

사용자 훈련 가능한 IHWA 전자상거래 검색 컴포넌트 시스템의 성능평가

백인천, 신재욱, 임경수
순천향대학교 정보기술 공학부

The Performance Evaluation of IHWA Electronic Commerce Search Component System

In-Cheon Paik, Jea-Wook Shin, Kyung-Soo Lim
Division of Information Technology Eng. Soonchunhyang University

요약

수많은 인터넷 쇼핑몰들과 상품검색 시스템이 생겨났지만 사용자가 원하는 상품 정보를 얻기 위해 많은 노력이 있어왔다. 각각의 쇼핑몰들의 개성과 가격, 그리고 부가 서비스를 비교하는 것 또한 일반 사용자의 입장에서 수많은 물들을 다 검색한다는 것은 매우 힘든 일이다.

제안된 IHWA(Information Harvest Warehouse) 상품검색 시스템은 정확하고 많은 정보를 사용자들로부터 직접 받아 자체 저장소에 저장할 수 있는 검색 컴포넌트 시스템이다. 본 논문에서는 IHWA 상품 검색 컴포넌트 시스템에 대한 성능평가 및 개선방향을 제안한다.

I. 서론

최근 인터넷을 중심으로한 가상공간의 형성으로 전자상거래의 새로운 문화들이 창출되면서 이에 접속하는 호스트와 사용자가 기하급수적으로 늘어나게 되었다. 따라서 인터넷상의 쇼핑몰들과 이곳을 이용하려는 사용자 또한 기하급수적으로 늘어남에 따라 보다 효과적으로 상품을 검색하는 시스템도 전세계적으로 개발되고 있다.

현재 상품정보를 제공하는 대표적인 시스템은 mysimon, excitestores, comparenet 등으로 일반 웹상에서 상품정보를 찾아 사용자에게 가격비교등의 부가 정보를 제공하지만 그 검색엔진들이 수집할 수 있는 정확하며 세부적이며 변화하는 정보를 채우는 데는 한계가 있다.

본 논문은 정보통신부의 대학 정보통신 육성·지원 사업에 의해 수행된 것임

본 논문에서는 이러한 문제점을 극복하고자 제안된 사용자 훈련가능한 IHWA 상품 검색 컴포넌트 시스템에 대해 사용자 훈련기능의 효율성과 시스템 성능을 고찰하여보고 개선방향을 제시한다.

II. 본론

1. IHWA 상품 검색 시스템 개요

먼저 IHWA 상품 검색 시스템에 대해 간략히 설명한다.

IHWA 시스템은 상품 정보 검색엔진으로 사용자가 자유롭게 인터넷을 통하여 IHWA 검색엔진을 자신의 취향대로 훈련시킬 수 있다는 것이 특징이다. 본 IHWA 검색시스템은 콜로라도 대학의 검색로봇 Harvest 시스템을 기반으로 만들어졌다. IHWA의 서버측 구조는 현재 서버측 컴포넌트의 표준인 EJB(Enterprise JavaBeans)로 제작되어 서로 다른 시스템 상에서 더욱 포괄적이며 견고한 환경을

제공한다. 따라서 IHWA 는 확장과 변형이 용이한 3-Tier 구조를 갖는 새로운 사용자 지향형 검색 시스템이다.

그림 1과 같이 IHWA 시스템은 서버와 클라이언트로 나누어진다. 클라이언트 부분은 사용자가 사용하게 되는 웹브라우저와 같은 부분이고 현재 IHWA2.0에서는 Java 어플리케이션으로 IHWA 서버 컴포넌트에 접근하게 되어있다.

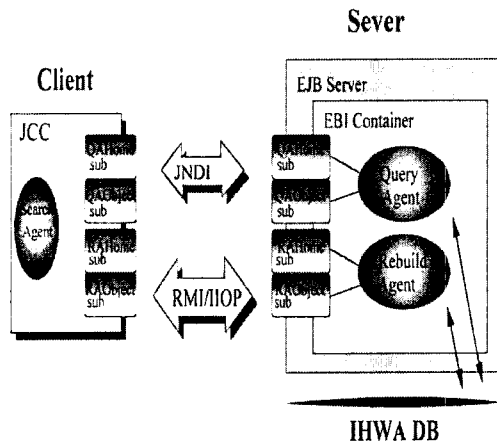


그림 1. IHWA 시스템의 구성도

서버부분은 검색엔진 전면에 서서 클라이언트 부분의 요구를 처리해 자료 탐색을 명령하는 부분이다. 서버는 QueryAgent와 RebuildAgent 두 개의 컴포넌트로 구성^[2] 되어있다.

현재 IHWA 시스템은 위에서 본 것처럼 서버측 컴포넌트의 표준인 EJB(Enterprise Java Beans)를 이용하여 설계 및 구현되어있다. 하지만 IHWA의 초기버전에서는 CORBA를 이용한 클라이언트와 서버의 통신을 하도록 했었다.

IHWA 검색 시스템의 자세한 구조는 논문^[1]에 상술되어 있다.

2. 성능분석

본 논문에서는 IHWA 상품 검색 시스템을 크게 두 가지 측면에서 성능측정을 하였다.

첫 번째로 IHWA의 가장 큰 특징이라 할 수 있는 사용자 훈련 가능한 검색 시스템이 사용자의 수와 훈련기간에 비례해 얼마나 향상된 성능을 제공하는지 테스트하였다.

두 번째로는 CORBA를 이용한 통신방법과 EJB를 이용한 서버측 컴포넌트 구조로 되어있는 IHWA를 기존의 시스템에서 제공하는 CGI와의 속도측면에서 성능에 대해 비교하였다.

2.1 사용자 수와 기간에 비례하는 IHWA 상품 검색 시스템의 성능향상

2.1.1 사용자에 의한 IHWA시스템의 정보의 증가

본 IHWA 상품 검색 시스템은 위에서 소개한 바와 같이 서버 측에서 기존의 검색엔진 위에 사용자가 개입하는 방식을 취하고 있다. 많은 사용자가 검색엔진의 효과적인 대화를 통한 능적인 탐색엔진의 훈련을 위해 참여한다. 그러므로 사용자 수와 기간에 비례하는 IHWA의 성능측정을 위해 다음과 같이 하였다.

먼저 성능측정을 위한 상품에 대한 질의어(Query)를 쇼핑몰에서 구매자가 흔히 찾을 수 있는 것들로 설정하였다.

· Query 리스트

- ① camera ② printer ③ earring ④ watch
- ⑤ bicycle ⑥ wallet ⑦ toy ⑧ telephone
- ⑨ chocolate ⑩ flower

다음으로 정상적인 사용자 50명을 선정하여 위의 10가지 Query에 대해 1주일에 3개 이상의 상품정보를 입력하도록 하였다. 여기에서 정상적인 사용자란 올바른 정보를 입력할 수 있는 사용자를 말한다.

사용자 수의 선정은 시간에 따라 선형적으로 증가하는 것으로 하였다. 사용자의 수를 시간에 비례하여 증가시킨 것은 실제 IHWA 검색 시스템이 시간이 지남에 따라 여러 사용자에게 알려진다는 가정 하에 사용자수를 증가시킨 것이다.

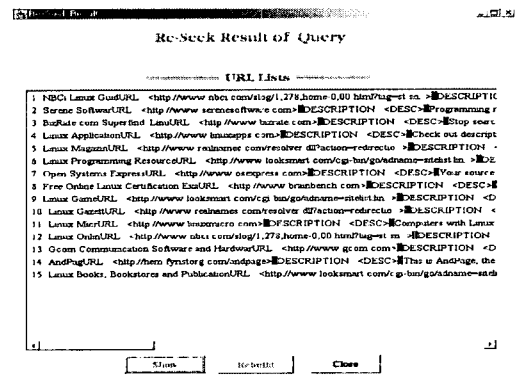


그림 2. Re-Seek Frame을 이용해 URL을 선택하는 창

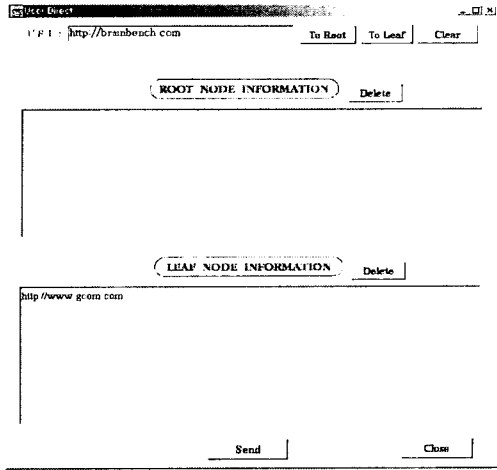


그림 3. 사용자가 IIWA 검색 시스템에 정보를 입력하는 창

그림2와 그림3은 사용자가 IIWA에 자신이 찾은 정보와 IIWA자체에서 제공하는 검색을 통해 DB에 직접 정보를 저장하는 과정이다.

위와 같이 사용자의 정보수집에 의한 IIWA의 상품 DB의 증가는 표1과 같은 결과를 도출하였다.

표 1. 사용자와 기간에 따른 수집된 정보의 양

기간 (누적된 기간)	사용자수	정보의양 (각 Query에 대한 평균)
1주	5	18
2주	10	50
3주	15	107
4주	20	178
5주	25	276
6주	30	386
7주	35	499
8주	40	651
9주	45	824
10주	50	1006

표1에서 볼 수 있듯이 시간이 경과함에 따라 사용자가 선형적으로 증가했을 경우 수집 기간의 제곱에 비례하여 정보가 수집됨을 볼 수 있었다. 이것은 수집리스트에서 검색로봇이 가져오는 정보는 제외하였고 순수한 사용자에게 의해 지정된 정확한 정보라고 할 수 있으므로 정확성있는 정보가 수집기간에 따라 급격히 증가함을 말해준다. 그림 3은 IIWA의 사용자에게 의한

IIWA 시스템의 수집량을 그래프로 나타낸 것이다.

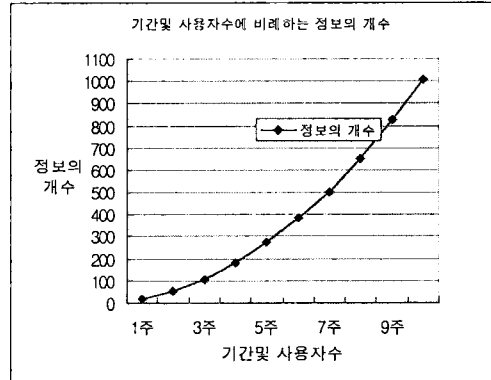


그림 4. 기간 및 사용자수에 비례하는 정보의 양

2.1.2 수집된 정보의 정확도 측정 및 분석

정상적인 사용자에게 의해 수집된 정보의 정확도를 측정하기 위해 아래의 공식을 사용하였다.

$$\text{정확도} = \frac{\text{해당 디렉토리에 속하는 상품 정보수}}{\text{검색된 총 상품 정보 수}} \times 100(\%)$$

본 논문에서는 한 상품에 대한 부가 혹은 연관되는 상품은 일단 정확하지 않은 것으로 간주했으므로 위에서 해당 디렉토리에 속하는 상품 정보수는 질의자가 원했던 정확한 상품 정보를 말한다.

위의 공식에 의거해 각 Query에 대한 정확도는 표3과 같다.

표 2. Query에 대한 정확도 분석 및 비교

query	정확한 정보의 개수	총정보의 개수	정확도(%)
camera	1084	1092	99.3
printer	946	949	99.6
earring	751	762	98.5
watch	1157	1159	99.8
bicycle	1084	1098	98.7
wallet	987	991	99.5
toy	1028	1029	99.9
telephone	1156	1167	99.0
chocolate	903	909	99.3
flower	857	865	99.0
평균	995.3	1002.1	99.3

표 3. 타 상품검색 시스템의 성능

Query	정확한 정보의 개수 / 총정보의 개수		
	mysimon	comparenet	excitestores
camera	422/500	178/239	281/449
printer	416/500	108/240	108/232
earring	494/500	136/240	41/41
watch	493/500	238/240	399/405
bicycle	192/211	103/231	11/19
wallet	185/224	164/239	36/51
toy	494/500	229/240	329/337
telephone	452/500	25/250	114/211
chocolate	403/500	213/240	181/338
flower	487/500	194/240	94/373
평균	91%	66%	65%

위의 성능평가 결과에서 볼 수 있듯이 IHWA 상품 검색 시스템은 사용자의 혼란에 의해서 타 상품검색 엔진에 비해 정확하고 풍부한 DB를 가질 수 있게 된다. 표3의 타 상품 검색 시스템에서 검색된 결과를 보면 각각의 검색시스템마다 특징이 있겠지만 관련상품이 적은 품목일수록 정확한 정보를 제공하는 것을 알 수 있다.

타 검색 시스템에 비해 IHWA 검색 시스템은 사용자가 이미 확인한 정확한 정보만을 저장하기 때문에 검색엔진의 정확성에 있어 향상된 성능을 가지고 오게 된다. 수집로봇에 의해 수집되는 정보가 포함되면 타 검색보다 정확성면에서는 높지 않게 될 것이다. 반면에 사용자의 수가 증가하면 할수록 IHWA 검색 시스템이 가질 수 있는 상품정보의 양은 더욱더 방대해진다.

2.2 IHWA 검색 시스템의 처리속도 분석

본 논문에서는 기존의 HTTP/CGI를 사용한 시스템과 CORBA를 이용한 시스템, 그리고 EJB를 이용한 서버측 컴포넌트 구조로 되어있는 IHWA 시스템을 각 Query에 대한 반응속도만을 비교 측정하였다.

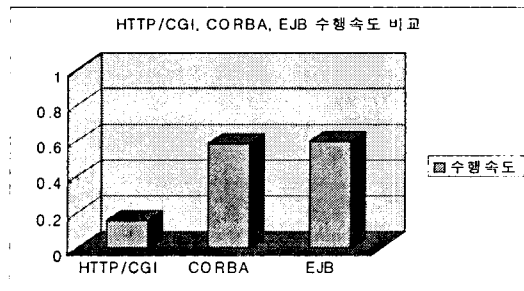


그림 5. CGI, CORBA, EJB의 수행속도 비교

그림5에서의 결과에서 알 수 있듯이 HTTP/CGI를 사용했을 때보다 EJB, CORBA에서의 시스템이 약 4배의 빠른 수행속도를 보였음을 알 수 있다.

EJB와 CORBA의 경우 거의 차이가 나지 않았고 EJB측의 미세한 우위성을 나타내고 있다.

이는 HTTP/CGI는 Apache Web Server를 통하여 데이터를 읽어오는 반면 CORBA나 EJB의 경우는 각각 Gatekeeper와 Inprise Application Server(IAS)를 이용하여 Client/Server간 통신을 하기 때문이다.

Gatekeeper와 IAS Server는 기본적으로 분산 객체 프로토콜인 CORBA IIOP를 사용하고 있고 여기서 EJB Server가 미세한 우위를 보이는 것은 네트워크의 트래픽과 접속 수를 줄여주는 세션빈으로 IHWA가 설계^{[2][4]}되었기 때문이다.

III. 결론 및 향후 연구 과제

IHWA 시스템은 사용자 혼란 가능한 상품검색엔진으로서 지금까지의 검색엔진과는 다른 개념의 검색엔진이다. 본 논문에서의 성능 평가에서도 알 수 있듯이 정상적인 사용자에게 의해 잘 훈련만 되어진다면 정확한 정보를 풍부하게 보유할 수 있으며 로컬 저장소에서 빠른 검색을 제공할 수 있다. 현재 본 실험을 위해 인위적으로 선택되었던 사용자 집단은 일반 웹에 애플릿 형태로 배치하여 좀 더 실질적으로 선택하여 시행되는 것이 필요하다. 또한 악의적인 사용자의 접근을 제한할 수 있는 체제의 보강이 필요하며 디렉토리 구조의 검색 결과를 제공해 줄 수 있는 분류처리기등이 보완되어야 할 것이다. 이미 발표한 새로운 IHWA의 컴포넌트 구조^[3]에 의해 구현을 함으로서 이러한 부분들에 대한 문제점들을 해결해 나갈 것이다. 또한 IHWA간 효율적인 정보 공유 메커니즘과 일반성을 띄는 상품검색 컴포넌트 구조의 설계가 절실히 요구된다.

IV. 참고 문헌

- [1] Incheon Paik, W Lee, "Design of Scalable User Oriented Internet Information Search System Using Distributed Object", Proceedings of TOOLS 24, Sep. 1997.
- [2] Incheon Paik, Tongwon Han, Design and Implementation of electronic commerce Search Engine Component, Proceedings of AoM/AoM 17, August. 1999.
- [3] Incheon Paik, Tongwon Han, "A Novel Component Architecture for Electronic Commerce Search System, To appear in the Proceedings of SNPD 00 Conference.
- [4] Richard Monson-Haefel, Enterprise JavaBeans, O'Reilly 1999
- [5] D.R. Hardy, M.F Schwartz, D. Wessels, "Harvest : A Scalable, Customizable Discovery and Access System", U. Colorado, Boulder, Tech. Report, Sept. 1995