

Ontology 기반의 북마크 정보 공유 에이전트: XML에 의한 접근

Ontology-based Bookmark Information Sharing Agent: XML Approach

정재은*, 조근식**

* 인하대학교 전자계산공학과

E-mail: jungjeff@orgio.net Homepage: www.intelligent.pe.kr

** 인하대학교 전자계산공학과 교수

E-mail: gsjo@dragon.inha.ac.kr

요약

기존의 BISAgent(Bookmark Information Sharing Agent)는 사용자에 의해 저장되는 북마크를 공유를 기반으로 협동적 웹 브라우징(Collaborative Web Browsing)을 통해 Information Overload를 해결코자 하였다. 하지만, 사용자의 북마크의 메타 정보 입력에 대한 부담은 Information Noise 발생의 원인이 되었고 그로 인한 부작용이 발생하였다. 본 연구에서는 BISAgent에 Ontology의 결합을 통해 사용자가 방문한 사이트의 내용 정보를 해석하여 개념화 하였으며, XML을 이용하여 표현하였다. 그럼으로써 북마크 메타 정보의 자동화 및 북마크 자체의 자동 삽입을 통해 사용자의 부담을 최소화하는 효과를 가져왔다. 실험으로써 Ontology를 이용한 일반 HTML문서의 개념화된 XML문서로의 변환을 수행하였으며 북마크 정보 저장의 자동화를 위해 사이트 방문 패턴 해석을 실시하여 시스템의 성능을 평가하였다. 양적으로 많은 사용자 북마크를 추출하지는 못했지만 추출된 북마크는 사용자의 재방문이 이루어짐으로써 상당히 높은 정확성을 얻을 수 있었다.

1. 서론

최근 웹에 대한 관심이 집중됨으로써 정보의 양이 급증하여 정보 검색에 어려움을 겪고 있다. 검색해야 할 정보의 양적 측면 뿐만 아니라 검색된 정보의 질적 측면도 심각한 실정이다. 이와 같은 정보 과적(Information Overload) 문제를 해결하기 위해 많은 연구들이 진행중이다. 사용자들의 관심사항이나 브라우징 패턴을 학습할 수 있는 개인 적응형 에이전트(Personal Adaptive Agent)나 관심사항이 유사한 패턴을 보이는 사용자들을 군집화(Clustering) 하여 해당 정보를 추천해 주는 협동적 정보 여과(Collaborative Information Filtering) 기술 등이 대표적인 연구들이며 각 기업별로는 사내에 존재하는 수많

은 정보의 효율적 관리를 위한 EIP(Enterprise Information Portal) 작업이나 데이터 웨어하우징(Data Warehousing)과 같은 엔터프라이즈 솔루션(Enterprise Solution) 개발들이 활발히 진행중이다. 이와 더불어 여러 사용자들의 북마크 정보를 공유함으로써 효과적으로 북마크 정보의 추천을 가능하게 하는 시스템에 대한 연구도 Information Overload를 효과적으로 해결하기 위한 연구들 중의 하나이다.

본 연구는 기존의 북마크 정보 공유 에이전트 시스템(BISAgent: Bookmark Information Sharing Agent System)이 가지고 있는 각 정보의 부정확성과 정보들 사이의 복잡한 상호 연관성등에 의한 정보 공유(Information Sharing) 상의 어려움들을 해결하고자

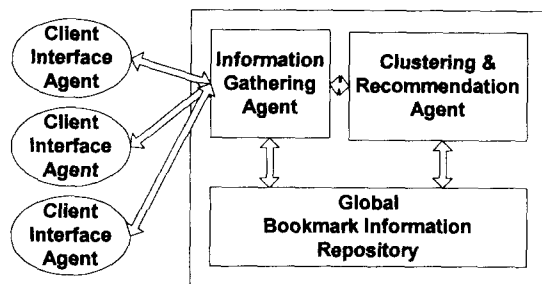
Ontology 기반 북마크 정보 공유 에이전트(Ontology-based BISAgent; OnBISAgent)을 설계 및 구현하였다. 그리고 현재 정보 시스템들의 표준으로써 각광 받기 시작한 메타 언어(Meta Language)인 XML(eXtensible Markup Language)을 이용하여 Ontology에 의해 획득되어진 북마크의 메타 정보 및 의미 정보(Meta & Semantic Information)를 표현하였다.

논문의 구성은 1장의 서론에 이어 2장에서는 연구 배경을 살펴보고 3, 4장에서는 본 시스템의 전반적인 구성 및 접근 방법을 설명한다. 5장에서는 실험을 통해 시스템의 성능을 평가하며 이어 6장에서는 관련연구를 마지막으로 7장에서는 결론을 내리고 향후 연구에 대해 기술한다.

2. 연구 동기

2.1 BISAgent의 소개

기존의 북마크 정보 공유 에이전트(Bookmark Information Sharing Agent: BISAgent) 시스템은 [그림 1]과 같이 사용자들이 웹 브라우저를 하면서 저장하는 북마크를 동일한 공간에 집중시켜 모든 사용자들이 공유함으로써 협동적 웹 브라우징(Collaborative Web Browsing)을 가능하도록 한다.



[그림 1] BISAgent 시스템 구조

협동적 웹 브라우징은 비슷한 취향과 관심분야를 가지고 있는 사람들이 서로 정보, 예를 들어 브라우징 패턴, 북마크 등을 공유함으로써 자신이 찾고자 하는 정보뿐만 아니라 예상치 못한 정보도 얻을 수 있도록 도와준다. 즉, 웹 서치 엔진을 이용하는 것처럼 능동적인 정보 검색 행위(active information searching behavior)가 아니라 협동적 웹 브

라우징 시스템의 지시를 따라가는 수동적인 정보 검색 행위(passive information searching behavior)라 할 수 있으며 시스템 측면에서는 사용자에게 정보 추천(information recommend)의 입장이라 할 수 있다.

사용자들은 BISAgent의 사용자 평가 모델(User Evaluation Model)에 의해 clustering되어지며 이는 협동적 정보 여과(Collaborative Information Filtering)를 위함이다. 관심 분야가 유사한 사용자들은 같은 그룹으로 Clustering 됨으로써 자신들이 기대하지 못했던 보다 만족스러운 정보를 추천 받을 수도 있다. 이 뿐만 아니라 특정 분야의 초보자들인 경우 동일 분야의 전문가들의 북마크 정보와 그들의 설명을 얻을 수 있다는 이유에서 Information Overload를 해결할 수 있었다(정재은 외. 2000; Jung et al. 2000).

2.2 BISAgent의 한계

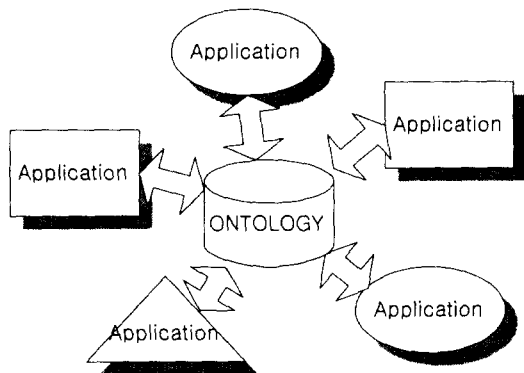
BISAgent는 사용자들로부터 북마크 정보를 입력 받을 때, 단지 사용자의 명시적인 방법에만 의존한다. 또한 사용자가 입력해야 하는 북마크의 단순한 URL 정보 뿐만 아니라 BISAgent에서 제안한 확장된 북마크 정보 구조(Extended Bookmark Information Structure)를 위한 부가적인 메타 정보(Meta-Information)들을 사용자들로부터 요구하게 된다. 이 작업 중에 사용자들은 많은 부담을 겪게 되었으며 그로 인한 Information Noise가 발생하게 되었으며 이에 따른 각각의 북마크에 대한 메타 정보의 정확성이나 북마크들 사이의 연관성 등을 신뢰할 수 없다는 것이 Information Noise 발생에 따른 문제점들이라 할 수 있다. 이와 같은 문제점들이 BISAgent 시스템의 성능을 저하시키는 치명적인 문제로 대두되었다. 이 문제는 정작 BISAgent에만 국한되지 않으며 대다수의 Information Sharing 시스템에서 겪게 되는 문제이다.

3. 문제 접근 방법

3.1 HTML 문서의 개념화를 위한 Ontology

근본적으로 Ontology란 “Explicit Specification of a Conceptualization”을 뜻한다(Vickery. 1997). 다시 말해, 현실 세계의 존재하는 지식이나 정보를 컴퓨터 시스

템 또는 Application들이 이해할 수 있는 별개의 지식 형태로 매핑(Mapping) 시킴으로써 개념화 할 수 있도록 하는 전달 함수(Transfer Function)라 할 수 있다. 이와 더불어 Ontology의 주된 목적이라 하면 어떤 시스템 구조(System Architecture), Application Domain, 또는 정보 구조(Information Architecture)에도 독립적으로 서로 커뮤니케이션이 가능하도록 하는 것이다(Ontology.Org). [그림 2]와 같이 분산 이종 호스트들의 Application들간의 정보 통합 및 공유(Information Integration & Sharing)를 위한 Content-based Agreement의 기능을 가지고 있다. 예를 들어, PSL(Process Specification Language) Ontology (Knutilla et al. 1997)나 CYCORP의 "Cyc Knowledge Server", Stanford KSL의 "Ontolingua"과 같은 프로젝트들이 정보 공유를 위한 Ontology 의 실례들이며 현재도 계속 진행중이다.



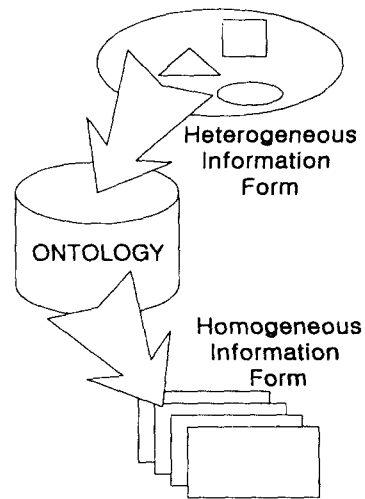
[그림 2] Ontology를 통한 정보 통합 및 공유

Ontology의 또 다른 기능으로 [그림 3]을 보면 현재 웹으로부터 얻게 되는 정보들은 매우 이질적인 형태들을 가지고 있으므로 Ontology를 통해 동질적인 정보의 형태로의 변환이 필수적이라 할 수 있다.

Ontology의 구축을 위해 사용되어지는 term들은 기본적으로 Domain Expert들의 "Brainstorming" 과정을 통해 추출되어 진다. 각 term들은 의미적 계층 구조(Semantic Hierarchy Structure)를 이루게 된다.

이와 같은 Ontology를 기반으로 BISAgent 시스템을 확장함으로써 BISAgent가 가지고 있던 한계를 극복할 수 있다. 사용자가 방문하는 웹 사이트들은 그 사이트의 Content를 분석함으로써 Clustering 하게

되며 사용자 모델을 학습하게 된다.



[그림 3] Ontology를 통한 동질의 정보 형태로의 변환

통계학적으로 사용자의 관심분야와 일치하거나 방문 빈도가 높은 사이트일 경우 자동적으로 서버에 북마크 저장 요구 메시지를 보내게 된다. 즉, 사용자에게 의한 명시적(Explicit) 북마크 저장 작업 뿐만 아니라 Ontology에 의한 묵시적(Implicit) 북마크 저장 작업이 이루어짐으로써 사용자의 부담을 줄인다. 그럼으로써 Information Noise의 발생 빈도를 줄일 수 있다.

3.2 웹으로부터의 정보를 형식화(formalization) 하기 위한 XML

현재의 HTML은 SGML(Standard Generalized Markup Language)의 한 응용이며, 이미 표준으로 정해진 DTD(Document Type Definition)에 따라 사용자는 제한된 범위내의 Instance 만을 작성할 수 있다. 따라서 현재와 같은 네트워크 상에서 복잡한 정보나 효과적인 검색, 재사용, 검증과 같은 능력에는 취약하다. W3C(World Wide Web Consortium)에서는 이러한 HTML의 취약점을 보완하기 위해 새로운 마크업 언어(Markup Language)인 XML을 1996년에 제정, 공개하였으며, 1998년 2월에는 W3C의 권고안(Recommendation)으로서 XML 1.0을 발표하였다.

XML은 웹 상에서 구조화된 문서를 전송 가능토록 한 마크업 언어로써 HTML과 SGML의 장점을

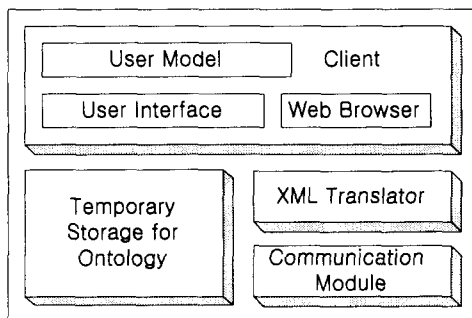
수용하고, 단점을 극복한 표준 언어이다. 특히 XML은 기능 면으로는 구조화된 문서의 작성을 용이하게 하며, 인터넷 기반의 마크업 언어이기 때문에 SGML보다 인터넷의 표준 문서로써 자리를 굳혀 가고 있다. 또한 XML은 인터넷 상에서의 문서의 내용(Content) 정의 기능과 정보 교환 및 공유(Information Exchange & Sharing)의 기능을 활용 가능하게 되었다. 이에 따라 분산된 이종 시스템들 사이의 정보 공유를 위한 Ontology의 실체를 위한 효과적인 표현 언어라고 할 수 있다.

OnBISAgent에서는 Ontology를 참조함으로써 목시적 북마크 저장 요구가 발생할 경우 해당 사이트의 HTML문서를 XML문서로 변환하는데 그 사이트의 Ontology에 의해 해석된 결과 및 메타 정보들을 첨가하게 된다. 그 다음 이 XML 문서는 서버측 북마크 저장소(Bookmark Information Repository)로 저장되어 Collaborative Filtering에 의한 북마크 정보의 추천을 위한 사용자들의 Clustering이 수행되어진다.

4. Ontology-based BISAgent; OnBISAgent

4.1 OnBISAgent 시스템 구조

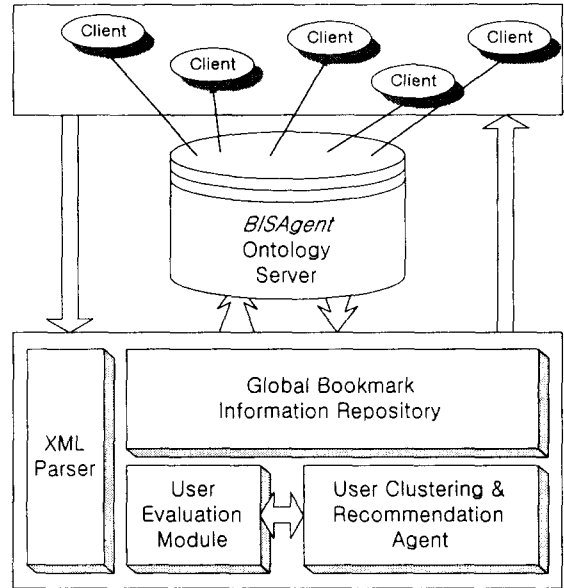
Ontology기반의 BISAgent 시스템(OnBISAgent)은 크게 클라이언트측 구조와 서버측 구조의 두 부분으로 나뉜다.



[그림 4] Client측 구조

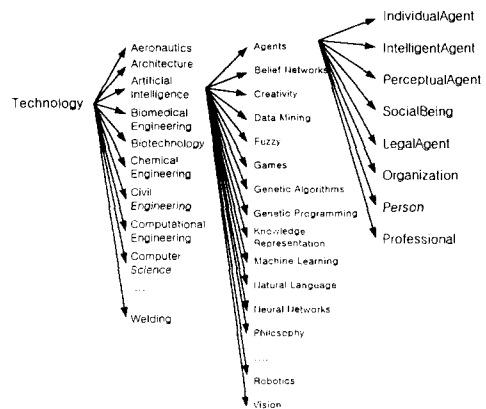
먼저 [그림 4]는 클라이언트 측의 구조를 나타내고 있으며, 사용자 인터페이스 모듈, Ontology 임시 저장소, XML 문서 번역기 모듈, 그리고 서버측과의 통신 모듈 등으로 나뉜다. 사용자 인터페이스 모듈의 사용자 모델(User Model)에 의해 사용자의 관심 분야

및 브라우징 성향(Browsing Behavior)을 학습할 수 있다.



[그림 5] 서버측 구조와 BISAgent Ontology 서버

그리고 서버 측의 구조와 BISAgent Ontology Server(BOS)은 [그림 5]와 같다. BOS는 클라이언트의 Ontology 임시 저장소에 해당 사용자의 사용자 모델에 해당하는 Ontology 정보를 제공하며 서버 측에서는 클라이언트로부터의 북마크를 Global Bookmark Information Repository(GBIR)에 저장하기 위해 참조하게 된다. GBIR에는 BOS의 참조에 의해 Clustering된 북마크들이 저장된다.



[그림 6] BISAgent Ontology 서버의 의미적 계층 구조

서버는 GBIR과 사용자의 전문성(Expert Rate; ER)과 북마크의 정확성(Accuracy Rate; AR)을 평가하

기위한 사용자 평가 모듈(User Evaluation Module; UEM)이 있다. 사용자 클러스터링 및 추천 에이전트(User Clustering & Recommendation Agent; UC&RA)는 UEM에 의한 ER 및 AR을 참조하게 된다(Jung *et al.* 2000).

다음으로 [그림 6]은 의미적 계층 구조를 이루고 있는 BOS 내부를 보여주고 있다. 각 노드(node)들은 그 노드의 정의 및 관련 키워드의 집합을 가지고 있다.

```

#IntelligentAgent
The collection of all intelligent agents. Each element of #IntelligentAgent is
an agent that is capable of having knowledge which it employs in its actions.
An intelligent agent #KnowsAbout certain things, and having #Beliefs (and
possibly #Goals) concerning those things may influence its actions. In
addition to persons, Cyc considers certain social beings, such as business
and government organizations, and intelligent machines, to be intelligent
agents. See also #Agent, #PerformedBy.
isa: #ExistingObjectType
genls: #Agent
some subsets: #SocialBeing #JudicialAgent #MedicalCareProvider
#Family-SocialEntity #LegalAgent #Organization #Court-Judicial
#MedicalCareProfessional #MedicalCareOrganization
#GeopoliticalEntity #SoleProprietorship #Partnership
#LegalCorporation #LegalGovernmentOrganization #Person
    
```

[표 1] BOS내에 저장되어 있는 "IntelligentAgent"에 관한 정보

예를 들어, [그림 6]의 엔드 노드 중의 하나인 "IntelligentAgent"에 관한 정보는 [표 1]과 같이 저장되어 있다(The Upper Cyc Ontology).

4.2 북마크 정보 공유 및 클러스터링

북마크의 공유를 위한 사용자의 북마크 저장은 [그림 7]과 같다. 해당 북마크의 HTML 문서는 개념화층(Conceptualization Layer)에서 BOS에 의해 제공되는 Ontology에 의해 Semantic 정보가 부가됨으로써 Conceptualization되며, XML 번역층(XML Translation Layer)에서는 개념화된 HTML문서를 OnBISAgent 시스템이 정의한 DTD를 참조함으로써 새로운 XML 태그를 추가하게 되며, [표 2]와 같은 형태를 취한다. 마지막으로 GBIR 삽입층(GBIR Insertion Layer)을 통해 GBIR에 삽입된다.

```

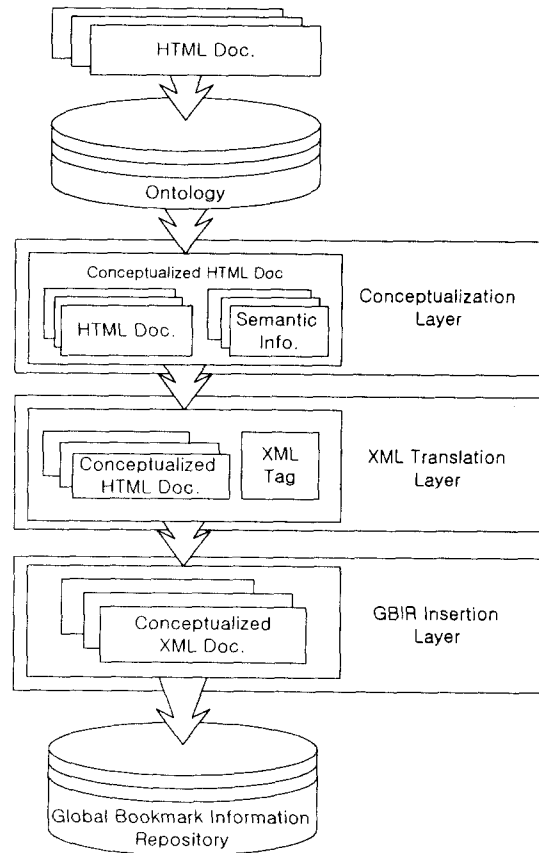
<?xml version=1.0?>
<!-- BISAgent XML -->
<!-- Developed by Jeff J. Jung -->
<Ontology-BISAgent>
  <meta-info>
    
```

```

<bisa-title> . . . </bisa-title>
<bisa-url> . . . </bisa-url>
<bisa-author> . . . </bisa-author>
<bisa-weight> . . . </bisa-weight>
<bisa-keyword> . . . </bisa-keyword>
</ meta-info >

<ht-doc>
. . . . .
</ht-doc>
</Ontology-BISAgent>
    
```

[표 2] 개념화된 XML 문서의 Scheme



[그림 7] 북마크 정보의 저장 과정

GBIR에 저장된 북마크는 User Evaluation Model 과 Conceptualized XML 문서의 Parsing과정에 의해 획득한 메타 정보에 의해 북마크 클러스터링 작업이 이루어진다.

4.3 북마크 정보 우선 순위 결정 전략

$$AR'_i = bif'_i \cdot \log \frac{M}{iaf'_i}$$

uy_j : The number of users that include particular bookmark information in a same area
 wy_j : The number of areas that include bookmark information

M : The number of total area registered in server

$$Global_k ER_j^i = N_k^i \cdot \log \frac{S}{P_k}$$

N_k^i : The number of bookmark information that i -th user insert into k -th area

S : The number of all users

P_k : The number of users that insert bookmark information into k -th area

$$private_k ER_j^i = \frac{1}{\log(A^i + 1)} \cdot \frac{N_k^i}{Total^i}$$

A^i : The number of all area that is stored by the i -th user

N_k^i : The number of bookmark information that i -th user insert into k -th area

$Total^i$: The number of all bookmark information that is stored by i -th user

$$R_j^i = \alpha \cdot AR_j^i + \beta \cdot w_j^i \cdot ER_j^i$$

$$ER_j^i = Global ER_j^i + \gamma \cdot Personal ER_j^i$$

w_j : Weight value assigned to the j -th bookmark information of the i -th user at the moment of insertion

α, β, γ : Experimental constant

k : The area of the j -th bookmark information

[표 3] BISAgent의 Bookmark Ranking Strategy

OnBISAgent에서는 여러 사용자에게 의해 공유된 북마크를 [표 3]와 같은 정보 우선 순위 결정 전략에 의해 해당 사용자에게 검색 및 추천한다. 일반적인 정보 검색 시스템(Information Retrieval System)에서의 TF-IDF (Term Frequency - Inverted Document Frequency) Scheme을 변형하여 정의하였다. 하단부의 $R_j^i = \alpha \cdot AR_j^i + \beta \cdot w_j^i \cdot ER_j^i$ 값이 높은 북마크일수록 사용자의 만족도가 높을 것으로 예상된다(Jung *et al.* 2000).

4.4 북마크 저장 자동화를 위한 사용자의 사이트 방문 패턴 분석을 통한 예측

사용자 북마크 저장의 자동화를 위한 예측은 사용자의 사이트 방문 패턴의 해석에 의해 이루어진다.

시간 (t)	방문 사이트	index #
00/04/06 01:15	http://www.ontology.org	1
00/04/06 01:19	http://eslab.inha.ac.kr	2
...
00/04/07 00:23	http://www.ontology.org	1
...
00/04/08 01:02	http://www.ontology.org	1

[표 4] 방문 사이트의 시간별 리스트

일반적으로 단위 시간당 방문 횟수(Site-Visit

Number per Unit Time)보다는 사이트의 방문 주기(Site-Visit Period)의 영향이 더 크므로 [표 4]과 같이 시간의 흐름에 따라 사용자가 방문하는 사이트의 리스트를 저장하고 있다가 반복되는 사이트의 주기를 조사한다. 다시 말해 우연한 계기로 인해 일정 기간 동안 방문이 빈번한 사이트보다 상당한 시간동안 주기적으로 방문한 사이트가 북마크로 저장된다는 의미이다.

$$\text{time_of_visit} = \{t_1, t_2, \dots, t_n\} \quad (\text{Equ-1})$$

$$\text{interval_of_visit} = \{\Delta t_1, \Delta t_2, \dots, \Delta t_{n-1}\} \quad (\text{Equ-2})$$

(단, $\Delta t_i = t_{i+1} - t_i$)

$$\text{mean}(\text{interval_of_visit}) \quad (\text{Equ-3})$$

$$= \mu_{\text{interval_of_visit}} = \frac{\sum_{k=1}^{n-1} \Delta t_k}{n-1}$$

$$\text{Variance}(\text{interval_of_visit}) \quad (\text{Equ-4})$$

$$= \sigma_{\text{interval_of_visit}}^2 = \frac{\sum_{k=1}^{n-1} (\Delta t_k - \mu_{\text{interval_of_visit}})^2}{n-1}$$

if $\sigma_{\text{interval_of_visit}}^2 \leq \delta$, this bookmark information is asserted. otherwise, wait additional visit of users.

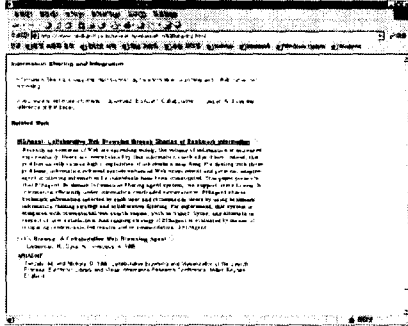
(단, δ 는 the threshold value for bookmark assertion이다.)

사이트 방문 주기는 시간 도메인(time-domain)의 이산 데이터(discrete data)들을 주파수 도메인(frequency-domain)으로 변환하는 푸리에변환(Fourier Transformation)과 유사한 작업으로써 시간별 사이트 리스트 정보를 사이트별 시간 리스트로 변환한다. 다시 말해 시간-도메인 이산 데이터가 사이트-도메인 이산 데이터로 변환된다는 의미이다. 위의 (Equ-4)에 의해 사이트 방문의 분산(variance)을 구할 수 있으며 이 값이 일정값(δ)보다 작을 경우 사이트 방문의 편차가 충분히 작아 일정한 주기를 가지고 방문이 이루어짐을 예측할 수 있으며 북마크 저장이 이루어진다.

5. 실험

5.1 HTML 문서의 개념화된 XML 문서로의 변환

사용자가 [그림 8]의 사이트를 북마크에 저장한다고 가정하자. 해당 사이트의 HTML문서는 [표 5]와 같고 ontology와 매칭된 키워드들은 굵은 글씨체로 나타내었다.



[그림 8] 실험을 위한 웹 사이트

```

<html>
<head>
<title> Information Sharing and Integration
</title>
</head>
<body>
...
<td width="65%"><font size="2"> Information
Sharing is very important technology for
information searching and collaborative web
browsing. </font>
<p>If you want to get more information,
download BISAgent: Collaborative ... paper.
And see the reference of that paper. </p>
</td>
...
<p><font size="2">Twidale, M. and Nichols, D.
1996. Collaborative Browsing and
Visualization of the Search Process,
Electronic Library and Visual Information
Research Conference, Milton Keynes, England.
</font></p>
</body>
</html>

```

[표 5] 키워드가 하이라이트된 HTML 문서

[표 5]의 HTML문서는 OnBISAgent 북마크 저장 과정을 거쳐 [표 6]과 같은 개념화 된 XML문서로 변환된다.

```

<?xml version=1.0?>
<!-- BISAgent XML -->
<!-- Developed by Jeff J. Jung -->
<Ontology-BISAgent>
<meta-info>
<bisa-title> information sharing and
integration </bisa-title>
<bisa-url>
http://www.intelligent.pe.kr/research/
information%20sharing.html</bisa-url>

```

```

<bisa-author> Jae-Eun Jung </bisa-author>
<bisa-weight> 5 </bisa-weight>
<bisa-keyword> information </bisa-keyword>
<bisa-keyword> intelligent </bisa-keyword>
<bisa-keyword> collaboration </bisa-
keyword>
<bisa-keyword> intelligent </bisa-keyword>
</meta-info >

<ht-doc>
<head>
<title> Information Sharing and
Integration </title>
</head>
<body>
...
<td width="65%"><font
size="2">Information Sharing is very
important technology for information
searching and collaborative web
browsing.</font>
<p>If you want to get more information,
download BISAgent: Collaborative ...
paper. And see the reference of that
paper. </p>
</td>
...
<p><font size="2">Twidale, M. and
Nichols, D. 1996. Collaborative Browsing
and Visualization of the Search Process,
Electronic Library and Visual Information
Research Conference, Milton Keynes,
England.</font></p>

</body>
</ht-doc>
</Ontology-BISAgent>

```

[표 6] GBIR에 저장되는 개념화된 XML 문서

5.2 사용자의 사이트 방문 패턴 분석에 의한 북마크 정보의 추출

1st week	1	2	1	1	2	1	1	1	1	1	1	1	1	1
2nd week	1	1	1	1	1	1	1	1	1	1	1	1	1	1
3rd week	1	1	1	1	1	1	1	1	1	1	1	1	1	1
4th week	1	1	1	1	1	1	1	1	1	1	1	1	1	1

1	www.chosun.com	16	www.x3.org
2	www.ontology.org	57	www.nortelnetworks.com
...	...	49	www.inha.ac.kr	58	www.comp.lancs.ac.uk

[그림 9] 사용자 A가 방문한 사이트 리스트

사용자가 방문하는 사이트의 통계적 해석을 통한 북마크 정보 추출을 위한 실험으로써 10명의 사용자가 만 20일 동안 방문하는 사이트의 리스트를 조사하였다. 이후 추출된 북마크의 정확성의 검증을 하기 위해 약 10일 동안 사이트 방문 패턴을 추가로 조사하였다.

[그림 9]는 사용자 A의 방문 사이트 리스트와 각 사이트의 방문 시간을 시간표에 표시한 것이다.

본 실험의 평가는 20일 동안 i 번째 사용자의 자동 저장된 북마크의 집합(D_i^{auto})과 이후 10일 동안 i 번째 사용자가 방문하는 사이트의 집합(S_i)의 비교를 통해 Coverage 과 Accuracy, 즉 정보검색에서의 각각 Recall과 Precision에 해당하는 값을 구하였다. 결과는 다음의 [표 7]과 같다.

User (i)	D_i^{auto}	S_i	Recall (Coverage)	Precision (Accuracy)
A	5	34	0.24	0.86
B	8	25	0.34	0.85
C	3	21	0.38	0.88
D	4	23	0.29	0.76
E	3	42	0.16	0.84
F	5	34	0.48	0.92
G	13	26	0.30	0.78
H	4	61	0.26	0.85
I	7	53	0.43	0.74
J	7	39	0.37	0.82
Average			0.325	0.83

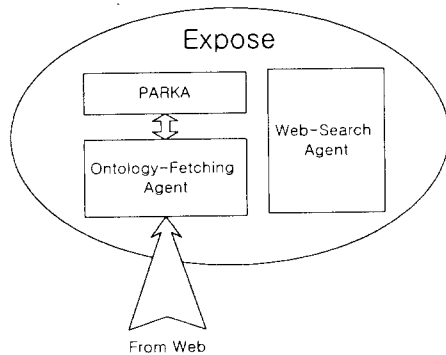
[표 7] 북마크 정보 추출의 정확성을 평가한 표

북마크의 자동 추출 기간 이후 사용자들이 방문하는 사이트들을 미리 예측하는 능력은 다소 떨어졌으나 추출된 북마크의 정확성 면에서는 상당히 높은 값을 얻음으로써 비교적 적은 수의 북마크라 하더라도 높은 신뢰성이 있다는 것을 알 수 있다.

6. 관련 연구

6.1 Ontology-based Web Agents, Expose

이 시스템은 웹에서의 정보 검색을 위한 Ontology 기반의 에이전트 시스템이다.



[그림 10] System Architecture of Expose

Ontology의 표현을 위해서는 SHOE를 이용하

였으며 [그림 10]와 같이 PARKA라고 하는 Ontology의 저장소가 있으며 두개의 독립적인 에이전트로 나누어지는데 하나는 웹으로부터 ontology를 수집해오는 에이전트와 웹에서 정보를 탐색하는 에이전트이다(Luke et al. 1997).

6.2 OML (Ontology Markup Language) & XOL (XML-based Ontology-exchange Language)

OML의 목적은 application component들이 서로 상호 운용(interoperable)가능토록 하는 것이다. 초기 버전은 기본적으로 SHOE의 XML로의 변환이라고 할 수 있으나, 최근 XML에 여러 기능이 더해지면서 OML 역시 새로운 기능을 제공하고 있다(OML). 이와 더불어 XOL은 XML 기반의 Ontology 교환 언어이며, 초기에 생체 정보 교환을 위해 설계되었으나 지금은 어떤 분야에서도 사용될 수 있다(XOL).

7. 결론 및 향후 연구

지금 인터넷 환경에서는 information overload를 극복하기 사용자의 많은 희생이 있어야 한다. 본 연구에서는 XML을 이용한 OnBISAgent(Ontology-based Bookmark Information Sharing Agent) 시스템을 제안하였다. OnBISAgent 시스템을 이용함으로써 사용자들의 효과적인 북마크 저장을 지원하였으며 높은 북마크 자동 저장의 예측률을 보임을 증명하였다. BISAgent이 근본적으로 목적인 저장된 북마크들을 공유함으로써 information overload 극복을 위한 협동적 웹 브라우징을 가능케 하는 것이다.

향후 연구 과제로는 보다 일반화된 Ontology의 구축이며 사용자 사이트 방문 패턴 인식력의 향상을 통한 빠르고 정확한 북마크 정보의 추출이라 할 수 있다.

References

- 정재은, 윤정섭, 조근식, 2000, "북마크 정보 공유를 통한 협동적 웹 브라우징," 제27회 한국정보과학회 춘계학술대회, 2000. 4.
- 이경하, 이강찬, 이규철, 2000, "XML 프로그래밍," 한국정보과학회, 제18권, 제4호, pp.4-12, 2000. 4.
- Cooper, C., "Benchmarking XML Parsers,"

- <http://www.xml.com/pub/Benchmark/article.html>
- Jung, J., Yoon, J., Jo, G. 2000, "BISAgent: Collaborative Web Browsing through Sharing of Bookmark Information," *IIP2000, IFIP*, Beijing, China, to appear.
- Knutilla, A., et al. 1997, "Process Specification Language: Analysis of Existing Representations," *NISTIR 6133, National Institute of Standards and Technology*, Gaithersburg MD, 1997. 11.
- Lenat, D., "CYC: A Large-Scale Investment in Knowledge Infrastructure," *Communication of ACM*, 38(11), pp. 33-38, 1995.
- Luke, S., Spector, L., Rager, D., Hendler, J. 1997, "Ontology-based Web Agents", *Autonomous Agent '97*, ACM, 1997.
- Vickery, B.C. 1997, "Ontologies," *Journal of Information Science*, 23 (4), pp.277-286, 1997. 1.
- Cyc Knowledge Server, *CYCOPR*, <http://www.cyc.com>
- KIF (Knowledge Interchange Format), <http://logic.stanford.edu/kif/>
- Ontolingua, *Stanford Univ. Knowledge Systems Lab*, <http://www.ksl.stanford.edu/software/ontolingua/>
- OML (Ontology Markup Language), *Oasis-Open.Org*, <http://www.oasis-open.org/cover/oml9808.html>
- Ontology.Org, *Enabling Virtual Business*, <http://www.ontology.org/main/papers/faq.html>
- The Upper Cyc Ontology, <http://www.cyc.com/cyc-2-1/index.html>
- W3C (World Wide Web Consortium), <http://www.w3.org>
- XOL (XML-based Ontology exchange Language), <http://www.ai.sri.com/pkarp/xol/>