역할을 하며, 새로운 학습규칙을 만들어내는 과정에서 많은 연구가 이루어지고 있다. 신경회로망의 대표적인 학습규칙으로 지도학습(Supervised Learning), 비지도학습(Unsupervised Learning), 강화학습(Reinforcement Learning) 등 크게 3가지가 있으며 신경회로망의 목표출력을 결정하느냐 결정

하지 않느냐에 따라 구분된다.

지도학습은 신경회로망의 출력이 결정되어야만 하고, 비지도 학습은 출력이 결정되지 않은 상태에서 입력데이터들의 특성을 스스로 학습해 나가는 방식인데, 이 비지도 학습은 패턴인식 분야에서 가장 활발히 응용되고 있으며 계속적인 발전을 거듭하고 있으나 인간의 언어습득 과정을 모방하기 위한 신경회로망 학습규칙은 아직까지 완성되지 않고 있다.

기존의 자연언어처리에서 응용되고 있는 신경회로망은 구문분석과 의미 분석에 이용된다. 신경회로망 알고리즘이 구문분석에서는 주성분을 분석한 병렬신호들을 처리하고 이를 기반으로 사전에서 제공하는 의미를 통해 의미 분석을 하는 병렬신호처리, 또는 패턴분류형태로 사용되었으므로 자연언어처리기법에 영향을 많이 받게 되는데, 신경회로망의 학습규칙이 자연어를 인식하는데 직접적으로 활용되지 못하고, 자연언어처리 기법이 가지고 있는 기능에 보조적인 역할을 하게 되어, 인간의 학습과정을 모방하는 신경회로망의 특성이 인간이 자연언어를 학습하는 과정과는 거리가 멀고 단지 패턴인식 기술의 일환으로 사용되기 때문에 인공지능 기법의 한계보다는 자연언어처리기법의 한계로 인한 단점에 직접적으로 영향을 받는 문제점이 있게 된다.

셋째는 자연언어처리 분야로서 자연언어처리 기술의 핵심 구성 요소는 형태소 분석기, 구문 분석기, 의미 분석기 3가지로 정리할 수 있다. 자연언어처리의 첫 단계인 형태소 분석은 형태소 분석기에 의해 수행되며 형태소 분석기는 형태소 분석 사전, 형태소 결합 규칙, 해석기로 구성되어 있다. 이 중에서 형태소 분석 사전은 사람의 수작업에 의해 형태소에 대한 정보를 입력한 것으로서 분석기의 기본이 된다. 형태소 분석기에 의하여 분석한 자연어를 구문 분석기에 전달하여 하나 이상의 자연어 단어로 이루어진 구나 문장을 구문 분석 규칙을 통하여 분석한다. 마지막으로 의미 분석 단계에서는 각 문장의 주어와 서술어를 인식하면서 미리 구축된 시소러스를 통하여 문장의 의미를 분석한다. 형태소 분석기, 구문 분석기, 의미 분석기 모두 많은 시간을 통해 데이터를 수집하고 모든 규칙을 개발자와 관리자가 계속적으로 입력해 주어야 하는 단점이 있다.

넷째는 패턴인식 분야로서 패턴인식이란 임의로 주어진 패턴을 적당한 부류로 대응시키는 기능으로 정의 할 수 있다. 이러한 패턴인식의 가장 주요한 목적은 주어진 데이터들 간의 관계를 통하여 여러 형태의 특징들을 파악하는 것이다. 패턴 인식 방법의 대표적인 방법으로는 선형대수를 이용하는 방법, 통계적 기법을 쓰는 방법, 신경회로망을 이용하는 방법 등 3가지가 대표적이라 할 수 있다. 패턴인식의 유형으로는, 주어진 데이터를 근간으로 인간이 일정한 측정 없이도 판단할 수 있는 능력을 모방한 신경회로망 기법과 많은 양의 데이터를 측정하고 통계적인 방법을 사용하는 기법이 주류를 이루었으나, 주어진 데이터들 간의 관계를 통하여 보다 많은 특징들을 파악하기 위한 새로운 패턴인식 기술의 개발이 요구되지만 그러한 패턴인식의 기술을 개발하기 위하여 그 기반이 되는 새로운 신경회로망 모델의 개발이 필수적으로 필요하게 되므로 기술의 구현이 쉽지 않게 되는 문제점이 있었다.

3) 인공지능 정보인식엔진의 기술의 개요

정보인식엔진은 수집한 정보에 대해 신경회로망 기술을 도입한 새로운 형태의 인식방식을 이용하여 인간의 도움 없이도 스스로 언어를 학습하고 주제를 인식하여 데이터를 자동으로 분류할 수 있도록 함으로써 성장성과 환경변화에 적응력을 갖고 정보처리의 한계성을 극복할 수 있도록 한 것이다.

종래의 정보를 처리하기 위한 방식은 관리자가 인위적으로 설정한 카테고리에 적합하도록 정보들을 분류함으로써 비용과 인력 및 시간상의 문제점을 가지고 있을 뿐 아니라 키워드 검색의 한계와 사전을 미리 구축해야 하는 단점을 함께 가지고 있었다. 그러므로 정보인식엔진은 사전, 일반 형태소 사전 또는 시소러스를 갖추지 않은 상태에서도 스스로의 학습을 통하여 단어와 구를 추출하는 단계와, 문장 패턴의 중심이 되는 중심어를 도출하여 언어 패턴의 유사성을 구하면서 단어와 단어의 밀접성과 관계성, 동의 및 복합명사를 반복적인 학습에 의해 그 구조를 파악하여 스스로 사전 및 시소러스를 구축하여 처리하는 단계와, 정보에서 주제어, 대주제어, 중주제어, 소주제어 및 배경지식을 뽑아내고, 주제어를 중심으로 정보를 구조적으로 분류하는 단계로 구성되며 이들 단계를 통해서 정보를 인식하고 처리하는 것이 가능하도록 한 것이다.

(1) 기술적 과제

종래 4가지 정보처리를 위한 기술의 문제점들을 해결하기 위한 방법으로 본 연구는 수집한 정보에 대해 신경회로망 기술을 도입한 새로운 형태의 인식방식을 이용하여 사람의 도움 없이도 스스로 언어를 학습하고 주제를 인식하여 데이터를 자동으로 분류할 수 있도록 하고, 성장성과 환경변화에의 적응력을 갖도록 하여 정보처리의 한계를 극복하도록 한 정보인식엔진을 개발하였다.

첫째, 수집된 데이터에서 색인어를 만들 때 색인어가 차지하는 비중을 정확히 알 수 없어 사용자가 검색어를 요청했을 때 제공되는 데이터들이 그 주제와는 관계없이 선택되어 서비스되고, 이로 인해 사용자가 제공된 데이터와의 관련성을 일일이 확인해야 하는 단점을 해결하기 위하여 새로운 신경회로망 기법을 이용하여 정보검색 분야의 한계를 극복하도록 설계하였고, 사용자에게 제공되는 데이터들의 주제를 파악하고 그 주제를 기준으로 데이터들을 분류하여 구조적으로 보여줄 수 있도록 수집된 데이터의 대주제, 중주제, 소주제를 추출하고 이들 주제와 관련된 관련 주제와 관련어들을 함께 추출하여 데이터베이스에 저장하는 시스템을 개발하였다. 따라서 사용자가 이 새로운 기술을 이용한 엔진을 이용하여 검색어를 요청하게 되면 시스템은 검색어와 관련된 대주제 분류와 이에 해당되는 중주제, 소주제 부류들을 구조적으로 보여주게 되어 사용자는 원하는 정보를 빠르고 쉽게 접근할 수 있게 된다.

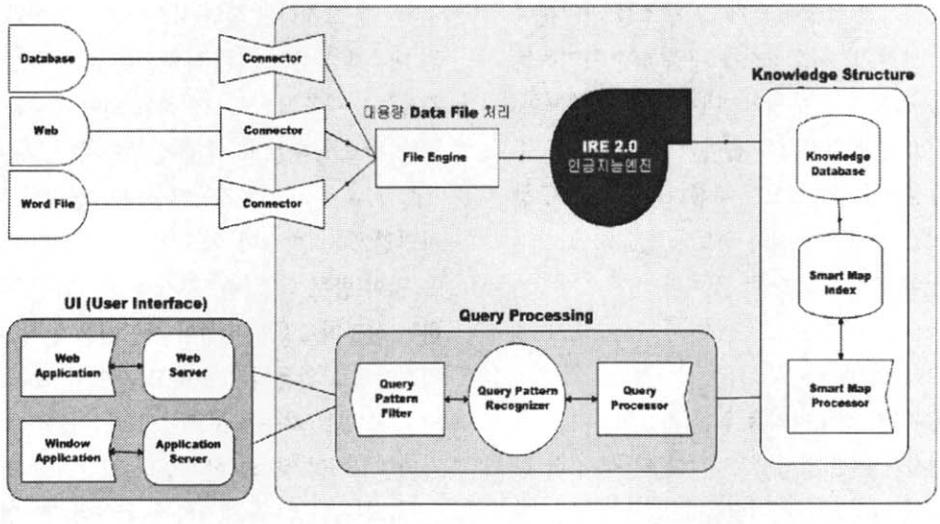
둘째, 기존의 인공지능분야에서 신경회로망의 자연언어처리방식이 자연언어처리기법에 영향을 많이 받는 단점(즉, 인간의 학습과정을 모방하는 신경회로망의 특성이 자연어를 인식하는데 직접적으로 활용되지 못하고 자연언어처리기법이 갖고 있는 기능에 보조적인 역할을 하게 되어 자연언어처리기법의 한계성의 직접적인 영향을 받게 되는 단점)을 해결하기 위하여, 본 연구는

기존의 비지도학습을 조합하여 새로운 형태의 학습규칙 즉, 인간이 자연언어를 학습하는 과정을 3개 이상의 조합된 신경회로망의 새로운 학습규칙을 활용함으로써 이 과정을 통해 스스로 자연언어를 학습하여 자연언어처리 기법의 한계에서 벗어나도록 설계하였다. 즉, 인공지능 기법을 자연언어처리 분야의 보조적인 요소로서 활용하는 것이 아니라 인공지능 기법이 주가 되어 자연어를 인간과 가장 유사하게 인식할 수 있도록 하여 자연언어처리 기법의 비효율성 등의 한계를 극복한 것이다.

셋째, 자연언어처리 분야에서는 형태소 분석에 영향을 미치는 사전과 의미 분석에 영향을 미치는 시소러스를 구축하고 처리용량을 계속 키워나가야 하는데 따르는 비용의 증가, 수작업과 인위적인 관리영역으로 인한 인력의 낭비, 시스템을 유지하거나 업그레이드시키는 데에 많은 시간이 소요되는 문제점을 해결하기 위하여, 본 연구가 제안하는 방식은 형태소 사전이나 시소러스 없이 스스로의 학습을 통해 단어나 구를 추출하는 신경회로망에 의해 인간이 자연어를 학습하는 과정을 모방함으로써 사전과 규칙 등의 구축 및 관리에 따른 기존의 자연언어처리기법의 한계를 극복하였다.

정보인식엔진은 처음부터 축적된 데이터들이 전혀 없는 상태에서 스스로의 학습으로만 자연언어를 처리하도록 하며 패턴인식 분야에서는 주어진 데이터들 간의 관계를 통하여 여러 가지 형태의 특징을 파악하는 능력을 배가시키기 위해 패턴인식 기술에 직접적인 영향을 미치는 새로운 신경회로망 모델이 요구되는 현실이므로, 하나의 신경회로망 모델만을 이용하던 기존의 방식에서 벗어나 각각의 특성에 맞는 학습규칙을 가지는 3개 이상의 조합된 신경회로망 모델을 정보인식엔진에 이용함으로서 패턴인식 기술을 발전시킨 것이다.

정보인식엔진은 패턴분류형태를 자연언어의 음절이 원소가 되는 새로운 형태의 패턴인식 기



〈그림 1〉 시스템 운영 구조도

술로서 기존의 패턴 형성과 매우 다른 형태로 신경회로망의 입출력 벡터를 제공하는 것이다.

(2) 핵심 기술

정보인식엔진을 구성하고 있는 핵심 요소는 신경회로망으로서 3개의 모델이 있으며 각각의 모델은 신경회로망 알고리즘인 ART(Adaptive Resonance Theory), SOM(Self Organized Map), LVQ(Learning Vector Quantization), BAM(Bidirectional Associative Memory)중에서 하나의 모델을 선택하여 이루어져 있고 이들 신경회로망의 학습규칙은 출력이 결정되지 않은 상태에서 입력데이터들의 특성을 스스로 학습해 가는 비지도 학습 규칙으로 설계하는 것이 관건이다. 본 연구가 제안하는 새로운 검색엔진의 시스템 운영 구조도는 <그림 1>에 요약되어 있다.

(3) 정보인식엔진의 특징

기존 정보인식 엔진은 인공지능을 기반으로 제작 되었으며 검색하고자 하는 자료가 내부 사이트에 없을 경우 가장 유사한 자료를 외부에서 검색하여 제공하며 기존 검색엔진들은 가능한 모든 조합을 갖는 인덱싱(indexing) 기술로 구현을 함으로써 불필요한 인덱스가 전체의 80%이상을 차지하는 매우 소모적이고 낭비적인 단점을 가지고 있다. 그러나 분류형 검색엔진에서의 인덱싱(indexing) 기술은 Pattern Matching 기술과 A.I 기술을 접목하여 가장 효율적이고 정확한 기술을 구현함으로 정보구분의 효율성을 제고하고 있다.

그리고 한 종류의 검색을 위해 매번 그에 적합한 모양의 인데스(index)를 만들어 내는 기존 검색엔진과는 달리 이미 나와 있는 인데스(index)들 간에 정규화를 갖는 연관 관계를 둠으로써 다차원적인 인덱스(multi-dimensional index)를 통해 새로운 인데스를 생성하는 시간과 공간들의 낭비를 제거하여 사용시 효율성과 편리성을 높이고 있다.

검색부분에서는 Query Processor의 속도 개선 Query Processing 시간을 획기적으로 줄였으며 구문검색 등의 연산속도 개선을 통해 기존 검색엔진들이 매번 복잡한 processing을 통해 제공하는 검색속도에 비해 월등한 속도 향상을 가져왔다.

3. 인공지능 분류 검색엔진의 구현 및 데모

1) 인공지능 분류 검색엔진의 기획

최근 DB 관련 기술의 발달 방향성을 살펴보면 최근 퍼지이론 및 신경망 이론 등 여러 방법을 이용하여 대용량 DB를 자동 분류하려는 노력을 기하고 있다. 그러나 알타이어계에 속하는 한국어는 영어 및 외국어에 비하여 프로그램화의 어려움을 가지고 있으며 또한 성능측면에서도 불만족스러운 결과를 나타내고 있다. 그러나 시스템의 통합화 및 DB량의 급증으로 인위적인 수작업을 통한 분류 작업은 한계에 이른 실정이며 엄청난 인건비 및 유지비에 의해 지속적 구축이 어려운 것이 현실이다.

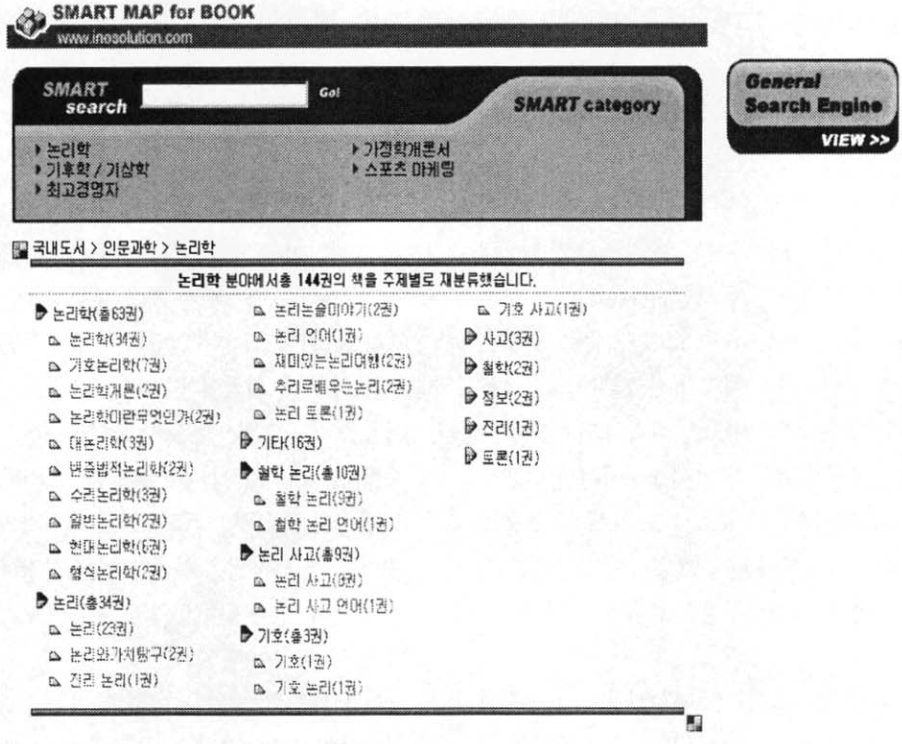
이를 해결할 방법은 어떠한 기술을 도입하든 최소한의 수작업을 통한 DB 구축 및 운영을 할

수 밖에 없다. 또한 모바일 하드웨어 장비와 호환 시점에 있어 모바일 시스템에 적용 가능하며 대용량 DB의 활용을 위해서는 형태소 사전 없이 검색기능과 함께 대용량 DB 분류가 가능한 분류형 정보인식엔진을 제작하는 것이 하나의 방안이 된다.

분류엔진의 기술기획을 통해 인공지능 전문가와 시스템 관리자, 웹 프로그램 전문가, DB 제작 전문가와 함께 인공지능 분류 검색엔진을 제작하였으며 우선 웹 사이트 중 DB량이 많은 도서 관련 사이트의 DB를 이용하여 데모를 구현하고 또한 모바일 하드웨어의 인터페이스 크기에 적합하게 프로그램을 수정 보안하여 축소 화면에서의 최적의 분류 구성을 결과를 도출하여 데모를 구현하였다.

2) 도서 검색 시스템에 분류 검색엔진의 서버스 구현

새로운 검색엔진의 기술적 가치를 가시적으로 보여주기 위하여 본 연구는 K문고의 텍스트 DB를 이용하여 새로운 시스템에 의한 검색엔지의 카테고리 분류 결과를 구현하였다. 그리고 인공 지능 데모엔진 제작과정에서 결과 카테고리 분류는 임의적인 사람의 도움이 없이 나온 결과이며 인공지능을 이용하여 형태소 사전 없이 구현 하였다.

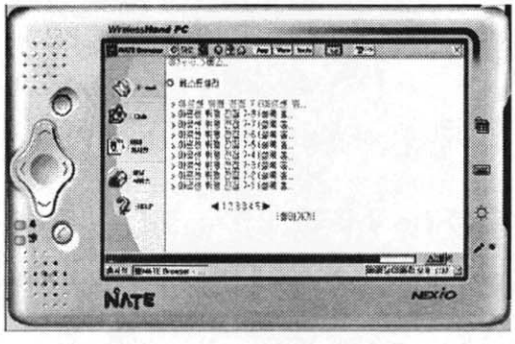


〈그림 2〉 도서시스템의 분류결과

<그림 2>는 K문고의 데이터를 이용하여 새로운 시스템의 인공지능에 따른 분류결과를 보여주고 있다.

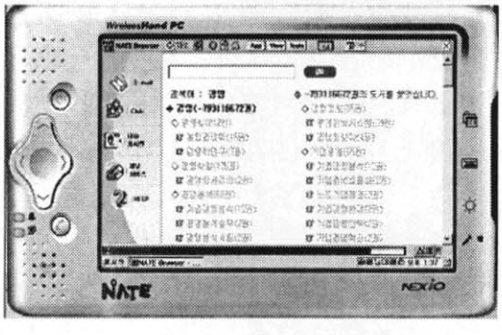
3) 모바일 인터페이스에서의 자동 분류 검색 엔진의 서비스 구현

데모 구현에서 모바일 업체인 네이트의 화면 크기 및 이미지, 카테고리 분류를 이용한 데모를 제작 하였으며 DB는 도서를 이용하여 제작하였다. 그리고 모바일 시스템에 호환성은 엔진의 특징 중 커스터마이징 지원 및 유연한 시스템 구축 지원이 가능하다는 점에 있다.



〈그림3〉 모바일 업체에서 서비스중인 리스트형식의 검색화면

자료 : 2003년 05월 NATE웹사이트



〈그림 4〉 자동 분류 검색엔진의 서비스 구현

기존 다른 시스템과의 완벽한 연동되는 이유는 검색엔진은 C언어 기반으로 API를 구현했기 때문에 UNIX기반 및 NT기반의 어떠한 시스템에도 구축이 가능하다. 추후 시스템 및 운영체제 변경에도 API를 통해 커다란 변경 없이 구축이 가능하다. 그리고 소스차원의 수정이 가능 하여 필요 시 새로운 모듈(함수)을 추가할 수 있다. 즉 검색 엔진 전체의 수정이 없이도 간단하게 관련 모듈만 수정하여 목적에 맞는 기능 및 성능 구현이 가능하여 모바일 시스템이나 어떠한 이 기종 시스템에도 적용이 용이하다

<그림3>과 <그림 4>는 N업체의 이미지 스킨을 이용한 데모를 구현한 것으로 도서 DB에 검색을 실시한 것이다. 그리고 분류는 수작업 없이 형태소사전을 사용하지 않고 인공지능에 의해 화면 크기에 적합하게 분류를 결과를 보여주고 있다.

4. 새로운 검색엔진에 대한 사용자 만족도 검증

1) 조사의 개요

인공지능 분류 검색엔진의 가치는 사용자의 평가에 달려 있다. 따라서 이 새로운 검색엔진의 편리성과 만족도를 평가하기 위하여 기존의 검색엔진과 새로운 검색엔진을 비교하기 위한 조사를 설계하였다.

비교의 정확성을 높이기 위하여 본 연구는 온라인상에서 실제 크기와 동일한 가상적 PDA 모바일 시스템에 일반적 엔진과 인공지능 분류엔진

〈표 4-1〉 설문지 배표 회수 및 이용율

구 분	설문지수	비 율 (%)
총 배부한 설문지	75	100%
총 회수한 설문지	68	90.6%
미 회수한 설문지	7	9.4%
Missing Data	8	10.6%
분석된 설문지	60	80%

을 설치한 후 응답자들에게 검색엔진의 사용을 요청하였고 응답자들은 사용 후 구성의 만족도와 편리성을 평가하였다. 현재 PDA 및 모바일 서비스에서 검색 컨텐츠가 활성화 되지 않는 상황이며 또한 검색이 부분적 컨텐츠에만 서비스 되고 있는 실정이다. 따라서 본 연구는 기술적인 성능 테스트와 광범위한 컨텐츠에 대한 사용자의 만족도 측정이 아닌 현제 서비스 되고 있는 도서 관련 사업 부분의 DB를 이용한 국내 업체의 PDA 인터페이스에 적용하여 업체에서 사용하는 일반적 키워드 형 검색엔진과 인공지능 분류 검색엔진에 대한 사용자들의 사용 만족에 대한 조사를 설계하였다.

또한 비교의 신뢰성을 높이기 위하여 한 응답자가 두가지 방식의 엔진에 대한 사용 후 평가를 하였고 그 측정치를 비교하였다. 일반엔진과 분류엔진의 제시순서에서 오는 효과를 제거하기 위하여 사용하는 순서를 통제하기 위하여 응답자의 반은 일반엔진을 나머지는 분류엔진을 먼저 사용하도록 설계하였다.

본 조사에 참여자는 하드웨어 장비 및 통신 속도, 인터넷 사용 수준이 비슷한 경북 포항의 고등학교 학생들의 협조를 통하여 온라인상으로 조사하였다. 고등학교 수업시간을 통하여 검색엔진의 사전 설명을 실시한 후 인터넷 교육 선생님의 협조와 통제에 의해 설문조사가 실시되었다. 본 설문조사의 응답자들은 기본적으로 인터넷 이용경험이 있고 일반 검색엔진을 활용하여 정보를 탐색한 경험이 있는 학생들로 제한하였다. 그 결과 본 설문조사에 모두 75명이 참가하였고 설문지 작성의 미숙으로 15명의 응답내용은 분석에서 제외되어 모두 60명의 응답이 분석에 사용되었다. <표 1>은 설문지 배포에 관한 결과를 요약하였다.

2) 사용자 만족도 요인 및 항목

검색엔진의 구현서비스에 대한 사용자의 만족도 조사에 관한 선행연구는 거의 없는 상황에서 만족도 조사에 대한 항목은 본 연구자들이 개발하였다. 본 조사에서는 만족도에 관한 항목들은 시스템적인 기술 평가가 아닌 사용자 관점에서 시스템 사용 시 시스템 구성에 대한 인지의 편리성 측면과 사용자 당사자의 사용 편리성 측면으로 나누어 개발되었는데 각 요인을 간략히 소개하면 다음과 같다.

(1) 시스템의 구성 요인

사용자가 처음 대하는 것은 전체적인 시스템의 구성이며 이것은 친밀감과 사용의 편리성에 관계된다. 또한 시스템적으로 기능적으로 우수 하여도 구성의 적정설이 뒤진다면 사용자의 시스템의 효율성이 저하될 것이다. 또한 아무리 시스템의 내용이 우수하더라고 구성적인 측면이 조잡하거나 항목의 적절한 구분이 미흡할 하다면 전체적인 시스템 자체에 대한 효율성이 떨어질 수도 있을 것이다. 시스템의 구성적인 측면에서 아래의 세부 항목들이 측정되었는데 이러한 항목들이 고려되어야 하는 이유를 아래에서 간략히 요약하였다.

• **시스템의 화면 구성**: 검색엔진의 사용자의 설문에서도 가장 처음 접하는 것이 화면구성이며 시스템을 구성하는 항목들의 간결성이 이용자의 만족감에 영향을 미치는데, 이는 항목의 중복 설계라든지 복잡성이 사용자의 불편을 초래할 수 있기 때문이다.

• **시스템을 구성결과에 사용된 용어의 적절성**: 시스템의 구성은 전문가의 작품이지만 사용자는 일반 사용자라는 입장에서 아주 민감한 문제이다. 만약 결과의 구성 용어에 부적절한 용어 사용과 어려운 전문용어의 사용으로 인하여 검색 결과의 사용이 용이하지 못할 수 있기 때문이다.

• **검색결과의 정보의 중복성 회피도**: 검색화면이 축소된 모바일 인터페이스에서 검색결과에서 중복성이 나타난다면 사용자는 일반 웹 상황에서 보다 더욱 많은 불편함을 호소할 것이며 정보검색 시간에 많은 불편이 있을 것이다.

• **하드웨어 화면크기에서 검색 이용 편리성**: 일반적 크기의 컴퓨터 모니터에서 편리성 평가가 아닌 크기가 제약된 모바일 PDA 하드웨어 인터페이스에서 사용자의 이용 편리성 측정은 화면크기에 사용 제약이 있기 때문이다.

• **검색결과의 내용 구성 형태의 만족도**: 일반적 검색결과 화면은 리스트 형식의 웹 검색결과와 같은 형태이지만 자동 분류형 검색엔진에서의 결과 화면은 내용의 이해를 통한 자료의 카테고리 형식의 결과 형태를 보여준다. 이는 축소된 화면 크기에서의 검색시스템에서 사용자의 중요한 사용 만족도의 기준이 된다.

(2) 시스템의 사용 편리성 요인

사용자의 기술적인 지원이 하드웨어 적인 시스템으로 지원 되는 상황에서 일반적 사용자는 시스템 표면에 나타난 이미지와 텍스트를 보서면 검색 시스템을 사용하고 사용시 사용자의 편리성 수준이 조사될 필요가 있는데, 조사된 세부 항목은 다음과 같다.

• **시스템의 처리 속도의 만족도**: 정보 검색시 가장 중요한 문제가 처리 속도이다. 이 항목은 누구나 공감이 가는 중요한 항목이지만 반면 사용자의 하드웨어 환경 및 거주지 지역 네트워크 수준 그리고 온라인 통신망의 부하 등 여러 불안전 요인이 있을 수 있다.

• **결과 화면에서 문항의 적절성**: 검색엔진의 사용한 도출된 결과 문항에서 간결한 단어선택을 이용한 적절한 문항 구성은 사용자에 문항 인지에 대한 만족도에 관한 것으로 인공지능을 이용한 사람의 인위적 도움 없이 프로그램에 의한 문항 결과의 적절성을 알아본다.

• 검색엔진 사용 시 사용 단어에 대한 결과 만

족도: 사용자는 학력 및 여러 요소에 의한 같은 내용의 검색을 하지만 검색단어 선정에 많은 차이가 있다. 이러한 이유로 개인별 단어 사용의 의한 만족스런 결과는 차이가 있으며 이에 대한 만족도를 측정할 필요가 있다.

• 검색결과에 대한 이동 편리성: 사용자는 검색결과를 그대로 유지하며 사용하는 것이 아닌 검색결과에서의 이동을 통하여 적절한 정보를 찾는다. 이는 미로에서 적절한 표지를 보며 빠르게 문을 찾는 것과 같은 것으로 적절한 결과 문항을 통한 결과의 이동 편리성은 중요한 만족도 요소이다.

• 검색결과에 대한 화면 구성 만족도: 검색결과에서 리스트 형식의 결과 화면과 자동 분류를 통한 카테고리 형식의 결과 화면은 기존의 하드웨어 인터페이스와 적절하게 조화를 이루어야 한다. 단지 프로그램의 성능의 우수성도 중요하지만 사용자는 하드웨어 장비와 기본 구성요소와 함께 검색결과를 보게 된다는 것은 중요한 요소이다.

5. 조사결과

1) 응답자의 특징

응답자는 모두 17세에서 20세 사이의 고등학

〈표 2〉 응답자의 특징

변수	집단 구분	빈도수(명)	비율(%)
연령	17-20	60	100
학력	고등학생	60	100
직업	학생	60	100
인터넷 사용기간	6개월미만	0	0
	6개월-1년미만	0	0
	1-2년 미만	15	25
	2-3년 미만	13	22
	3년 이상	32	53
하루 인터넷 이용시간	1시간 미만	22	37
	1-2시간 미만	32	53
	2-3시간 미만	5	8
	3-4시간 미만	1	2
	4시간 이상	0	0
인터넷 사용시 검색 사이트 이용 빈도	1-2회 미만	16	27
	2-5회	28	47
	6-10회	14	23
	10회이상	2	3

생이었으며 이들은 모두 1년이상 경험이 있었으며 하루인터넷 사용시간은 53%가 1시간에서 2시간이었고 응답자 모두는 검색사이트 이용빈도도 있는 것으로 나타나 응답자의 적합성이 문제가 없는 것으로 판단되었다. <표 2>는 응답자 특징의 개요를 보여준다.

2) 측정항목의 신뢰도 및 타당성 분석

본 연구는 수집된 자료를 통하여 검색엔진에 따른 비교분석에 앞서 측정에 사용한 만족도에 관한 항목들의 요인구조를 평가하기 위하여 요인분석(주성분 분석: principal component analysis, 회전: varimax)을 실시하였다. 본 연구에서는 각 응답자가 일반엔진과 분류엔진에 대한 만족도를 순차적으로 평가하였기 때문에 요인분석은 엔진 타입별로 수행되었다. 다음 <표 3>은 요인부석 결과와 신뢰도 계수 측정결과를 요약하였다.

위의 요인분석결과는 검색엔진 타입과 관계없이 사용자 만족도가 시스템구성요인과 시스템 편리성 요인으로 구분되고 있으며 각 요인에 속하는 항목들의 신뢰도계수가 높아 측정항목의 타당성과 신뢰성을 확인하여 주고 있다.

〈표 3〉 엔진 타입별 요인분석 결과

요인이름	측정항목	요인 1	요인 2
일반검색 시스템구성	결과구성 만족도	.763	–
	화면구성의 만족도	.749	–
	용어의 적절성	.701	–
	이용편리성	.651	–
	중복성의 회피도	.564	–
일반 검색 시스템 편리성	문항인지 만족도	–	.835
	사용단어 만족도	–	.760
	화면구성의 편리성	–	.674
	결과이동의 편리성	–	.554
아이겐 값/설명비율		2.377/47.5%	2.037/30.9%
신뢰도 계수		.932	.917
요인이름	**측정항목**	**요인 1**	**요인 2**
분류검색 시스템구성	결과구성 만족도	.859	–
	화면구성의 만족도	.853	–
	용어의 적절성	.665	–
	이용편리성	.751	–
	중복성의 회피도	.581	–
분류검색 시스템 편리성	문항인지 만족도	–	.808
	사용단어 만족도	–	.800
	화면구성의 편리성	–	.799
	결과이동의 편리성	–	.510
아이겐 값/설명비율		2.100/52.5%	1.765/34.2%
신뢰도 계수		.951	.928

〈표 4〉 엔진 타입별 사용자만족도에 대한 비교결과

구 분	기존시스템	제안 시스템	t 값	유의수준
구성에 대한 만족도	3.29	5.45	13.0	0.00
편리성에 대한 만족도	3.14	5.49	13.1	0.00

3) 검색엔진별 사용자 만족도에 대한 비교평가

검색엔지별 사용자 만족도에 대한 비교는 각 요인별 항목들의 평균값으로 평가하였다. <표 4>에서 보는 바와 같이 분류엔진의 시스템구성과 편리성이 일반엔진의 시스템구성과 편리성보다 통계적으로 유의하게 우수하게 평가되고 있는 것으로 나타났다.

6. 결 론

1) 연구결과의 요약

21세기는 디지털 정보기술의 발달로 인하여 각 분야의 융합화 현상이 가속화 되고 있으며 검색 시스템 역시 정보 저장 시스템 및 네트워크 시스템 기술의 발달 그리고 하드웨어 기술 발전 그리고 무선 시스템의 발전 및 모바일 시스템의 급속 성장 과 저변 확대 사용으로 인한 사용자의 검색 시스템의 사용 용도와 구현 환경의 변화를 격고 있다.

그러나 현제 대표적 모바일 하드웨어인 PDA에서 검색시스템은 초기적 접목 시점에 있으며 아직 구현 기술 및 컨텐츠가 도서 부분 및 몇몇 컨텐츠에만 적용되는 현실이다. 이에 이번 연구는 모바일 시스템에서의 검색엔진의 구현 방향성에 대한 연구를 통하여 최소화된 하드웨어 장비의 인테페이스 크기에 적합한 검색시스템의 개발과 사용자의 편리성에 대한 성향을 조사하였다.

이를 통하여 화면이 작은 모바일 시스템의 인터페이스에서 일반적 검색엔진의 결과 도출 화면인 나열형의 결과화면 보다 분류형 검색결과를 가진 자동분류엔진에 대한 사용자의 만족도가 우수하다는 것이 확인되었다.

2) 연구의 한계점 및 향후 연구방향

본 연구는 모바일 환경의 통신 시스템 환경에 초기 접근 시점이라는 사회 환경과 아직 검색 컨텐츠의 비 활성화된 시점에서 실험을 진행 한 점은 객관적 조사 분석을 힘들게 하는 점이 있었다. 또한 검색 시스템 역시 분류형 검색 시스템은 일반적 키워드 검색 및 문답식 검색엔진보다 일반 사용자에게 생소한 시스템이라는 점과 분류형 검색 시스템의 구현 방법이 현재 연구개발 단계이며 초기 상용화 시점이라는 점과 시스템 설계 방법의 여러 방법 중 인공지능을 이용한 주제 분류 검색엔진을 제작 방안이 있어 시스템 제작과 성능에 따른 다른 결과가 도출될 수 있는 가능성이 있다. 그리고 검색 시스템에 적용한 도서 DB가 텍스트만 18기가를 사용 한 것과 모바일 서비스가 활성화 되지 않았다는 점은 DB크기 및 서비스 종류에 따른 연구의 한계가 있다. 그러나 본 연구

의 결과는 자동 분류 솔루션에 관한 추후 컨텐츠 개발 및 회사 운영 방안에 따른 환경의 개선에 중요한 시사점을 제공하고 있는 것으로 보인다.

참고문헌

김영보(1977), "인터넷 탐색엔진의 분류체계에 관한 연구", 성균관대학교 석사학위논문.

김형주(2002), "데이타베이스 시스템", 경문사.

손소현 & 정현, "웹기반 검색엔진의 한글 색인 및 질의 처리 기술 동향"

오병택, 송종철, 손소현(2001), "검색시스템에서의 정보 시각화 기술 연구동향및 전망", 전자통신동향분석,1 6 (6): p. 83-91.

이경배(2000), "검색엔진의 현재와미래", 삼성SDS 정보기술연구소 보고서.

와이즈넷(2002), "분류형 검색엔진의 특징과 비교분석", 와이즈넷

요한 헬름(2001), "무선정보 서비스 디자인", 대청사.

정지원(2002), "인터넷 인공지능 검색엔진 보고서", 이너솔루션 보고서.

정지원(2002), "인공지능 알고리즘을 이용한 분류엔진 설계", 이너솔루션 보고서.

채서일(2000), "사회과학조사방법론", 학현사.

EC연구회외(2000), "i모드 비즈니스의 모든 것", 김영사.

광고학연구 : 제14권 5호.
The Korean Journal of Advertising, Vol.14, No.5 (Winter 2003).
pp.89~104.

Study of reference engine's characteristics and user satisfaction by mechanical brain automatic assortment system.

Junghoon Lee
Graduate Student
Dept. of Multimedia
Dongguk University

Changjo Yoo, Ph.D.
Professor
Dept. of Advertising
Dongguk University

This study introduced newly developed information recognition engine's characteristics and core technology, realized service results after applying to books reference system, and empirically analyzed user satisfaction toward the new reference engine.

Since this newly developed engine uses new type of recognition method, the system itself learns language without human's support. As a result, it raises efficiency of information assortment, removes waist with respect to time and space, and provide superior assortment speed. We also analyzed user satisfaction toward this new system. The survey results showed that this new system was superior to the existing system with respect to the satisfactions toward composition and convenience.

Key Word : Mechanical Brain Engine System, Information Assortment, Information Recognition Method.