

키워드 기반의 다중 사이트 항해 시스템 개발

황준식[†] · 김희진[†] · 정호숙^{††} · 유수진[†] · 박성빈^{†††}

요 약

본 논문에서는 사용자의 질의어에 대한 검색 결과에서부터 출발하여 연관된 키워드를 지속적으로 제공하는 시스템을 제안하였다. 본 논문에서 제안하는 시스템의 가장 큰 목적은 사용자가 자신이 찾고자 하는 것을 정확히 설명할 수 없는 정보를 찾도록 도와주는 것이다. 본 논문에서 제안하는 시스템은 많은 정보 제공처와 연동을 통해 유용한 정보를 확보하여 그 정보들을 통합된 스키마로 표현하고, 사용자들에게 크로스 서치 환경을 제공한다. 또한 관련 키워드들간의 관계도를 보여줌으로써 사용자는 관심 있는 키워드를 선택하여 원하는 자료를 쉽게 검색 할 수 있게 된다. 따라서 사용자들은 특정 사이트에 국한되지 않은 환경에서 원하는 정보를 쉽게 찾을 수 있다.

주제어 : 크로스 서치

Development of a system for keyword-based navigation on multiple sites

Chunsik Hwang[†] · Heejin Kim[†] · Hyosook Jung^{††} ·
Sujin Yoo[†] · Seongbin Park^{†††}

ABSTRACT

In this paper, we propose a system that starts with the searching result of user's query and continuously offers new keywords related to the query. The primary purpose of the proposed system is to help users find information that the users cannot explain clearly. The proposed system obtains useful information by connecting different information sources, represents the information under an integrated schema, and offers a cross searching environment. In addition, it presents a visualized diagram that shows relations between keywords so that users can easily find the desired information. Therefore users can find information easily in an environment that is not limited to a certain site.

Keywords : Cross search

† 정 회 원: 고려대학교 컴퓨터교육과
 †† 종신회원: 고려대학교 컴퓨터교육과 박사
 ††† 종신회원: 고려대학교 컴퓨터교육과 교수(교신저자)
 논문접수: 2010년 12월 06일, 심사완료: 2011년 01월 03일

1. 서론

현재의 인터넷에는 매우 많은 자료가 있으며, 동시에 뛰어난 검색엔진이 있지만 잘 알지 못하는 내용에 대한 자료를 찾고자 노력하는 시점에서 그것을 정확히 설명하는 키워드를 안다는 것은 어려운 일이다. 이 경우 사용자는 자신이 알고 있는 연관된 키워드부터 검색을 시작하여 점진적으로 지식을 얻으며 정답에 접근해 가게 된다.

본 논문에서는 사용자들이 자신이 알고 있는 부정확한 묘사를 사용해서 정확한 질의어를 찾아주는 시스템을 구현하였다. 먼저, 사용자가 관련성이 있다고 생각되는 단어로 시작하고, 시스템의 판단 하에 연관성이 있는 다른 단어들이 제시되며 이들 후보 단어 중 사용자가 선택한 단어를 중심으로 관계도를 확장하면 자연스럽게 다수의 단어들과 관계가 중첩되는 단어들이 등장한다. 이들 단어들이 바로 사용자가 선택한 단어들과 밀접한 관련성을 가지고 있는 단어들로, 이 단어들이 사용자가 몰랐지만, 사용자가 찾고자 하는 단어들이 나타나는 자리이다.

본 논문에서 제안하는 시스템은 여러 개의 데이터 소스에 대한 검색을 동시 진행하여 그 결과를 통합된 형태로 보여준다. 시스템은 서로 관련 있는 키워드들을 관계를 나타내는 선으로 연결한다. 또한 화면에 출력할 때 연결된 키워드들을 모으고 그렇지 않은 키워드들은 거리를 두어 표현한다.

본 연구에서 사용한 데이터들은 다음과 같다. 첫째는 Wikipedia [1] 가 보유한 데이터들이다. 전세계 누구나 제작에 참여할 수 있는 이 백과사전은 그 본문 내용에 있어서 등장하는 주요단어들이 다른 페이지의 색인어로서 연결되어 있다. 즉 제공되는 본문 내용을 처리하면 단어와 단어 간의 관계를 알 수 있다. 또 다른 데이터로는 DBpedia [2], OWL [3]이나 RDF [4]와 같은 자료가 있다. 각 분야의 데이터를 체계적으로 분류, 관리하기 위해서 또는 컴퓨터에 의해 의미론적인 처리를 하기위해서 연구된 이 분야의 결과물들은 변별력 있는 단어들과 그 사이의 연관 관계를 함께 가지고 있기 때문에 이 연구에 매우 적합한 자료이다.

본 논문은 다음과 같이 구성되어 있다. 2 장은 관련 연구를 기술하고 3 장은 본 논문에서 제안하는 시스템의 구조를 설명한다. 4 장은 시스템의 작동 과정 및 본 논문에서 제안하는 시스템을 활용한 실험 결과를 설명하고 5 장은 본 시스템이 어떻게 학습에 사용되는 지에 대한 구체적인 시나리오들을 제시하며 6 장은 결론을 서술한다.

2. 관련연구

본 논문에서 제안하는 시스템과 비슷한 기능을 가지는 시스템들은 Google Wonder Wheel [5], IHOP [6], NovoSeek [7], 및 Bio2RDF [8] 등이 있다.

Google Wonder Wheel은 “키워드 연관맵” 검색 서비스이다.[5] <표 1> 은 구글 원더휠과 본 논문의 시스템간의 차이를 나타낸다.

<표 1> 구글 원더휠과 본논문의 시스템 비교

	Google Wonder Wheel	본 논문의 시스템
사용자	웹에 존재하는 일반적인 정보를 이용하려는 사람들	Wikipedia등 특정 데이터베이스의 정보를 이용하려는 사람들
연관 검색어 제시 범위	두 개의 원까지의 연관 검색어만 보여준다. 즉, 여러 단계를 거치게 되면 이전의 검색 쿼리의 연관 검색어는 안 보여준다.	사용자가 선택한 모든 쿼리의 연관 검색어를 모두 여러 단계를 거치더라도 이전의 검색 쿼리의 연관 검색어를 계속 보여준다.
연관 검색어 연결 범위	여러 단계를 진행하다가 임의의 검색 쿼리의 연관 검색어가 이전에 선택한 다른 쿼리의 연관 검색어와 관련이 있더라도 이를 연결시켜 보여주지 않는다.	여러 단계를 진행하다가 임의의 검색 쿼리의 연관 검색어가 이전에 선택한 다른 쿼리의 연관 검색어와 관련이 있다면 연결시켜 보여준다.

구글 원더휠과 달리 본 논문에서 제안하는 시스템은 집단 지성 또는 전문가에 의해서 매우 잘 정제된 데이터를 사용함으로써 신뢰할 수 있는 데이터를 사용한 연산의 결과를 제시한다. Wikipedia, OWL, 심지어 특정한 이슈만 다루는 사이트이라 할지라도, 거기에 등재된 내용은 지식을 갖춘 사람에 의해 충분히 검토된 데이터들이

다. 반면, 구글 원더 휠에서 사용되는 사용자의 질의를 재가공하는 방식은 근본적으로 이것을 찾던 사용자들은 이것도 찾았었다는 맥락에 제시되는 단어이다. 이것은 비슷한 검색을 하던 사람들의 검색어를 알려주는 것으로 체계적으로 정리되어 있는 데이터를 탐색하는 것이 아니라 비슷한 처지의 다른 사람의 것을 보는 것이기 때문에 신뢰도에 잠재적인 문제가 항상 존재한다.

iHOP [6] 은 gene과 protein을 기반으로 생물학 문서들을 서로 연결시킨다. 즉, 임의의 gene이나 protein entity를 관련된 PubMed abstract와 연결시킨다. 생물학 관련 문서들을 연결시켜 마치 인터넷과 같은 방법으로 탐색할 수 있도록 gene과 protein의 network로 표현한다. <표 2> 는 IHOP 과 본논문의 시스템의 비교를 나타낸다.

<표 2> IHOP 과 본논문의 시스템 비교

	iHOP	본 논문의 시스템
유사 점	abstract 에서 추출한 gene 과 protein 을 network 로 표현함 특정 gene 과 관련된 abstract 를 연결시킴	검색된 문서에서 주요 term 을 추출하여 관련된 concept 를 연결시킨 concept network 를 생성함 특정 term 과 관련된 Wikipedia article 이나 검색 결과를 연결시킴
차이점	abstract 를 indexing 하여 gene synonym 추출 gene 과 abstract 에 대한 XML 문서를 생성하여 이를 HTML 문서로 제시함	검색된 page 의 title 을 파싱하여 term 추출 검색된 개념들의 연결 관계를 생성하여 concept map 으로 보여줌

NovoSeek [7] 은 검색 결과를 이해하기 쉽게 필터링하기 위해서 PubMed에 있는 entity identification을 수행하는 생물학 문서 검색 엔진이고 full text indexing을 통해 생물학 문서에서 gene, disease, drug, chemical과 같은 생물학 개념을 추출한다. NovoSeek 의 특징은 사용자 질의어가 언급된 문서를 검색해줄 뿐만 아니라, 검색된 문서에서 그 질의어와 관련된 개념들도 추출하여 제시한다. <표 3> 은 NovoSeek 과 본논문의 시스템과의 비교를 나타낸다.

<표 3> NovoSeek 과 본논문의 시스템 비교

	NovoSeek	본 논문의 시스템
유사 점	여러 데이터베이스 (PubMed, CDC, NIH 등) 에 저장된 문서들을 한꺼번에 검색 질의어와 관련된 개념들을 추출하여 제시함	여러 데이터베이스 (Wikipedia, DBpedia, drugbank, chemistry) 에 저장된 문서들을 한꺼번에 검색 질의어와 관련된 개념들을 추출하여 제시함
차이점	text indexing 을 통해서 개념 추출 관련 개념들을 리스트로 제시 사용자 질의어와 관련된 문서가 검색된 후 관련 개념을 선택하였을 때 새로운 검색이 아니라 검색된 문서들이 선택된 개념을 기준으로 필터링 됨	검색된 page 의 title 을 파싱하여 term 추출 개념들 간의 연결 관계를 map 형식으로 제시 사용자의 질의어와 관련된 문서 검색 후 관련 개념을 이용하여 새로운 질의 생성 및 검색이 가능

Bio2RDF [8] 는 의학, 생명공학 관련 정보를 RDF로 통합하고 이를 RDF Browser들을 사용하여 검색할수 있게하는 프로젝트이다. Bio2RDF는 외부 데이터 소스로부터 XML, RDF, Text, HTML등을 받아들이며, 이를 RDF로 재가공하기 위해 RDFizer[9]를 JSP로 제작하여 입력된 데이터를 처리하였다. 이렇게 가공한 데이터를 내부 TripleStore에 보관하여 Sesame[10], Piggy Bank[11], Tabulator[12,13,14] 등 RDF Browser들의 요청에 따라서 제공할수 있다.

Bio2RDF는 External Data Source들로부터 데이터를 수집해서 이를 RDF로 변환하여 Bio2RDF.org의 중앙의 TripleStore에 대부분의 자료를 보관해두고 있다. 여기서 데이터들은 가공되지 않은 고유의 포맷으로 각각 다른 저장소에 존재하다가 RDFizer에 의해서 통합하기 좋게 가공된다. 이와 유사하게 본 연구에서도 웹 상에 흩어져있는 데이터를 수집하여 고유 데이터 구조로 변환하여 통합을 수행한다. 위키피디아 같은 경우는 그 HTML문서나 XML Export 포맷을 통해 자료를 받아온다. 이렇게 받은 자료를 RDFizer에 상응하는 역할을 해주는 Advisor를 통해서 고유의 데이터 스키마로 변환하여 통합한다.

3. 시스템 구조

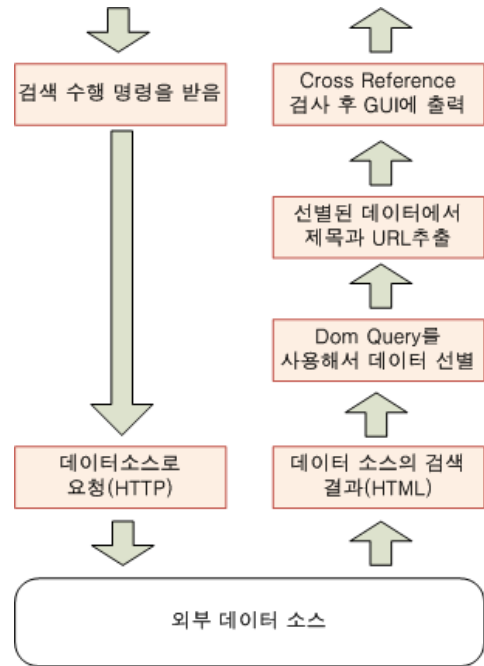
본 논문에서 제안하는 시스템은 서로 상이하게 정의된 복수의 데이터 소스로부터 통합된 데이터 구조를 만들어 탐색을 수행한다. 시스템은 통합된 데이터 구조를 키워드, 문서, 관계 세가지 요소로 정의한다. 키워드는 검색어(Search Term)를 의미하고 문서는 검색의 결과(Search Result)이다.

시스템은 크게 Platform과 Advisor로 구성되어 있다. Platform은 지식지도를 구성하기 위한 단일화된 환경을 구축하는 것이 목적이다. 여기서 단일 환경은 키워드, 관계, 문서로 정의되며 Platform은 이 자료들을 입력 받아 적절하게 표시, 탐색하고 해당 문서를 연결해주는 기능을 가진다. 또한 Advisor를 관리하며, 사용자의 질의를 Advisor로 보내는 기능을 수행한다. Advisor의 역할은 개별 데이터 소스를 해석하여 Platform이 처리할 수 있도록 변환하는 것이다.

Advisor로부터 Platform으로 데이터가 전송되면 Platform은 데이터 등록을 위한 일련의 과정을 수행한다. Platform을 지금까지 사용자와 Advisor의 활동을 통해 데이터를 누적해온 하나의 데이터 스트럭처를 가지고 있다. Advisor로부터 제공받은 데이터는 이처럼 통합되어 있는 큰 데이터에 삽입되게 된다. 그 과정을 위해서 Platform은 우선 노드를 데이터 구조에 추가를 한다. 그리고 Advisor가 제공하는 관계정의에 따라서 연관된 노드를 검색한다. 이때 연관성이 있는 노드를 발견하며 두 노드사이에 엣지가 존재함을 표기해 넣는다. 이러한 과정을 통해서 여러 데이터들과 연관 관계를 저장하게 된다. 데이터 통합 작업이 끝나면, 해당 데이터를 전송하여 화면 출력할 것을 지시한다. 이러한 일련의 과정은 다수의 프로세스를 통해서 동시에 빠르게 진행됨으로, 사용자는 순간적으로 많은 노드가 추가되는 것을 볼 수 있다.

각 Advisor는 질의어를 받아 특정 데이터 소스에서 검색을 수행하고, 그 결과를 해석하여 키워드, 문서, 관계를 추출한다. 따라서 Advisor가 사용하고자 하는 데이터 소스는 검색 기능을 지원해야 하고 데이터 소스가 보관하고 있는 고유의 구조를 키워드, 관계, 문서의 정형화된 형태로 해

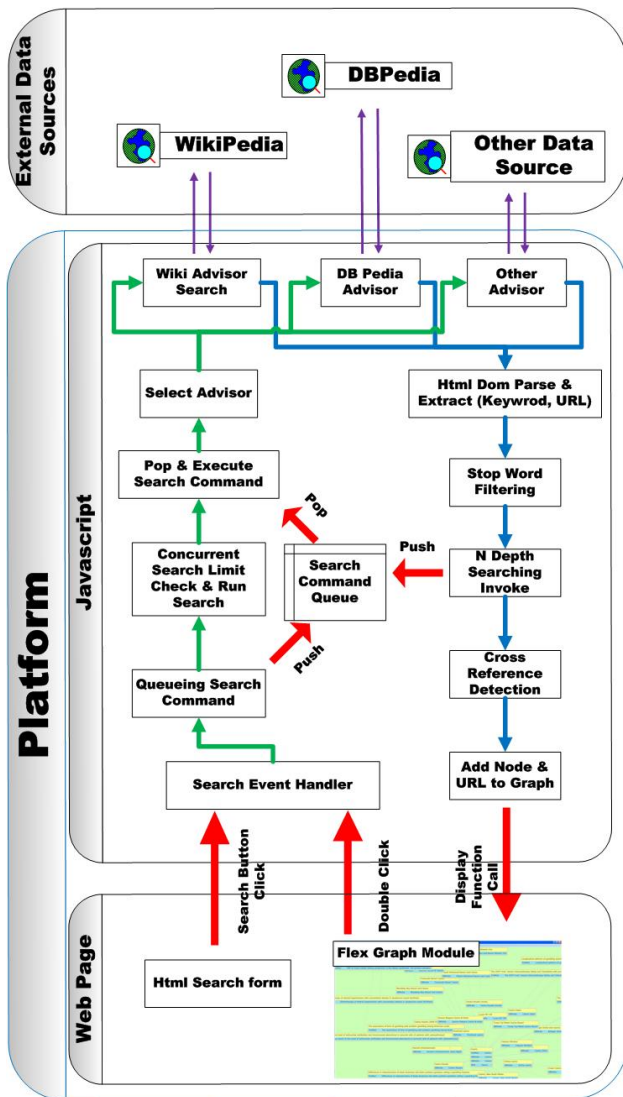
석 할 수 있어야 한다. Advisor의 개별 구현은 각 데이터 소스에 따라서 상이 하지만 그림 1과 같이 공통적인 구조를 갖게 된다.



<그림 1> Advisor의 공통 구조

현재 구축된 Advisor들은 전부 HTML을 사용하고 있으며, 따라서 HTTP GET을 사용하여 검색을 하고 있다. 이렇게 검색된 데이터들은 해석의 과정을 거쳐서 Platform의 적합한 구조로서 재해석된다. 수집된 HTML은 Dom Query를 통해서 간단하게 정보를 추출할 수 있다. Platform이 통합하려고 하는 외부 데이터 소스의 HTML들은 웹 프로그램이 데이터베이스에 저장된 자료를 사용하여 자동 생성한 것이므로 매우 정형화되어 있고, 데이터의 구분이 비교적 명확하다. 기본적으로 자료의 출처가 DB에 저장되어 있던 것이기 때문이다.

그림 2는 프로그램 내의 모듈 구성 단위와 모듈간의 통신을 보여주고 있다.



<그림 2> 시스템 구조

4. 시스템 작동 과정 및 실험

시스템은 기본적으로 사용자의 질의에 대하여 일련의 검색과정을 수행한다.

시스템은 다음 세가지 방법을 이용하여 질의어와 연관된 개념을 얻는다.

첫째, 시스템은 사람의 판단으로 정리된 데이터를 사용한다. 온라인 상에는 수 많은 데이터들이 있는데, 상당수의 데이터들은 사람이 활용하기 위해서 사람의 손으로 만들어지거나 정리해둔 데이터들이다. 특히 분명한 목적성을 가지고 구축된 사이트의 경우는 더욱 그렇다.

둘째, 시스템은 군집성을 가진 단어의 특성을 이용한다. 보통 특정한 분야나 하나의 개념과 연

계되는 단어 그룹은 상당히 한정적이다. 따라서 이러한 단어가 질의어로서 언급되는 문서는 그 연관성의 비중의 크고 작음이나 종류와 관계없이 연관되어 있다. 일반적으로 문서는 설명하고자 하는 개념과 연관된 단어가 많이 분포되어 있다. 그리고 가장 대표성을 갖는 단어가 해당 문서의 제목이 된다. 제목은 문서가 설명하고자 하는 개념을 대표한다. 따라서 하나의 질의어가 포함된 다수의 문서를 얻은 뒤, 문서 집합의 제목만 수집하여도 질의어와 연계되며 어떠한 개념을 대표하는 키워드를 획득 할 수 있다. 이를 이용하면 기본적인 단어 일치 기술만 사용하더라도 충분히 많은 숫자의 문서를 얻을 수 있으며, 문서의 갯수만큼의 키워드를 얻을 수 있다.

셋째, 시스템은 다양한 관점의 교차 적용을 통해서 더 좋은 결과를 얻고자 하였다. 접근방법의 특성상 시스템은 매우 많은 데이터를 다루게 되는데 연결된 다수의 데이터베이스를 대상으로 하는 여러 질의어에 대한 검색 결과물에서 일정한 숫자 이상의 연결성을 발견할 경우 연관관계를 인정한다. 검색 결과로 나온 키워드는 일정 숫자 이상의 검색 결과 집합에 동시에 소속되어야 함으로 빈출 키워드 및 공통성을 가진 키워드를 선별해 낼 수 있다.

구체적으로 시스템이 질의어를 처리하는 과정은 다음과 같다.

첫 단계는 사용자가 검색어를 넣는 부분이다. 사용자는 직접 키보드를 사용하거나 검색 결과물로 등장하는 키워드를 더블 클릭하여 선택함으로써 시스템에 검색을 지시할 수 있다. 시스템은 사용자로부터 검색어를 받으면 해당 검색 명령을 큐에 추가 해둔다. 웹 환경에서 다수의 검색을 동시 수행하기 어려워, 큐잉을 해두고 점진적으로 처리를 한다. 큐에 저장된 명령은 Advisor의 최대 실행 갯수의 한도 내에서 동시에 처리 될 수 있다. 시스템은 Advisor를 더 실행 할 수 있는지 상황을 판단하여 최대한 효율적으로 검색을 수행한다. 각 검색 명령이 처리될 때 우선 적합한 Advisor를 선별하고, 해당 Advisor에 검색어를 넘긴다.

시스템에서는 검색을 HTTP GET으로 수행함으로써, 각 Advisor에 재정의된 규칙에 따라서 질의

URL을 생성하고, 이를 AJAX로 요청한다. 현재의 웹브라우저들은 다른 도메인에 대한 AJAX 요청을 차단하기 때문에 이 시스템에서는 서버에 proxy를 준비하여 AJAX를 요청하고 있다. 서버에서는 클라이언트를 대신하여 페이지를 요청하고 결과를 받아서 클라이언트, 즉 웹 브라우저에게 보내준다. 웹 브라우저는 검색 페이지의 결과를 HTML으로 다운로드 받게 되면 Advisor는 이를 지정된 Dom Query문을 사용해 적절한 데이터를 선별한다. 이 시스템에서는 Dom Query를 이용하기 위해서 JQuery [15] 를 사용하고 있다. 또한 이러한 과정은 개별 Advisor별로 재정의 할 수 있게 함수들로 구성되어 있으므로 HTML이나 XML가 반환값이 아닌 다른 데이터 소스와 연동하기 위한 변형된 Advisor를 제작하는 것도 가능하다.

데이터를 선별한 뒤에는 정확한 제목(=키워드)와 해당 URL을 추출하게 된다. 제목이 추출되면 제목이 의미가 있는 데이터인지 아닌지 불용어 여부를 판단한다. 현재는 간단히 등록된 불용어중 포함되어 있는지를 테스트 하지만 확장 가능한 함수로 구성되어 있어 테스트의 조건을 수정 할 수 있다. 해당 함수가 불필요한 자료라 판단하여 결과로 False 값을 반환하면 해당 데이터를 버린다. 여기서 시스템은 한 가지 판단을 더하게 되는데 충분한 숫자의 백그라운드 검색을 위해서 검색결과를 이용한 재검색을 수행할지 여부를 결정하게 된다.

검색된 데이터는 우선 메모리에 존재하는 데이터 스트럭처에 등록이 된다. 그 후 이렇게 등록된 데이터가 기존에 검색이 되었던 다른 데이터들과 Cross Reference를 형성하는 지 검토한다. Cross Reference를 형성한다면, 여러 검색 결과 셋에서 동시에 등장하는 중요한 데이터임으로 이를 화면에 출력한다.

이러한 결과물은 사용자가 정확한 질의어를 모르고 있을 때 주변부 단어들을 이용해서 중심어로 접근 할 수 있음을 보여주는 것이다. 시스템은 직접적인 연관성이 없는 단어 집합 사이에 징검다리를 놓음으로써 사용자가 인지하지 못했지만, 검색에 사용했던 단어 모두를 결과로서 포용 할 수 있는 새로운 단어를 제시하게 된다. 시스템의

동작은 물리적으로 보았을 때는 여러 검색결과에서 교차되는 내용을 선별하여 제시하는 형태를 취하고 있지만, 인간이 정리해둔 제목 - 상세 내용의 자료 체계와 여러 데이터 검색 결과를 동시에 고려하는 접근이 대안이 되는 키워드를 제시하는 것을 가능하게 만들어준다. 여기에 각 데이터 소스에 특성에 따라 구현된 Advisor가 Platform이 항상 일관된 형태 데이터를 사용한 데이터 처리를 할 수 있게 해줌으로써 매우 높은 확장성 및 환경 적응성을 가지게 되어 새로운 도메인에 대한 적용가능성이 크게 열려있다.

실제 본 논문에서 제안하는 시스템이 사용자 관점에서 어떻게 도움이 될 수 있는 지를 보이기 위해 해당 분야에 대한 배경 지식이 낮은 학생이 시스템을 이용하여 달의 운동에 대한 학습을 할 경우에 대한 실험을 하였다.

먼저 학습자가 알아야 할 중심 개념들로 다음과 같은 개념들을 가정한 후 본 논문의 시스템과 유사한 기능들을 제공하는 Google 의 Wonder Wheel 및 본 논문에서 제안하는 시스템을 사용할 경우 어떤 결과를 얻는 지를 분석하였다.

satellite, planet, perigee, eccentricity, apogee, lunar phase, new moon, full moon, orbit, synodic period, sidereal period, lunar eclipse, umbra, penumbra

Google의 Wonder Wheel과 본 시스템을 이용하여 검색된 단어들을 비교하는 실험 과정은 다음과 같다.

① Google의 Wonder Wheel과 본 시스템 각각에서 처음에 'moon'을 질의어로 입력하여 단어를 검색한다. 그리고 검색된 단어들 중에서 위에 제시된 중심 개념에 해당하는 것을 찾는다.

② 앞서 검색된 단어 중에서 'moon phase' 또는 'lunar phase'를 질의어로 선택하여 단어를 다시 검색하고 검색된 단어들 중에서 위에 제시된 중심 개념에 해당하는 것을 찾는다.

③ 검색된 단어들 중에서 'moon eclipse' 또는 'lunar eclipse'를 질의어로 선택하여 단어를 다시 검색하고 검색된 단어들 중에서 위에 제시된 중심 개념에 해당하는 것을 찾는다.

먼저, Google의 Wonder Wheel에서 동일한 키

워드 검색할 때, 검색되는 단어를 살펴보면 다음과 같았다.

planet, moon eclipse, full moon

검색된 단어들 중에서 ‘moon phase’을 다시 검색하였을 때, 위의 중심 개념에 해당하는 것이 존재하지 않았으며 검색된 단어들 중에서 ‘moon eclipse’를 다시 검색하였을 때도 위의 중심 개념에 해당하는 것이 존재하지 않았다.

본 논문에서 제안하는 시스템을 사용할 경우 달의 운동에 대한 학습이므로 학습자는 ‘Moon’을 입력하여 단어를 검색했을 때, 검색된 단어들 중에서 위의 중심 개념에 해당하는 것은 다음과 같았다.

satellite, perigee, eccentricity, apogee, lunar phrase, new moon, full moon

검색된 단어들 중에서 ‘lunar phase’을 다시 검색했을 때, 다음의 중심 개념들이 검색되었다.

orbit, lunar eclipse

검색된 단어들 중에서 ‘lunar eclipse’을 다시 검색했을 때, 다음의 중심 개념들이 검색되었다.

umbra, penumbra

실험 결과 처음에 지정된 열두개의 중심 개념들 중 Google wonder Wheel 은 총 3개의 중심 개념을 검색했고 (약 21%), 본 논문에서 제안하는 시스템은 11개의 중심 개념을 검색하였다 (약 79%). 따라서 특정 분야 내에서 관련된 개념들을 검색하여 제시하는 기능은 본 논문의 시스템이 더 우수하다고 볼 수 있으며, 이를 학습에 활용할 때 학습자에게 더 도움이 된다고 볼 수 있다.

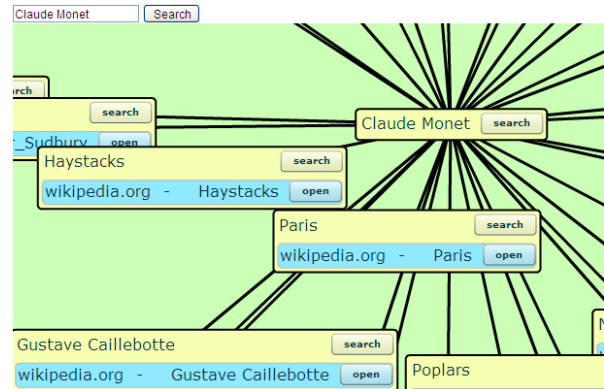
5. 사용 시나리오

본 장에서는 본 논문에서 제안하는 시스템이 어떻게 학습에 사용될 수 있는지를 보여주는 시나리오들을 설명한다.

5.1 시나리오 1

학습자는 미술시간에 미술 사조의 특징과 대표적인 화가에 대하여 공부했다. 하지만 학습자는 미술 사조는 생각이 나질 않고, 대표적인 화가인 모네(Claude Monet)와 드가(Edgar Degas)만 생

각이 난다. 관계도 기반 탐색 시스템을 이용하여 대표적인 화가를 검색하여 검색결과의 관계를 통해 학습자는 미술 사조 검색할 수 있다. 우선 학습자는 시스템에 “Claude Monet” 키워드를 입력하여 검색한다. 그림 3 은 “Claude Monet”를 키워드로 검색한 결과 화면이다.



<그림 3> Claude Monet 검색 결과

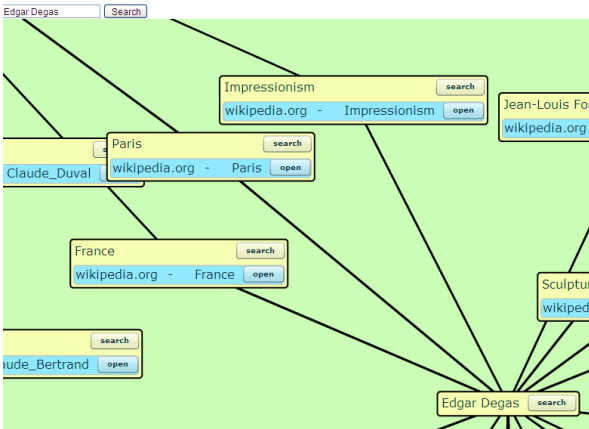
이 예제에서는 데이터 소스로서 Wikipedia 와 DBpedia 가 사용되고 있는데 모두 다수의 참여자에 의해서 가공 정리된 것이다. 화가인 모네는 미술 관련 주제의 문서에서 자주 언급이 될 수 있기 때문에 자연스럽게 미술 관련 용어, 동시대의 미술가, 작품의 명칭등과 같은 내용이 검색이 되게 된다. 이것이 적용되어 모네의 검색결과 그림 3 에서 Haystacks와 같은 모네의 작품명이나 같은 인상주의 화가이며 모네등의 후원자였던 Gustave Caillebotte의 이름을 볼 수 있게 된다.

최초의 검색이 된 이후 사용자가 관계도에 출력된 노드에 대해서 더 자세한 결과를 얻고 싶을 경우 관계도를 확장할 수 있다. 관계도를 확장하는 방법은 특정 질의어에 대하여 검색하여 그 결과를 추가하는 것이다. 또한 화면에 표시되는 노란색 노드를 더블 클릭하는 것이다.

예를들어 학습자는 Claude Monet의 검색결과만으로 모네의 미술 사조는 알 수 없다. 그래서 “Edgar Degas“ 키워드를 입력하고 search 버튼을 클릭한다. 그림 4는 “Claude Monet”와 “Edgar Degas“키워드로 검색된 결과 화면이다.

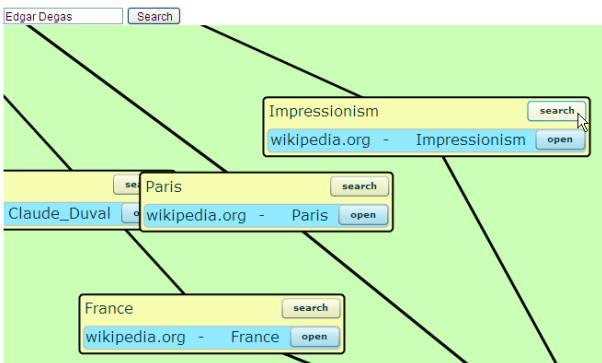
드가의 검색결과 또한 드가와 연관된 여러 단어들 존재하고 있으며, 두 검색 결과 사이에서

는 공통점을 가질 수 있다. 실질적으로 두 화가는 인상주의 화가를 대표하는 인물들도 알려져 있으므로 검색 결과의 교집합에는 이러한 점이 반영되게 된다. 마찬가지로 프랑스 파리에서 작품 활동을 하였으므로 이 내용 역시 교집합에 포함될 수 있다. 즉, 그림 4와 같이 모네와 드가는 파리, 인상주의 등의 노드로 서로 연결되어 있다. 이로써 모네와 드가를 연결하는 노드를 통해 학습자는 모네와 드가의 미술 사조가 인상주의 (Impressionism)임을 알 수 있다.



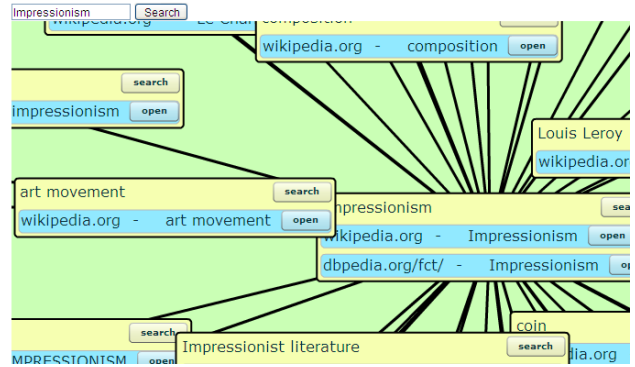
<그림 4> Edgar Dega 검색 결과

또한 학습자는 키워드를 입력하지 않고, 그림 5와 같이 Impressionism노드의 search버튼을 통해 검색할 수 있다.



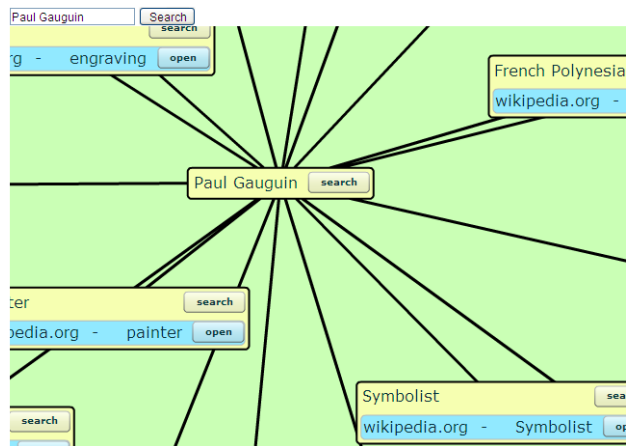
<그림 5> Impressionism노드의 search 버튼

학습자는 "Claude Monet"와 "Edgar Degas" 키워드와 "Impressionism" 키워드로 검색된 결과를 그림 6과 같이 확인할 수 있다.



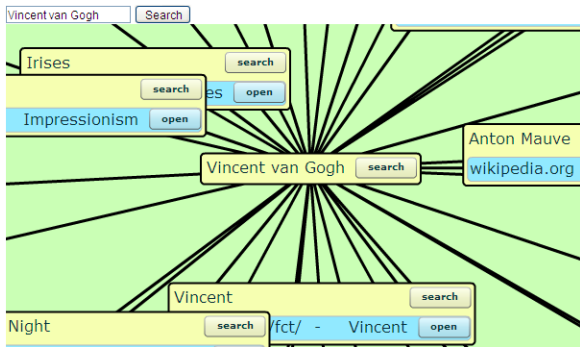
<그림 6> Impressionism 검색 결과

학습자가 고갱(Paul Gauguin)와 고흐(Vincent van Gogh)의 공통점에 대해 알고 싶다면 우선 사용자는 시스템의 입력창에 "Paul Gauguin"을 입력하고 검색한다. 그림 7은 "Paul Gauguin"키워드로 검색된 결과 화면이다.



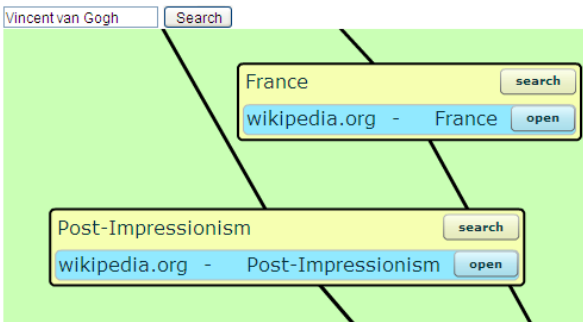
<그림 7> Paul Gauguin 검색 결과

학습자는 고갱과 관련된 많은 검색노드를 볼 수 있다. 다음으로 제안하는 시스템의 입력창에 "Vincent van Gogh"키워드를 입력하고 검색버튼을 클릭한다. 그림 8은 "Vincent van Gogh"키워드로 검색한 결과 화면이다.



<그림 8> Vincent van Gogh 검색 결과

학습자는 제안하는 시스템을 통해 고갱과 고흐가 프랑스에서 태어났거나 작업활동을 했고, 미술사조가 후기 인상주의(Post-impressionism)임을 그림 9를 통해서 알 수 있다.

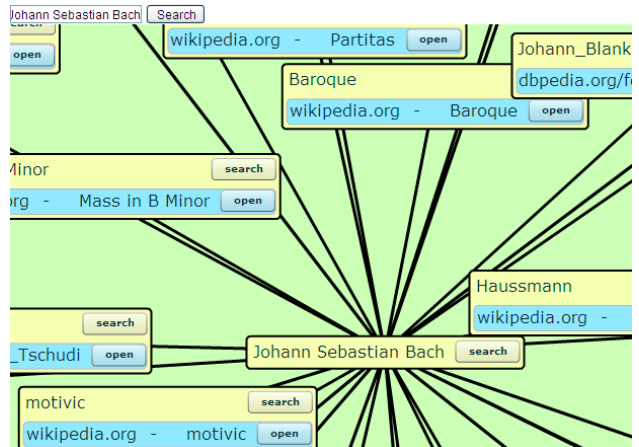


<그림 9> 공통 사항

5.2 시나리오 2

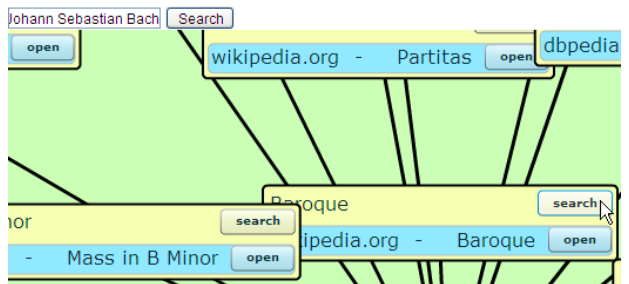
교수자는 시대별 음악사로 바로크 시대의 작곡자(Johann Sebastian Bach, Antonio vivaldi)에 대한 수업을 하고, 다음 차시로 로코코(Rococo) 시대의 음악의 특징을 안내했다. 하지만 학습자는 다음차시로 예고된 것이 기억이 나질 않아 예습을 하지 못하는 상황이다. 제안하는 시스템은 학습자에게 주변지식을 통해서 관련된 정보를 검색하여 제공한다.

우선 학습자는 “Johann Sebastian Bach”키워드를 제안하는 시스템에 입력하고 검색버튼을 클릭한다. 그림 10은 “Johann Sebastian Bach”키워드에 대한 검색 결과를 나타낸다.



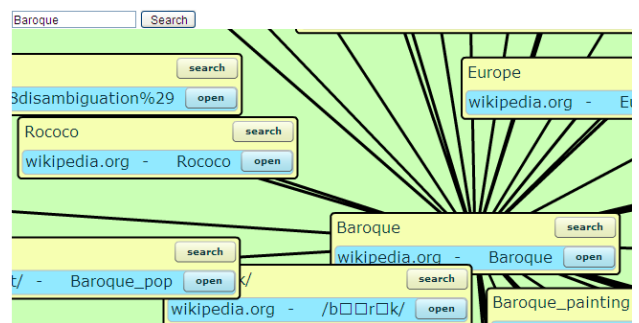
<그림 10> Johann Sebastian Bach 검색 결과

학습자는 그림 10의 검색결과에서 바로크(Baroque)노드를 찾을 수 있다. 이후 학습자는 Baroque노드의 검색버튼을 그림 11과 같이 클릭한다.



<그림 11> Baroque 노드의 search 버튼

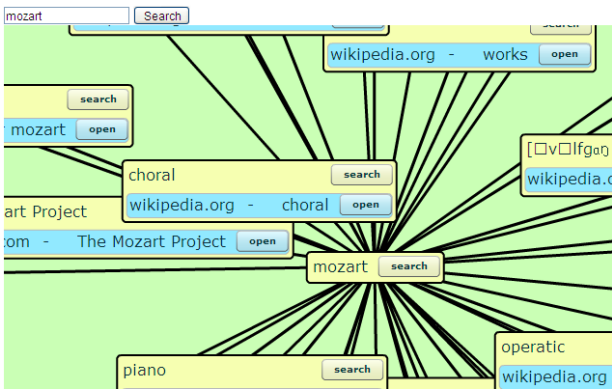
제안하는 시스템은 “Johann Sebastian Bach”과 “Baroque”를 키워드로 검색된 결과를 그림 12와 같이 확인할 수 있다. 학습자는 그림 12의 검색결과에서 “Rococo”라는 노드를 발견하고, 다음차시로 예고했던 것이 로코코시대였음을 상기할 수 있다.



<그림 12> Baroque 검색 결과

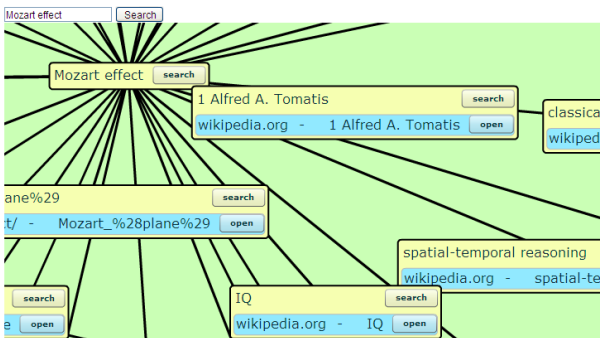
5.3 시나리오 3

어떤 학습자가 Mozart에 대한 보고서를 작성하고자 한다. Mozart와 관련된 내용 중에서 어떤 것을 주제로 보고서를 작성할 것이지를 결정하기 위해서 본 논문에서 제안한 시스템을 사용하여 Mozart와 관련된 용어들을 우선 알아본다. 학습자는 “Mozart”라는 키워드를 입력하고 “Search” 버튼을 클릭한다. 이때, 그림 13 화면과 같이 “Mozart”와 관련된 주요 개념들을 볼 수 있다.



<그림 13> Mozart 를 검색한 후의 관계도

“Mozart”와 관련된 개념을 제시된 것을 살펴보면 대부분 Mozart가 작곡한 음악의 제목이거나 관련된 사람들의 이름이다. 그 중에서 독특하게 “Mozart effect”는 개념을 발견한 학습자는 이에 대한 정보를 찾기 위해서 “Mozart effect” 개념을 더블클릭하여 이와 관련된 개념을 추가로 찾는다. 그림 14 는 “Mozart”와 “Mozart effect”키워드로 검색한 결과 화면이다.

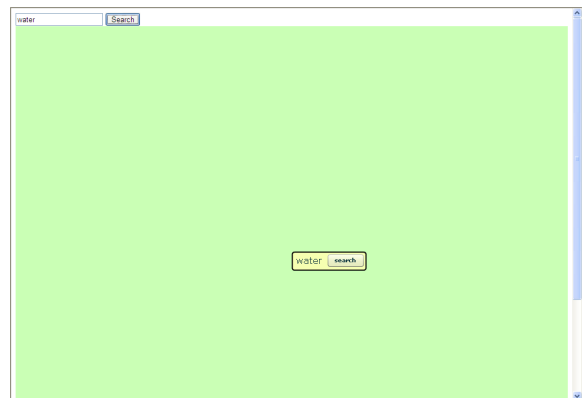


<그림 14> Mozart effect 를 검색한 후의 관계도

학습자는 “Mozart effect”와 관련된 개념들을 보면 IQ나 spatial-temporal reasoning이라는 능력과 관련되어 있다는 사실을 알게 된다. IQ는 일반적으로 잘 알려진 능력이지만 spatial-temporal reasoning이 어떤 능력을 말하는 것인지 알기 위해서 “spatial-temporal reasoning” 개념을 더블클릭하여 이와 관련된 개념을 추가로 찾는다. 검색된 결과를 살펴보면, “spatial-temporal reasoning”이 art, mathematics, science, engineering, computer science 등과 관련되어 있음을 그림 15를 통해서 확인할 수 있다. 이러한 과정을 통해서 학습자는 “Mozart effect”가 인간의 지적 능력과 관련되어 있음을 알게 되고 이와 관련된 내용을 더 조사하여 “Mozart”의 음악이 인간에게 미치는 영향에 대한 보고서를 작성한다.

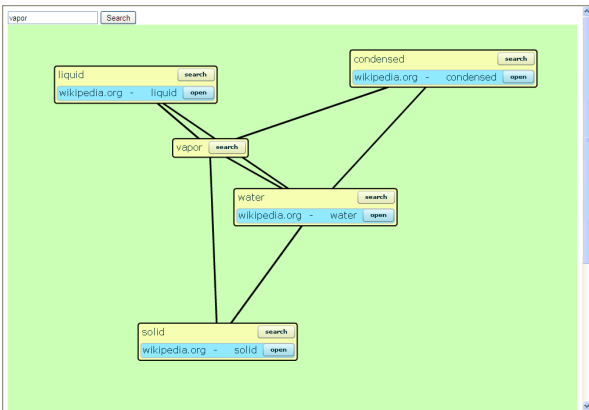
5.4 시나리오 4

어떤 학생이 과학 시간에 물의 세 가지 상태 변화에 대하여 학습한 후, 물질의 상태 변화에 영향을 주는 요인에 대하여 알아보려고 한다. 교과서에서 제시된 물의 상태 변화에서 시작한다. 먼저 ‘water’라는 키워드를 입력한다. (그림 15)



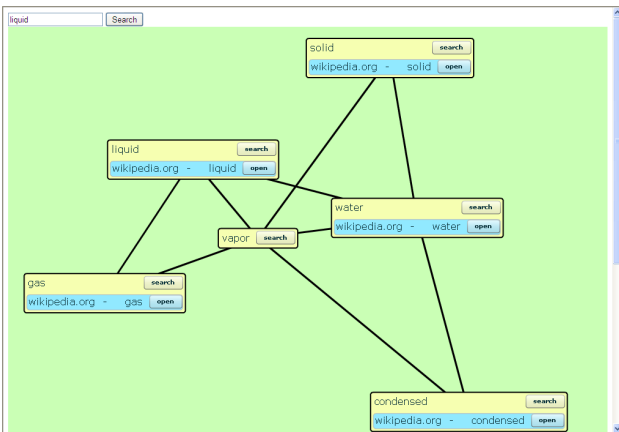
<그림 15> “water” 키워드 검색 결과

물의 첫 번째 상태변화로서 ‘vapor’를 입력한다. (그림 16)



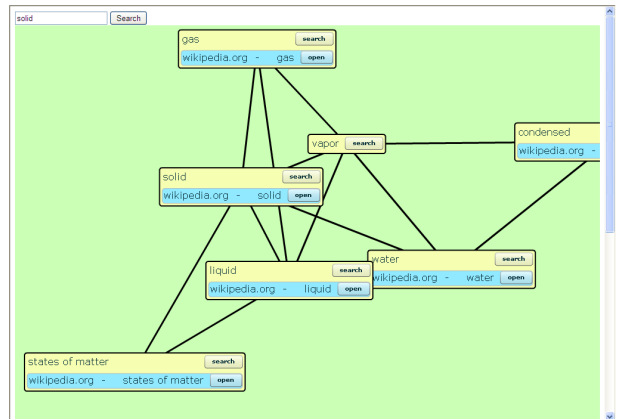
<그림 16> “vapor” 키워드 검색 결과

이때, 'vapor'와 'water'가 연결되고 물의 또 다른 상태인 'liquid'와 'solid'라는 키워드가 함께 검색된다. 두 키워드는 'vapor'와 'water'에 연결된다. 또한 기체가 액체 상태로 변화하는 현상인 'condensed'라는 키워드도 검색되고 'vapor'와 'water'에 연결된다. 각각의 물의 상태에 대한 정보를 검색하기 위해서 'liquid'의 'search'버튼을 클릭하면, 'gas' 라는 키워드가 검색된다. (그림 17)



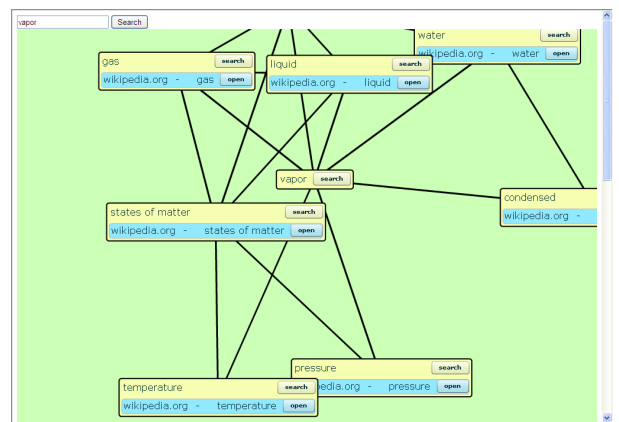
<그림 17> “liquid” 키워드 검색 결과

또한 'solid'의 'search'버튼을 클릭하면, 'states of matter'라는 키워드가 검색된다.



<그림 18> “solid” 키워드 검색 결과

학생은 물질의 상태 변화에 영향을 끼치는 요인에 대하여 공부하고자 하므로, 'states of matter'의 'search'버튼을 클릭한다. 이때, 'temperature'와 'pressure'라는 키워드가 검색된다. (그림 19) 이로써 학습자는 물질의 상태 변화가 온도와 압력과 관계됨을 알게 된다.



<그림 19> “states of matter” 검색 결과

6. 결 론

본 시스템은 웹에 게시된 방대한 데이터를 크로스 레퍼런스 시켜서 이 데이터를 기반으로 사용자가 자신이 원하는 정확한 질의어를 모를 때 웹에서 이 질의어를 찾아내는 것을 첫 번째 목적으로 한다. 보통의 검색 시스템이 자료를 찾기 위해서 존재하는 것과는 달리 이 시스템은 검색어를 찾기 위한 목적으로 고안되었다.

시스템은 사용자가 입력 또는 선택하는 단어들과 공통적으로 연결되는 단어를 찾아서 키워드

연결망에 추가한다. 이는 사용자가 알고 있는 목표 검색어와 연관성을 가진 단어를 통해서 최종 검색어를 찾아내는 것이다. 또한 시스템은 사용자가 입력 또는 선택한 키워드들과 많은 관계를 가지는 단어를 화면에 그래프로 출력하여 시각적으로 연결성을 인식할 수 있게끔 출력해준다. 연결이 복잡한 즉, 찾고자 하는 단어일 가능성이 높은 단어는 연결관계도의 중앙부에 위치하게 된다.

논문의 시스템은 검색의 최종 목적을 찾아내기 위한 표본 데이터를 선택 할 수 있다. 이는 어드바이저의 선택을 통해서 실현된다. 데이터 소스를 선택할 수 있는 응용사례로는 동일한 관심사를 지닌 사람들이 만들어낸 문서만을 활용하여 경우이다. 이때는 도메인 한정환경이 만들어지게 되면서 특정 분야에 국한된 지식을 더 효과적으로 찾을 수 있다. 이때 시스템은 그 분야의 전문가와 같은 배경 지식을 가지게 되는 것이다.

또한 논문에서 제안하는 시스템은 다양한 정보를 실시간으로 취합하여 사용자에게 탐색의 방향성을 지속적으로 제시하였다. 이는 찾고자 하는 정보를 정확히 알지 못하는 상황에서 더 많은 정보를 토대로 보다 나은 도움을 주기 위함이다. 이 프로그램은 구글 원더 휠 [5] 과는 달리 다양한 데이터 소스를 사용할수 있는 일반화된 프로그램이다. 어떤 데이터 소스를 사용하는가, 어떠한 데이터 추출 방법을 사용하는가에 따라서 다양한 응용이 가능하다. 가령, 도메인 분야에 대하여 권위있는 웹사이트들의 자료들을 수집하여 해석한다면, 해당 도메인 분야에 대해 최적화시킨 도구로 발전할 수 있다. 이를 통해 의학이나 교육 등의 분야에서 지금보다도 훨씬 효과적으로 탐색을 할 수 있는 시스템으로 발전 할 것이라 기대한다.

참 고 문 헌

- [1] <http://Wikipedia.org>
- [2] <http://wiki.dbpedia.org/About>
- [3] <http://www.w3.org/2004/OWL>
- [4] Klyne, G., &Carroll, J. J. (2004). Resource description framework (RDF): Concepts and abstract syntax. W3C recommendation.
- [5] <http://www.googlewonderwheel.com/>
- [6] <http://www.ihop-net.org>
- [7] <http://www.novoseek.com>
- [8] Belleau, F., Nolin, M. A., Tourigny, N., Rigault, P., &Morissette, J. (2008). Bio2RDF: Towards a mashup to build bioinformatics knowledge systems. *Journal of Biomedical Informatics* 41(5): 706-716
- [9] <http://simile.mit.edu/wiki/RDFizers>
- [10] <http://www.openrdf.org/index.jsp>
- [11] Huynh, D., Mazzocchi, S., &Karger, D. R. (2007). Piggy Bank: Experience the Semantic Web inside your web browser. *J. Web Sem.* 5(1): 16-27.
- [12] <http://dig.csail.mit.edu/2005/ajar/release/tabulator/0.7/tab.html>
- [13] Berners-Lee, T., Chen, Y., Chilton, L., Connolly D., Dhanaraj, R., Hollenbach, J., Lerer, A., &Sheets, D. (2006). Tabulator: Exploring and Analyzing linked data on the Semantic Web. *Proceedings of the The 3rd International Semantic Web User Interaction Workshop (SWUI06)*.
- [14] Berners-Lee, T., Hollenbach, J., Lu, K., Presbrey, J., Pru d'ommeaux, E. &schraefel, m. c.(2007) Tabulator Redux: Writing Into the Semantic Web. *Technical Report ECSIAMeprint14773, Electronics and Computer Science, University of Southampton.* (Unpublished)
- [15] <http://jquery.com/>



황준식

2007 고려대학교 사범대학
컴퓨터교육과(이학사)
2009 고려대학교
컴퓨터교육학과(이학석사)

2011~현재 고려대학교 컴퓨터교육학과
박사과정

관심분야: 컴퓨터교육, 소셜웹, 메쉬업

E-Mail: vollfeed@korea.ac.kr



유수진

2010 고려대학교
컴퓨터교육과(이학사)
2010~현재 고려대학교
컴퓨터교육과 석사과정

관심분야: 시맨틱 웹, 컴퓨터교육, 하이퍼미디어,
형식체계

E-Mail: mynameislydia@korea.ac.kr



김희진

2005 관동대학교
컴퓨터교육과(이학학사)
2008 고려대학교
컴퓨터교육과(이학석사)

2008~현재 고려대학교 컴퓨터교육과 박사과정

관심분야: 시맨틱 웹, 하이퍼미디어, 소셜 컴퓨팅,
컴퓨터교육

E-Mail: prin@korea..ac.kr



박성빈

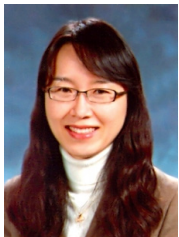
1990 고려대학교 전산학과
(이학사)
1993 University of Southern
California(전산학 석사)

1999 University of Southern California
(전산학 박사)

2006~현재 고려대학교 컴퓨터교육과 부교수

관심분야: 하이퍼텍스트, 컴퓨터과학교육, 알고리즘,
계산이론

E-Mail: psb@comedu.korea.ac.kr



정호숙

1998 서울교원대학교
컴퓨터교육과(교육학학사)
2001 서울교원대학교
컴퓨터교육과(교육학석사)

2010 고려대학교 컴퓨터교육과 이학 박사

관심분야: 컴퓨터교육, 시맨틱 웹, 소셜 네트워크

E-Mail: est0718@comedu.korea.ac.kr