

사용자의 인지부담 절감을 위한 인지 기반 지식 구조 및 정보 탐색 프레임워크*

박 호 건[†] 맹 성 현 김 경 민 장 관 최 종 욱

한국정보통신대학교 공학부

웹과 디지털 라이브러리가 대중화되면서, 사용자들은 이를 정보 조사나 학습 같은 높은 인지부담을 수반하는 작업에 활용하기 시작한다. 그러나 대부분의 검색엔진 및 브라우저는 사용자에게 단순한 질의 기반 검색 및 항해탐색 기능만을 제공하므로 정보를 해석하고 연계하며 통합하는 등의 역할은 사용자에게 전가한다. 본 논문에서는 인간-웹 상호작용을 위한 이단 모델(Two-level Model)이라고 부르는 개인화된 지식구조를 제안한다. 이 구조는 지식공간과 정보공간으로 구성되어 있으며, 새롭게 지식공간을 구성하고 이를 이용하는 명령(Operation)을 제공한다. 지식 공간은 웹에서 상호작용을 통해 검색된 결과에 대한 개념적인 해석이라고 할 수 있으며, 토픽 및 토픽간의 관계를 통해서 이 지식 공간이 구성된다. 이러한 지식 공간은 정보공간의 정보객체와 연결되어 있어 지식공간으로부터 원하는 정보에 대한 접근이 가능하다. 이 모델에 기반 하여 구현된 시스템을 통해 정보탐색 태스크 중 가장 높은 인지부담을 요구하는 탐색적 검색 환경 아래의 실험을 수행하여 사용자 인지부담 절감 효과를 확인 할 수 있었으며, 향후 개인의 효과적인 정보 탐색을 위한 가치 있는 메타데이터로서의 활용이 기대된다.

주제어 : 정보탐색, 인지 기반 지식 구조

* 본 연구는 지식경제부 및 정보통신연구진흥원의 IT핵심기술개발사업의 일환으로 수행 [2008-F-047-01, Urban Computing Middleware 기술 개발] 하였으며, 과학기술부(K2071100007-07A0100-00710) 및 한국과학기술협력재단의 지원과 2008년도 2단계 BK 21 사업의 부분 지원으로 수행되었습니다.

† 교신저자: 박호건, 한국정보통신대학교 공학부, 연구세부분야: 정보검색
Email: gsgphg@icu.ac.kr

서론

웹과 디지털 라이브러리¹⁾가 널리 보급되면서, 사용자들은 이를 단편적인 정보를 찾는 것만이 아닌 학습이나 조사 같은 일정량 이상의 인지 부담을 요하는 정보 탐색 작업에 활용하기 시작하였다. 그러나 저마다 다른 배경지식과 정보 요구를 가진 사용자에게 효과적인 정보 탐색을 지원한 다는 것은 매우 어려운 일이기 때문에, 대부분의 검색엔진 및 브라우저는 정보를 해석하고 연계하는 등의 역할을 사용자들에게 전가하고 있는 실정이다. 이러한 문제를 해결하고자, 연구자들은 사용자들의 인지 분석을 통해 이들의 탐색 활동을 보조하기 시작하였다.

가장 많이 활용되어 온 것은 사용자의 인지기반 행동 양식 정보이다. 이 정보는 Click-through 데이터나 기타 다른 피드백을 통해 추출할 수 있으며, 이는 사용자 프로파일 생성에 반영되고 도메인 온톨로지와 함께 활용됨으로써 개인화된 검색 [1] [2] 이나 추천 시스템 [3] 에 널리 활용되어 왔다. 인지 기반 행동 양식과 함께 인지 기반 지식 구조(Cognitive Knowledge Structure) 또한 정보 탐색 지원에 중요한 역할을 해왔다. 인지 기반 지식 구조란 정보를 접한 사용자들이 머릿속에 가지게 되는 개념적 해석을 의미한다. 특히, 도메인 온톨로지나 ODP²⁾ 같은 일반적인 지식 구조들이 질의 확장 [4] 이나 문서 분류 [5]에 활용되기 시작함으로써 더 효율적인 정보 검색 시스템을 구현할 수 있게 되었다. 그러나 인지 기반 행동 양식에 비해 인지 기반 지식 구조에 대한 연구는 사용자 개개인의 대한 메타데이터나 유저 모델을 추출하고 사용하기보다 일반적인 개념체계를 활용하는 수준에 머물러 왔다. 본 연구에서는 개개인이 정보를 습득함으로써 가지게 되는 인지기반 지식 구조를 추출함으로써, 이를 통해 사용자의 정보 탐색을 지원하려고 한다. 이 개인화된 지식 구조는 향후 사람들 간에 공유될 수 있으며, 검색 및 브라우징 성능 향상을 가능하게 하는 메타데이터로서의 잠재력을 가지고 있다.

구체적으로, 우리는 이단 모델(Two-level Model)이라고 불리는 개인화된 인지기반 지식구조와 이에 기반 한 웹-인간 상호작용을 촉진하는 프레임워크를 제안한다.

1) 비록 디지털 라이브러리의 일반적인 정의는 모든 데이터가 시스템의 관리 하에 있어야 한다는 전제를 가지지만, 본 논문에서는 웹을 디지털 라이브러리에 포함하고 기술한다.

2) Open Directory Project (<http://www.dmoz.org/>)

이 모델은 이름에서 제시하는 바와 같이 두 단계로 구성된다. 첫 번째는 사용자의 개념적 이해를 표현하는 지식 공간(Knowledge Space)이고 두 번째는 웹 페이지 같은 정보 객체를 표현하는 정보 공간(Information Space)으로 정의한다. 여기서 지식공간은 정보공간 위에 위치하며, 정보공간에서 웹 페이지 혹은 디지털 객체를 읽으면서 사용자의 인지 공간 속에 생성되는 개념들이 토픽(Topic) 과 그들 간의 관계(Association)로 표현되도록 하는 장소로 사용된다. 토픽은 하나 또는 그 이상의 정보 객체와 연결되어 있으며 이를 자원(Occurrence)라고 한다. 여기서 사용하는 토픽, 관계, 자원 같은 용어들은 토픽맵 프레임워크(ISO/IEC 13250)로부터 빌려왔으며, 본 논문에서 제안하는 방법은 이 프레임워크에 기반을 둔 새로운 개념의 정보 탐색 모델 이라고 할 수 있다.

제안하는 모델이 가장 유용하게 사용될 수 있는 영역은 앞서도 언급된 인지 부담의 경감에 있다. 특히 과거에 찾았던 정보의 대한 재사용성을 높여주는 장점을 가지게 되는데, 검색 히스토리를 단순히 활용 [1] 하는 과거 연구와는 달리 지식 공간의 구성과 그에 대한 활용에 초점을 두고 있다는 것이 본 연구의 주안점이다. 지식공간 및 지식공간과 정보공간 사이에 존재하는 명령(Operation)들은 과거에 접했던 정보들 간의 관계를 형성하고 이에 접근 가능하게 하여 사용자가 정보탐색 과정에서 가질 수 있는 인지적 부담을 경감시킬 수 있다.

본 연구에서는 이단 모델의 검증을 위해 새로운 상호작용 프레임워크와 도구를 구현하였다. 이 도구는 기존의 검색엔진과 함께 접합되어 구현되었으며 사용자가 일반적인 검색과 브라우징을 할 수 있도록 할 뿐만 아니라, 과거에 검색한 정보를 기반으로 새롭게 지식공간을 구성하는 시각화된 인터페이스를 제공한다.

실험을 위해서 탐험적 검색(Exploratory Search) [6] 을 주요 태스크로 선택하였다. 이는 정보 검색에서 가장 많은 인지부담을 요구하는 태스크 중의 하나이며, 이를 위해 사용자는 정보를 모으고(Scanning/Viewing) 분석하고(Analyzing), 의사결정(Qualitative/Quantitative Judgment)을 하는 일련의 과정을 거치게 된다. 이 과정은 긴 검색 및 브라우징 활동을 수반 [7] 하며, 정보를 읽고 습득해서 문서로 정리하거나 북마킹 툴을 사용하는 등의 방법이 적용되어왔다. 특히 이전에 접한 정보에 다시 접근하는 경우에 다른 간단한 검색 태스크 이상의 인지부담이 주어지기 때문에, 기존의 검색엔진이나 브라우저로는 이런 종류의 작업을 지원하기에는 적합하지 않

다. 본 논문에서는 이 태스크를 실험자에게 환경으로 제공하고 프로토타입을 사용하게 함으로써 얼마나 큰 인지부담 경감이 생기는지 실험하였다.

관련 연구

정보 검색 시스템에서 사용자의 인지부담을 덜어주기 위하여 개인화된 지식 구조를 활용하려는 시도는 과거에도 있었다. 1970년대, 최초로 BRS (Bibliographic Retrieval Services)와 Orbit 같은 상용 검색 엔진에서 지난 질의를 저장하고 향후 다시 사용할 수 있는 기능을 제공하였고, 1980년대에 이르러 I3R과 MARIAN (Multiple Access and Retrieval of Information with ANnotations)에서 사용자들이 자신의 과거 질의를 이용하여 습득한 지식을 표현하는 것을 가능하게 하고, 그를 위한 도구를 제공하였다. 특히, MARIAN에서는 처음으로 검색결과에 기반한 문맥기반 패시 (Conceptual Facet)을 활용하여 습득한 지식을 표현할 수 있도록 해 주었다. 그러나 위에서 언급한 기존의 시스템들은 개인의 인지기반 지식 구조의 추출은 시도하지 않았다. 즉, 그들의 연구는 과거 질의와 브라우징 했던 페이지만 활용하였기 때문에, 온톨로지와의 연결고리를 찾기가 힘들며, 향후 사용을 위한 메타데이터로서는 실용적이지 못하였다.

최근 들어서 Belhassen 의 연구팀 [8] 은 사용자의 정보 요구를 표현하고 관리하기 위한 사용자 모델을 제안하였다. 이들은 검색엔진을 통한 사용자와의 상호 작용을 통해 사용자 모델을 만들 수 있다고 보고 사용자의 정보 탐색 전략과 목표를 반영한 프레임워크를 개념적으로 제안하였다. 그리고 그 뒤를 이어 Tian 연구팀 [9] 이 문맥 기반 정보검색 시스템을 위해 사용자의 인지기반 지식 구조를 모델링하고, 그것을 도메인 온톨로지 위에서 확립하려는 시도를 하게 된다. 그 방법으로 Spreading Activation 기법을 도메인 온톨로지에 사용하여 개인의 인지 기반 지식 구조를 만들려고 하였다. 하지만, 이 연구는 도메인 온톨로지에 의존적이었기 때문에 활용도가 많이 떨어졌었다. 이에 반해, 본 논문에서 제안하는 지식 구조 모델은 사용자와 시스템이 협업을 통해 개인화된 지식 구조를 생성하며, 이를 정보 탐색에 활용한다.

본 논문의 3절에서 이단 모델(Two-Level)에 대해서 소개하고, 4절에서는 이에 기반 한 프레임워크를 제안하며, 마지막으로 5절에서는 구현한 프로토타입 시스템에 대한 실험 및 평가를 기술한다.

이단 모델(Two-level Model)

그림 1은 정보 공간(Knowledge Space)과 지식 공간(Information Space) 사이의 관계를 개념적으로 보여준다. 정보 공간은 웹 페이지나, 문서, 이미지 같은 정보 객체의 집합으로 구성되어 있으며, 이는 웹 페이지의 하이퍼링크나 디렉터리 정보도 포함한다. 기존에 행해왔던 검색이나 브라우징이 이 공간 안에서의 가능한 행동들이다. 반면에, 지식 공간은 정보공간의 상위에 존재하며 링크를 통해 정보 공간의 객체들과의 관계를 형성한다. 지식 공간은 사용자가 해석한 정보의 의미를 반영하여 적절한 토픽과 관계들로 그 개념 체계를 표현한다. 또한 토픽 타입, 관계의 타입을 정의할 수 있으므로 이들과 온톨로지를 위한 명령 역시 지식 공간에 정의된

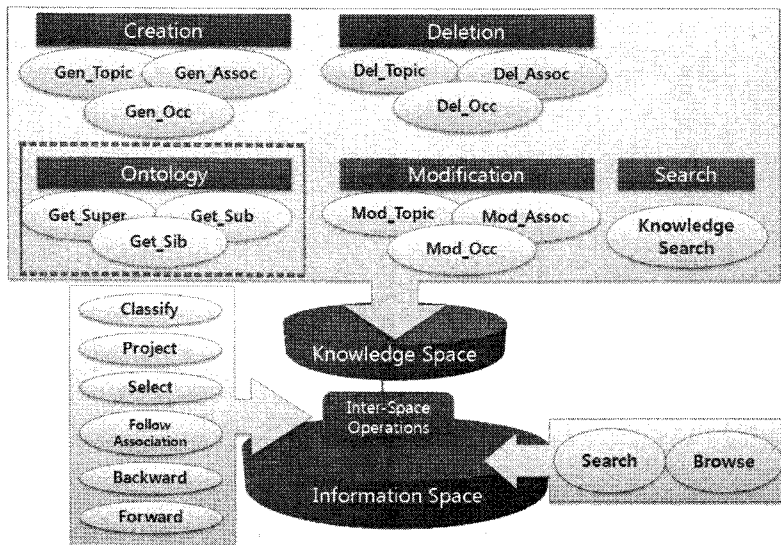


그림 1. 이단 모델의 개요

다. 이 공간은 시스템의 추천과 사용자와 참여를 통해 구축, 유지, 보수된다. 정보 공간과 지식 공간 사이에서는 정보객체가 토픽의 자원(Occurrence)로 연결되어 있어, 공간간 명령(Inter-Space Operation)들을 통해 서로간의 자연스러운 연결을 지원한다. 그림 1에서 보여주고 있는 여러 명령(예: Classify, Project 등)들은 향후 추가적으로 확장될 수 있으며, 본 논문에서는 필수적인 기능들만 소개한다.

이론적 기반

우리가 제안하는 이단 모델은 디지털 라이브러리의 5S Theory [10] 와 조화롭게 사용할 수 있다. 5S Theory는 디지털 라이브러리를 구성하는 기본 개념들을 설명하기 위해 제안된 이론 체계로, 5S는 Streams, Structures, Spaces, Scenarios, Societies를 의미한다. Streams은 문자나 이미지, 비디오 같은 콘텐츠들의 연속을 의미한다. 이것들은 동적인 요소들의 연속 그리고 어떤 동적인 정보의 흐름을 표현할 수 있다. Structures는 정돈된 정보를 표현하는 방법이다. 이는 Directed Graph와 레이블 집합의 튜플, 그리고 레이블 함수를 포함한다. Scenario는 함수에서 어떤 필수적인 요소를 성취하기 위해 연산의 상태를 수정하는 행동이나 이벤트를 정의한다. Societies는 디지털 라이브러리에서 가장 상위 요소이며, 나머지 요소들의 집합과 그것들의 관계들로 구성된다.

본 논문에서는 이 개념과 정의에 추가로 이단 모델을 위한 개념과 명령들을 기존의 개념들을 이용하여 정형화하려고 한다. 이단 모델에서 정보공간은 이미 정의되어 있는 개념으로, [10]에서처럼, digital object, do 의 집합으로 구성된 collection, C 의 집합, 구조적 메타데이터의 상세(specification) ([10]의 Def. 11 참조)의 집합인 streams의 집합, 그리고 structuredStream ([10]의 Def. 15 참조) 으로 표현할 수 있다. 이에 추가로, 지식 공간의 개념은 아래의 Knowledge map을 통해 정의될 수 있다.

- Definition: $KM = (KS, Contents, P)$ 을 Knowledge map이라고 한다.

여기서, $KS = \{(T_{KS}, A_{KS}), (L_{KS}, F_{KS})\}$ 는 Knowledge structure를 의미하며, T_{KS} 와 A_{KS} 는 토픽과 토픽 사이 관계(Association)를 의미한다. 각 A_{KS} 또는 T_{KS} 는 레이블, L_{KS}

을 가지고 각 구조에 알맞은 레이블을 찾아주는 함수, F_{KS} 도 존재한다. *Contents*는 정보 공간 속 디지털 오브젝트(digital object), *do*, 모든 (sub)streams, 그리고 구조와 노드와 Stream의 segment들을 연계해주는 가능한 모든 structuredStream을 포함한다.

P 는 KS 와 *Contents*를 연결하는 함수이다. 즉, 이 함수는 지식 공간과 정보 공간을 연결한다. 이는 앞에서 언급한 자원(Occurrence)을 의미한다.

A. 지식 공간 구성을 위한 명령

- Creation (Gen_Topic, Gen_Assoc, Gen_Occ): 지식 공간은 Knowledge map, $KM = (KS, Contents, P)$ 을 만드는 것이다. 이 명령은 KM 의 핵심부분으로 토픽, 관계, 자원 (Occurrence)을 생성하여 주며, 사용자의 마음속에서 개념화할 수 있는 모든 것들이 토픽과 관계가 될 수 있다. 자원은 토픽에서 *Contents*로 연결하는 함수 P 로 정의한다.

- Deletion & Modification: Repository 제어 함수 (Def. 19 in [10]) 가 이 명령을 관리하며, collection을 제어, 관리하는 get, store, del, modification같은 함수를 가진다.

- Knowledge Search: $OP_{knowledge_search}: Q \times K \rightarrow 2^K$, 여기서 Q 는 질의 (Def. 21 in [10] 참조)이며, K 는 Knowledge collection이다. 이 함수는 K 의 부분 집합을 질의나 Knowledge object, ko 와 관계 짓는다. 즉, 이 명령은 질의를 받고 적합한 지식 객체를 찾아 주는 역할을 한다.

• Ontology-Related Operations

$OP_{get_super-type}: T_K \times O \rightarrow O$, 여기서 T_K 는 토픽이고, O 는 온톨로지 객체이다. 이 명령은 토픽의 상위 개념을 찾기 위해 온톨로지 객체와 토픽을 관계 지어준다.

$OP_{get-sibling}: T_K \times O \rightarrow O$, 여기서 T_K 는 토픽이며 O 는 온톨로지 객체이다. 이 명령은 토픽의 같은 상위노드를 가진 개념(Sibling)을 찾기 위해 온톨로지 객체와 토픽을 관계 지어준다.

$OP_{get-sub-type}: T_K \times O \rightarrow O$, 여기서 T_K 는 토픽이며 O 는 온톨로지 객체이다. 이 명령은 토픽의 하위 개념을 찾기 위해 온톨로지 객체와 토픽을 관계 지어준다.

B. 공간 간(Inter-Space)명령

- Project : $OP_{project}: T_{KS} \times P \rightarrow 2^{Contents}$, 여기서 T_{KS} 는 토픽이다. 이 명령은 $Contents$ 를 토픽과 관계 지어주며, 사용자로부터 토픽을 받아서 표시된 포인터, P 를 사용해서 해당되는 $Contents$ 를 찾아준다. Project 명령으로 사용자는 특정 다른 조건에 만족하는 그 정보 객체의 타입을 기술할 수 있다. 포맷이나 장르 같은 타입 정보가 여기서 자원(Occurrence) 타입으로서 사용 가능하다.

- Select : $OP_{select}: 2^{KS} \times P \rightarrow 2^{Contents}$, 여기서 2^{KS} 는 지식구조의 부분집합이다. 이 명령은 토픽과 $Contents$ 를 관계 지어 주며, 사용자가 정의한 토픽 표현은 입력 받은 토픽 표현에 해당되는 토픽으로부터 발생하는 포인터, P 를 사용해서 상응하는 $Contents$ 를 찾아 준다. Select는 불린(Boolean)표현이나 수동으로 선택한 조건을 받아서 선택 받은 토픽에 연결된 정보 객체를 찾아주며, 이는 의미 기반 검색(Semantic search)의 효과를 지닌다.

- Classify : $OP_{classify}: T_{KS} \times 2^{Contents} \rightarrow P$, 여기서 T_{KS} 는 토픽으로 그 토픽의 토픽 타입을 분류 스키마로 사용한다. 이 명령은 토픽과 $Contents$, 또는 디지털 객체, do 의 집합을 관계 지어 자원(Occurrence)을 효과적으로 생성하는 연결함수 P 를 돌려준다.

- Follow_Association : $OP_{follow_association}: T_{KS} \times A_{KS} \rightarrow 2^{Contents}$, 여기서 T_{KS} 는 토픽의 집합이고 A_{KS} 은 관계이다. 이 명령은 $Contents$ 의 부분집합을 돌려준다. T_{KS} 의 토픽과 A_{KS} 으로 연결된 토픽들의 자원(Occurrence)의 $Contents$ 의 부분집합을 돌려준다.

- Forward : \mathcal{J} 를 KS 들의 연속이라고 했을 때, 검색에 따르는 Knowledge-level 명령들 후에 $\{KS_0, KS_1, \dots, KS_n\}$ 가 만들어지고 저장된다.

$OP_{forward}: succ(KS_i)$, 여기서 $succ(.)$ 는 \mathcal{J} 안에서 KS_i 뒤의 KS_{i+1} 를 반환하는 함수이다. $succ(KS_n)$ 은 정의되지 않는다.

- Backward : $OP_{backward}: pred(KS_i)$, 여기서 $pred(.)$ 는 \mathcal{J} 안에서 KS_i 앞의 KS_{i-1} 를 반환하는 함수이다. $pred(KS_n)$ 은 정의되지 않는다.

예제

다음 두 가지 시나리오는 이단 모델을 적용한 도구의 다양한 명령들이 정보탐색 태스크에 어떻게 적용될 수 있는지 소개한다.

- **시나리오 1:** 프린터에 관심을 가지고 있는 사용자는 먼저 레이저 프린터에 관한 페이지를 찾기 위해 “레이저 잉크 프린터”를 질의로 입력한다. (Search) 그 후, 사용자는 탐색 중에 레이저 잉크 프린터가 레이저 프린터와 다르다는 것을 인지하게 되고 (Browse), 질의를 수정한다. (Search) 사용자는 이 새로운 검색 후에 반환된 결과들의 처음 몇 개의 페이지를 읽고 (Browse), 토픽 “레이저 프린터”(Gen_Topic)와 관련된 페이지를 주목하게 되고 그것들을 저장한다. (Gen_Occ)

레이저 프린터에 대한 배경지식을 얻게 된 사용자는 구체적인 정보를 얻기 위해 (Browse) 상품명을 클릭하고 특정 프린터 모델의 사진을 볼 수 있는 링크를 따라간다. (Browse) 몇 분의 추가적인 탐색 후, 사용자가 저장한 페이지로부터 링크된 사진을 한꺼번에 보기 위해, 제안한 도구를 사용하여 사용자는 JPEG이라는 자원 (Occurrence) 타입을 가진 파일을 찾아 주는 Project 명령을 실행한다. (Project) 사진을 본 사용자는 X라는 회사의 제품을 마음에 두게 되고, 위의 활동과 유사한 방법으로 관련된 페이지를 찾아, “X”라는 토픽 아래에 관련된 페이지를 연결한다. X에 대한 추가적인 정보를 모은 후, 사용자는 X가 생산한 레이저프린터에 관련된 페이지를 보기 위해 Select 명령 (“X”와 “레이저 프린터” 두 토픽의 인터섹션) 을 실행하고 (Select), “X가 생산한 레이저 프린터”라는 토픽으로 그 결과를 저장한다. (Get_Sub & Gec_Occ) 마지막으로 사용자는 “manufactured by”라는 두 토픽 사이의 관계를 생성하고 탐색을 마무리한다. (Gen_Assoc)

이 시나리오는 기존에는 고립되어 있던 검색과 브라우징 행동이 제안된 프레임워크 위에서 어떻게 자연스럽게 연결되는지 보여준다. 특히, 지식 단계와 공간 간 (Inter-space) 명령들은 체계적인 방법으로 정보 단계 명령들 사이의 간극을 연결하여 준다.

- **시나리오 2:** 한 주 후에, 사용자는 조사한 상품 중 레이저 프린터를 주문하기

로 결심한다. 그는 이전에 구성된 자신의 지식 공간에 접근하고, 구성된 토픽을 살펴보고 탐색 때 습득한 정보를 회상한다. 그는 X사의 레이저 프린터를 떠올리고, “X가 생산한 레이저 프린터”라는 토픽을 선택한다. 토픽의 자원(Occurrence)들을 살펴보면, 가격이 생각보다 꽤 높다는 것을 알게 되고 이전에 방문했던 페이지를 다시 방문하여 가격을 비교해보겠다고 생각하게 된다. 사용자는 토픽을 만들면서 방문했던 페이지들의 세션 정보를 찾아봄으로써 방문했던 페이지들의 정보를 얻을 수 있었다. (Backward & Forward).

이 시나리오에는 제안한 프레임워크가 사용자가 과거에 탐색한 정보를 어떻게 재 활용할 수 있게 해주는지를 보여준다. 제안한 프레임워크의 지식 공간이 검색결과를 사용자의 인지 구조에 맞게 지식 단위에서 조직할 수 있도록 하기 때문에, 더 효율적으로 재사용을 지원해 줄 수 있다. 이때, Backward와 Forward 명령이 이를 위한 도움을 제공하며, 이는 과거 Walden’s Path [11] 부터 사용된 기능으로 지식의 재사용에 높은 효과를 보여 왔다.

위 시나리오에서 살펴보았듯이, 제안한 프레임워크의 지식 공간을 이용한다면, 많은 검색 세션 후에도 만들어진 토픽-관계 네트워크를 통해 수많은 정보 들 속에서 사용자들은 원하는 정보를 좀 더 효율적으로 찾고 활용할 수 있을 것이다.

구 현

본 연구에서는 위에서 소개한 이단 모델에 기반 한 새로운 상호작용 프레임워크를 설계하였고, 그에 따른 프로토타입 시스템을 구현하였다. 그림 2에서 볼 수 있듯이 본 프레임워크에서는 상호작용 관리모듈을 통해 사용자를 웹에 연결하여 준다. 사용자 쪽에는 지식 공간 구성을 위한 워크스페이스가 존재하고, 웹 쪽에는 일반적으로 정보를 얻는 데에 사용되는 검색엔진 및 브라우저가 존재한다. 검색엔진이나 브라우저를 이용하여 사용자가 웹을 검색하고 항해탐색 할 때, 본 프레임워크는 웹과 사용자 사이에서 상호작용을 제공하여 정보탐색 과정 중에 발생하는 인지부담을 경감시켜준다. 상호작용관리 모듈에서는 검색/브라우징, 워크스페이스 생성 보조 등을 위한 다양한 모듈이 존재한다.

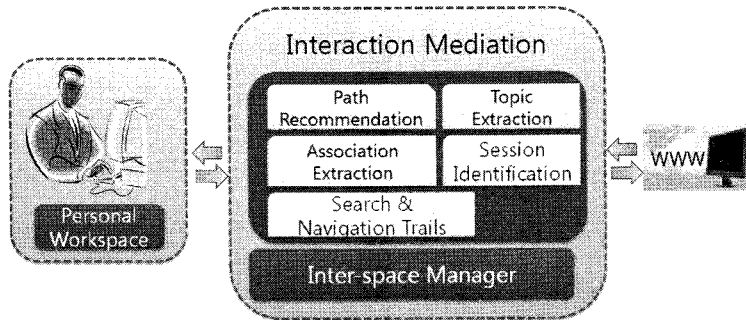


그림 2. 상호작용 프레임워크

개인 워크 스페이스

그림 3은 프로토타입 시스템의 인터페이스를 전체적으로 보여주고 있으며, 그 왼쪽이 사용자의 워크스페이스 부분이다. 워크스페이스 내 (1) 에서 보여 지는 것은 지식 공간의 토픽과 관계들이다. 이들은 프레임워크에서 제공하는 상호작용을

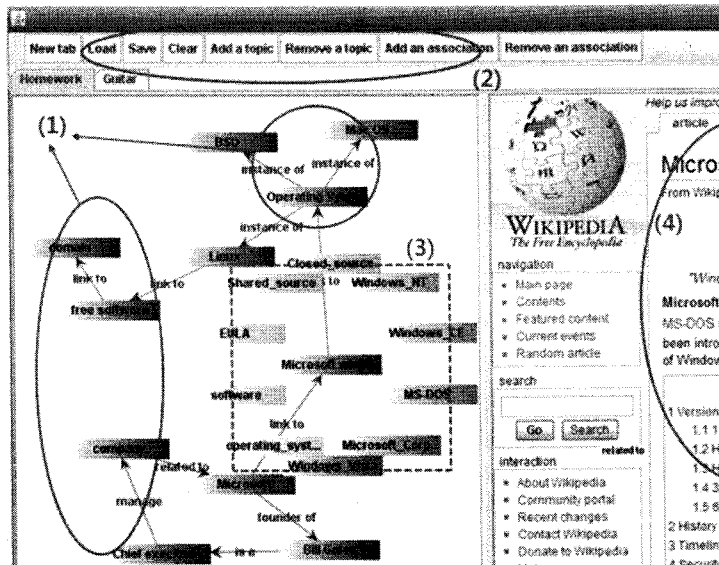


그림 3. 프로토타입 시스템 화면

통하여 1차적으로 브라우저 (4) 에서 이루어지는 검색/항해탐색 궤적들이 추출되고 토픽 및 관계 추출 모듈을 통해 아래와 같은 모습을 보이게 된다. 여기서 (2) 같은 조작 틀을 이용해서, 1차적으로 만들어진 토픽과 관계들을 수정하고, 새로운 지식 객체들을 만든다. 이때 사용자 인지부담의 추가적 경감을 위해서, 기존에 만들어진 지식 공간의 정보들을 이용하는 경로 추천 알고리즘을 통해 향후 사용자가 방문할 만한 경로를 추천하여 준다. 추가로 방문할 만한 페이지의 우선순위는 색깔로 나타내주게 되며, 빨강색이 가장 높은 순위이다.

상호작용 지원

경로 추천 모듈

경로 추천 모듈은 [12]에서 고안한 경로추천 알고리즘을 이용하며 사용자의 액션들을 바탕으로 동적으로 방문할만한 경로를 추천하여 준다. 시맨틱 네트워크에서 추론에 널리 쓰이고 있는 Spreading Activation 기법 [13]을 사용하였으며, 최고 10개의 경로를 워크스페이스 내 인터페이스를 통해 추천하여 준다.

검색 및 항해탐색 궤적 추적

이 모듈은 지식 공간의 구성을 위해, 질의 궤적과 질의 후 항해탐색 궤적을 클릭 스트림을 기준으로 관찰하여 워크스페이스에 토픽 및 관계의 형태로 기록하여 준다. 이는 사용자가 일일이 워크스페이스를 생성하지 않아도 되도록 도와주기 위하여 고안되었으며, 해당 궤적을 바탕으로 사용자가 지식 공간을 구성할 수 있도록 도와준다. 과거의 몇몇 응용(예: 구글 노트북[14], 야후 마이웹[15], Pathway[16])들이 이 같은 방법으로 검색 및 항해탐색 궤적들을 기록하는 기능을 제공하였고, 그 유용성을 보여주었다.

기타 모듈

이 밖에도 프레임워크 상에는 나타나 있지만, 프로토타입 시스템에서 완전히 구현되지 않은 네 가지 모듈이 있다. 토픽 추출기(Topic Extraction Module), 관계 추출기(Association Extraction Module), 세션 식별기(Session Identification Module), Inter-space

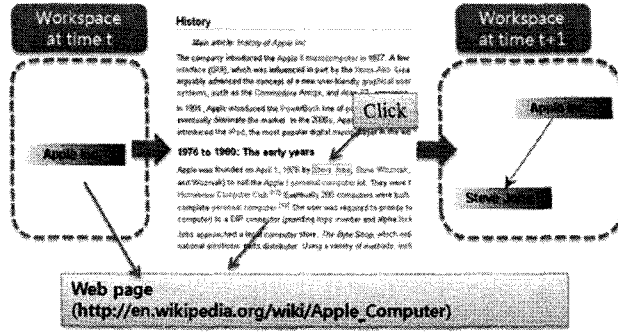


그림 4. 항해 탐색 궤적

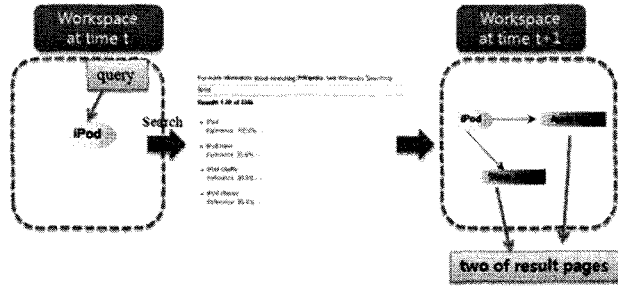


그림 5. 질의 궤적 기록의 스냅샷

Manager가 바로 그것이다. (1) 토픽 추출기는 사용자가 지식 공간의 토픽을 생성할 때 관계를 형성하는 정보 객체의 내용을 바탕으로 해당 토픽의 레이블을 추천하여 준다. (2) 관계 추출기는 자동으로 새로운 토픽과 기존에 존재하던 토픽 사이의 관계를 추천하여준다. 위의 (1)~(2) 모듈들은 검색 및 항해 탐색 궤적 추적 모듈과 함께 지식 공간을 조직하는 과정을 지원하기 위한 모듈이며, 자동화된 방법을 통하여 사용자가 참여하는 과정에서 발생할 수 있는 부담을 절감시키는 데에 도움을 준다. (3) 세션 식별 기는 자동으로 다른 세션들과의 경계를 찾는다. 여기서 세션은 관심사의 단위로, 경계를 식별한다 함은 다른 관심사로 전환되는 시점을 찾는 것을 의미한다. (4) Inter-space Manager는 두 공간 사이를 관계 지어주는 역할을 하며, 이단 모델의 공간 간 명령들을 수행한다. 현재 이 모듈들의 부재로 사용자들의 추

가적인 작업 부담이 존재하는 상태이지만, 제안한 프레임워크가 인지부담의 경감에 주는 영향을 실험(5 절)하는 데에는 큰 문제가 없었다.

실 험

본 논문에서 제안한 이단 모델과 그에 기반 한 프레임워크를 평가하기 위하여, 구현한 도구에 대한 두 가지 사용자 테스트를 수행하였다. 첫 번째에서는 제안한 도구가 사용자의 인지부담 경감에 주는 영향에 대하여 실험하였다. 즉, 탐험적 검색(Exploratory search) 시 제안한 도구가 검색과 브라우징을 통해 습득한 정보를 지식 공간을 통해 조직하고 사용하는 것이 인지 부담의 경감에 어떤 영향을 주는지 확인하였다. 두 번째 실험에서는 사용자가 정보 탐색을 하는데 있어 필수적으로 거치게 되는 정보의 재사용 과정에 제안한 프레임워크가 어떤 영향을 주는 지를 실험하였다.

인지 부담 경감

이 실험은 이단 모델에 기반 한 시스템이 사용자의 인지부담 경감에 주는 영향을 알아보기 위한 실험이다. 이 실험에서는, Subject Workload Assessment Technique (SWAT) [17]라는 특별한 측정수단을 사용하였다. 이 측정법은 시간(Time), 심적 노력(Mental effort), 그리고 스트레스(Stress) 세 가지 변수를 평가하게 되며, 이 측정법은 정보 탐색 작업을 하는 동안 사용자의 구체적인 인지 부담 경감 정도를 알아볼 수 있기 때문에, 구현한 도구의 성능 평가에 적합한 측정 수단이다.

5명의 실험참가자들은 다양한 정보들에 대한 배경지식을 습득하고 정보를 탐색하는 데에 적합한 환경을 제공하는 위키피디아³⁾에서 총 10가지 탐험적 검색을 수행하였고, SWAT을 통해 인지 부담의 경감 여부를 측정하는데 참여 하였다. 토픽은 10가지 서로 다른 위키피디아 카테고리로부터 선택하였고, 10가지 카테고리의

3) Simple English Wikipedia (http://simple.wikipedia.org/wiki/Wikipedia:Simple_English_Wikipedia)

분류 스키마는(http://en.wikipedia.org/wiki/Wikipedia:Requested_articles#Other_classification_schemes)에서 찾아 볼 수 있다. 실험하는 동안 참여자들은 주어진 토픽에 대하여 정보를 모으고 분석하면서 결과물로 주제에 대한 요약 및 여러 가지 분석 자료(예: 주어진 토픽에 대한 여러 중요 요소들과 범주의 목록)(18)를 작성하였다. 그리고 좀 더 실제에 가까운 탐험적 검색 환경을 제공하기 위해서 미리 정해진 문제가 아닌 사용자가 채워야 하는 빈 템플릿만을 제공하였으며, 학습효과를 최소화하기 위해 사용자에게 두 실험 변인 (인터페이스의 사용과 비사용)을 번갈아 가면서 실험하게 하였다.

참가자의 평가는 1 점(최고점) 부터 3 점까지 부여하게 하였으며, 그 결과는 표 1이 보여주고 있다. 제안한 도구는 인터페이스가 없는 경우를 현저하게 앞선 4.6의 총점을 얻었다. 특히, 의도한 것처럼 심적 노력의 경우에서 가장 큰 차이를 보여주었다. 이런 결과는 제안한 툴이 세 가지 면에서 모두 인지부담을 줄이는데 도움을 주었음을 보여주었다.

표 1. SWAT 테스트 결과

	제공하는 인터페이스 미사용 (Average SD)	제공하는 인터페이스 사용 (Average SD)	차이
시간	1.8 (0.45)	1.6 (0.55)	- 0.2
심적 노력	2.4 (0.55)	1.2 (0.45)	- 1.2
스트레스	2.2 (0.45)	1.8 (0.45)	- 0.4
계	6.4 (0.89)	4.6 (0.89)	- 1.8

정보 재사용

이단 모델과 그에 기반 한 도구의 구현은 정보탐색과정에서의 인지부담을 줄이는 것을 돕기 위해 고안되었다. 이 정보 탐색 과정은 검색/브라우징 활동으로 구성된 단순한 정보 탐색 활동뿐만 아니라 시간 후 지난 후 지난 정보의 재활용 또한 포함하며, 이 역시 향후 사용자의 인지부담에 영향을 주게 된다. 이 실험에서는 웹

환경에서 사용자가 과도한 정보의 양으로 인해 과거에 접한 정보의 재사용에 어려움을 겪게 될 때, 본 연구에서 제안한 도구가 주는 영향을 테스트하였다. 이 실험에서 제안하는 도구는 같은 목적으로 만들어진 다른 두 가지 도구(Stuff I've Seen(SIS) [19] & 즐겨찾기)와 그 효율성 면에서 비교되었다. “즐거찾기”는 PVR (Post-valued Recall) [20] 관점에서 유용한 틀이며, “SIS”는 과거에 찾은 정보를 인덱싱하여 검색하도록 도와주는 데스크톱 검색 틀이다.

실험 디자인

본 실험에서는 10개 그룹의 학생 (총: 30명) 들이 참가하여 6가지 태스크를 통해 세 가지 방법(즐거찾기, SIS, 제안한 도구)에 대한 효율성을 평가하였다. 세 명의 참여자가 한 그룹을 형성하여 실험에 임하였으며, 이들은 세 가지 방법으로 개개인에게 주어진 태스크를 수행하였다. 각 태스크는 5개의 질문으로 구성되었으며, 태스크는 의학이나 스포츠 같이 서로 다른 6개의 도메인으로 구성하였다. 참여자들은 하나의 태스크에서 주어진 30개의 웹 페이지를 읽었으며, 한 페이지 당 1분의 시간을 할당 받아 주어진 세 가지 틀을 이용해서 1분 동안 정보 탐색 결과를 정리하였다. 이 태스크 종료 후 사용자는 3가지의 질문을 받았는데 모두 읽은 페이지로부터 얻은 정보를 재사용하여야 답을 낼 수 있는 것들이었다. (예: 뉴욕 사마리안 병원의 이사회의 새로 선임된 두 명의 이사의 이름은?) 참가자들은 제한된 시간 내에 기존의 태스크를 거쳐 정리한 정보를 이용하거나 다시 페이지를 방문해서 질문에 답 하도록 하였다. 한 문제당 제한 시간은 5분으로 주었고, 그 이후의 소요시간은 무시하였다. 사용자의 학습효과와 의존도를 최소화하기 위하여, 아래 그림 6과 같은 순서로 번갈아 가며 실험하게 하였다. 그리고 정확한 실험을 위해, 실험 전에 모든 사용자가 세 가지 도구에 익숙해 질 수 있도록 같은 자리에서 사용법 강의와 함께 10분의 연습시간을 주었다.

다음은 비교 대상 방법들에 대한 간략한 소개이다. 즐겨찾기의 경우 미래에 참조하기 위하여 방문한 페이지를 저장하기 위하여 사용되며, 사용자는 폴더를 만들고 기억할 만한 페이지를 해당 폴더에 집어넣는다. 폴더는 본 논문의 토픽과 유사한 개념으로 트리 형태의 구조를 가지게 된다.

SIS의 경우 정보의 재사용을 돕기 위해 고안되었으며, 데스크톱 내의 다양한 파

일을 함께 색인 할 뿐만 아니라 질의하고 및 방문한 페이지의 색인 기능 역시 제공한다. 이 소프트웨어는 현재 윈도우 데스크톱 내에서 사용 가능하다.

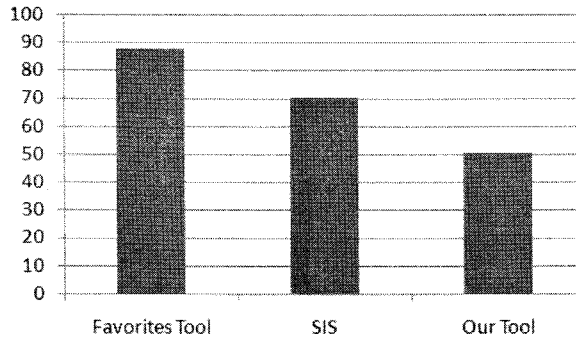
결과와 분석

10개의 그룹에서 수집한 180개 데이터의 사용도구 별 평균은 그림 7과 같다. 제안한 도구를 사용했을 때 태스크의 수행시간은 평균 50초가량 걸렸지만, 즐겨 찾기와 SIS를 사용했을 경우, 각각 88초 (76% 더 오래 걸림)와 70초 (40% 더 오래 걸림)의 시간을 소모하였다. 검색 기능만을 사용하는 SIS의 경우, 정보 탐색 결과를 사용자 스스로가 참여해서 정리하고 조직화할 수 없기 때문에, 사용자들이 주어진 시간 안에 적절한 질의를 떠올리지 못하여 답을 못 찾는 경우가 존재하였다. 비록 SIS는 페이지 정리를 위한 추가적인 노력을 요구하지 않는 도구이지만, 제안한 도

	T ₁			T ₂			T ₃			T ₄			T ₅			T ₆			
	M ₁	M ₂	M ₃	M ₁	M ₂	M ₃	M ₁	M ₂	M ₃	M ₁	M ₂	M ₃	M ₁	M ₂	M ₃	M ₁	M ₂	M ₃	
U ₁	○				○				○	○					○				○
U ₂		○				○	○					○				○	○		
U ₃			○	○				○				○	○						○

T: Task, M: Method, U: User

그림 6. 각각의 그룹을 위한 실험 설계



Unit: seconds

그림 7. 도구 별 태스크 완수 시간 비교

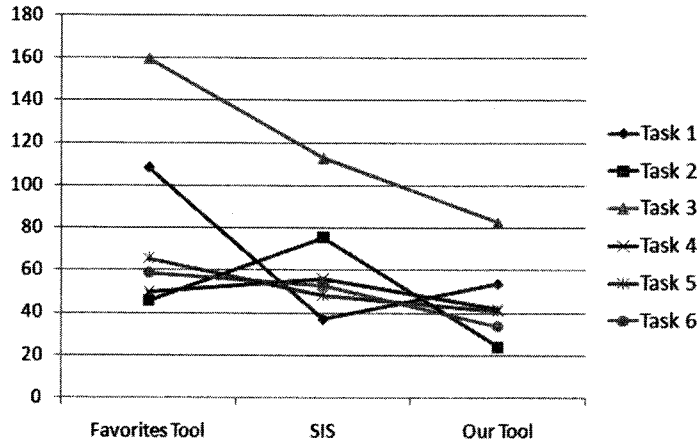


그림 8. 태스크 별 완수 시간 결과 비교

구와 즐겨찾기 역시 크게 많은 시간을 요구하지 않았으며, 특히 탐색 초기에 투자하는 시간이 많았지만 그 구조가 계속 유지되는 경우, 향후 추가적인 구성에 필요한 시간은 많지 않았다.

그림 8에서는 6가지 태스크 단위로 태스크 완수시간을 비교한 결과를 보여주고 있다. 제한한 도구는 거의 모든 태스크에서 즐겨찾기나 SIS보다 더 좋은 성능을 보여주고 있다. 예외적으로 SIS의 경우 태스크 1에서만 가장 좋은 성능을 보여주고 있는데, 이는 고유 명사 질의를 통해 주어진 질문에 쉽게 응답할 수 있는 경우였다.

추가적인 분석을 통해 각 방법들의 상대적인 장단점을 확인한 결과, 즐겨 찾기의 경우 다루기가 쉬워 계층적인 방법으로 페이지를 관리하는데 효율적이지만, 명확하지 않은 특정 정보를 찾기는 어려웠다. 특히 얇은 계층으로 즐겨찾기를 구성할 경우, 너무 많은 폴더를 생성하게 되어 필요한 정보를 포함하고 있을 지도 모를 모든 폴더를 방문하여야 하기 때문에 많은 시간이 필요하게 되었으며, 계층이 깊어지거나 뒤틀린 경우에도, 반복적으로 계층을 이동하는 시간이 추가적으로 필요하다는 단점이 생겼다. 더불어 두 페이지 사이의 의미 관계를 나타내는 것이 불가능하였다.

SIS의 경우, 사용자가 자신들만의 지식 공간을 만드는 데에 추가적인 노력이 들지 않는다는 장점이 있지만, 어떤 문맥적인 단서가 있어야 질의를 만들고 검색을 시작할 수가 있었다. 뿐만 아니라 개인화된 지식 공간을 만든다기보다 정보 공간을 검색하는 것에 더 의존적이었다.

이단 모델에 기반 한 도구의 경우 그 두 가지 면에서 최고의 면모를 보여주었다. 지식 공간 명령으로 과거에 접한 정보에 쉽게 접근할 수 있었으며, 특히, forward/backward 명령을 이용한 Sequence guide는 사용자기 쉽게 과거의 질의 세션으로 돌아가서 간단하게 해당되는 주변 지식들을 쉽게 살펴볼 수 있도록 도와주었다. 하지만, 실험에서 1분의 시간이 주어졌듯이 지식 공간의 구축에 사용자들의 노력이 필요하기 때문에, 사용자들이 여기에 들이는 노력이 추가된 부담으로 작용할 지도 모른다는 것이 단점으로 작용하였다.

위 실험 결과를 바탕으로 본 논문에서 제안한 도구가 통계적으로 유의한 차이를 보이는지를 확인하기 위해, 추가적으로 ANOVA(Analysis of variance) 검사를 통해 변수들에 따른 태스크 수행시간의 통계적 의미를 알아보았다. 그 결과는 표 2에서 보여주고 있다.

표 2. ANOVA 검사 결과

Methods	Mean	Std. Deviation	95% Confidence Interval for Mean	
			Lower Bound	Upper Bound
1: Favorites	87.69	98.82	62.16	113.22
2: SIS	70.09	67.67	52.61	87.57
3: Our Tool	50.33	43.78	39.02	61.64

ANOVA는 모든 데이터의 분산 비교를 통해 만들어진 F분포를 이용하여 가설을 검증한다. 그 결과로 F 값은 $F(2,177)=3.866$ ($p < 0.05$) 을 얻을 수 있었고 그 차이는 95% 수준에서 신뢰 할 수 있었다. 즉, 제안하는 도구가 사용자에게는 더 높은 재사용성을 제공하였다고 확인할 수 있었다.

결 론

본 논문에서는 새로운 개인화된 지식 구조화 방법과 그에 기반 한 상호작용 프레임워크를 제안하였다. 이는 검색과 브라우징을 사용한 전통적인 정보 탐색 행동에 지식 공간을 도입한 새로운 지식 구조화 방법이며, 비록 프로토타입이 제한된 기능만을 제공하였지만 실험을 통해 높은 인지부담 절감 효과를 확인할 수 있었다.

그래서 우리는 향후 방향으로 다음과 같은 것들을 계획하고 있다. 먼저, 제안한 프레임워크 위의 모듈의 성능 향상을 통해서 사용자들이 더 쉽게 지식 구조로 시스템화하는 것을 도와주려 한다. 이는 사용자들이 더 효율적으로 지식 구조를 생성할 수 있도록 도와주어, 향후 사용을 위한 양질의 데이터를 확보할 수 있도록 지원하여 준다. 그리고 모아진 지식 구조는 다른 이용자와 공유를 통해 메타데이터로서 전자 도서관이나 웹에서 가능한 새로운 상호작용이나 검색 및 질의응답 시스템을 위해 연구 될 수 있다. 특히, 지식 공간내의 객체들을 이용한 새로운 질의응답 시스템은 사용자에게 유용한 정보 탐색 수단으로 사용될 수 있을 것으로 기대된다.

참고문헌

- [1] Sugiyama, K., Hatano, K., and Yoshikawa, M. (2004), Adaptive Web Search Based on User Profile Constructed without Any Effort from Users, Proceedings of WWW 2004.
- [2] Pretschner, A. and Gauch, S. (1999), Ontology Based Personalized Search, Proceedings of ICTAI 1999.
- [3] Kelly, D. and J. Teevan. (2003), Implicit feedback for inferring user preference: a bibliography, Proceedings of SIGIR Forum 2003.
- [4] Bhogal, J., A. Macfarlane, and P. Smith. (2007), A review of ontology based query expansion, Information Processing and Management, 43(4), 866-886.

- [5] Xue, G. R., et al. (2008), Deep classification in large-scale text hierarchies, Proceedings of SIGIR 2008.
- [6] Marchionini, G. (2006), Exploratory search: from finding to understanding. Communications of the ACM, 2006. 49(4), 41-46.
- [7] White, R. W. and S. M. Drucker. (2007), Investigating behavioral variability in web search, Proceedings of WWW 2007.
- [8] Belhassen, A. S., Ben Abdallah, N., and Ben Ghezala, H. H. (2000), Cognitive approach for building user model in an information retrieval context (poster session), Proceedings of SIGIR 2000, 337-338.
- [9] Tian, X., et al. (2007), Modeling User's Cognitive Structure in Contextual Information Retrieval. Proceedings of Fourth International Conference on Fuzzy Systems and Knowledge Discovery.
- [10] Gonçalves, M. A., et al. (2004), Streams, Structures, Spaces, Scenarios, Societies (5S): A Formal Model for Digital Libraries, ACM Transactions on Information Systems, 2004. 22(2): p. 270-312.
- [11] Furuta, R., et al. (1997), Hypertext paths and the World-Wide Web: experiences with Walden's Paths, Proceedings of Hypertext '97.
- [12] Park, H., Myaeng, S. H., Jang, G., Choi, J.W., Jo, S., and Roh, H.C. (2008), Interactive information seeking interface for exploratory search, Proceedings of ICEIS 2008.
- [13] Anderson, J.R. (1983), A spreading activation theory of memory, Journal of Verbal Learning and Verbal Behavior.
- [14] Google Notebook, <http://www.google.com/notebook/>
- [15] Yahoo Myweb, <http://myweb.yahoo.com/>
- [16] Pathway, <http://pathway.screenager.be/>
- [17] Reid, G. B. a. N., T. E., ed. (1988), The subjective workload assessment technique: A scaling procedure for measuring mental workload, P.A. HANCOCK and N. MESHKATI, eds. Human Mental Workload.
- [18] Jijkoun, V. and M. de Rijke. (2006), A Pilot for Evaluating Exploratory Question

Answering, Proceedings of EESS 2006.

- [19] Dumais, S., et al. (2003), Stuff I've seen: a system for personal information retrieval and re-use, Proceedings of SIGIR 2003.
- [20] Wen, J. (2003), Post-Valued Recall Web Pages: User Disorientation Hits the Big Time, IT & Society, 1(3), 184-194.

1 차원고접수 : 2008. 10. 27

2 차원고접수 : 2008. 11. 18

최종게재승인 : 2008. 12. 1

(Abstract)

Cognitive Knowledge Structure and Information Seeking Framework to Reduce Cognitive Burden

Hogun Park Sung-Hyon Myaeng Kyung-Min Kim Gwan Jang Jong-Wook Choi
Information and Communications University

As the Web and digital libraries have become a commodity, they are used for a variety of purposes and tasks that may require a great deal of cognitive efforts. However, most search engines in the Web and digital libraries support users with only searching and browsing capabilities, leaving all the cognitive burdens of manipulating information objects to the users. We propose a two-level model for human-Web interactions, consisting of knowledge and information spaces, and a tool that provides knowledge space and inter-space operations in addition to searching and browsing at the information level. Knowledge space is an explication of user's conceptual view of the information objects being explored through interactions with the Web or a digital library. Topics are created and related with associations at the knowledge level and connected to information objects in information space. The tool implemented using the Topic Maps framework has been tested for efficacy as an aid to reducing cognitive burden under exploratory search task.

Keywords : Information Seeking, Cognitive Knowledge Structure