

# 정보추출결과의 시각화 표현방법에 관한 이용성평가 연구\*

## A Usability Evaluation on the Visualization of Information Extraction Output

이 지 연(Jee-Yeon Lee)\*\*

### 목 차

- |                        |                     |
|------------------------|---------------------|
| 1. 서론                  | 2. 2 시각화 브라우저 환경 구현 |
| 2. 추출된 정보의 시각화 브라우저 이용 | 2. 3 시각화 브라우저 이용 연구 |
| 2. 1 정보추출과정            | 3. 결론               |

### 초 록

이 연구의 목적은 자동적으로 추출된 정보를 시각화 브라우저를 통해 표현하고 접근하는 방법에 대해 이용성평가를 수행하는 것이다. 특정 주제 분야에 국한되지 않는 정보추출시스템을 사용하여 신문기사에서 의미정보를 자동추출하였고, 그 결과로 시각적으로 브라우즈할 수 있는 지식베이스를 구축하였다. 이 연구에 이용된 정보추출시스템은 신문기사의 텍스트정보로부터 다양한 자연언어처리 기술을 적용하여 개념-관계-개념의 형태로 표현되는 정보를 자동적으로 추출하였다. 지식베이스에 저장된 정보를 시각적으로 표현하는 시각화 브라우저를 구현하였는데, 이때 인터페이스 부분은 쌍곡면형 정보시각화 시스템의 하나인 PersonalBrain을 이용하여 개발하였다. PersonalBrain을 이용하면 표현하고자 하는 정보를 논리관계의 네트워크 형태로 연결할 수 있다. 지식베이스 정보의 시각화 접근방법에 대하여 15명의 이용자를 대상으로 질적 평가를 수행하였다. 15명의 연구대상자로 하여금 시각화 브라우저를 이용하도록 하고 관찰과 인터뷰를 통하여 이용성평가 데이터를 수집하였다. 내용분석결과, 이용자들은 정보추출결과의 시각화 브라우저 이용가능성에 대해 긍정적인 반응을 보인 동시에 몇 가지 이용성 관련 문제점을 지적하였다.

### ABSTRACT

The goal of this research is to evaluate the usability of visually browsing the automatically extracted information. A domain-independent information extraction system was used to extract information from news type texts to populate the visually browsable knowledge base. The information extraction system automatically generated Concept-Relation-Concept triples by applying various Natural Language Processing techniques to the text portion of the news articles. To visualize the information stored in the knowledge base, we used PersonalBrain to develop a visualization portion of the user interface. PersonalBrain is a hyperbolic information visualization system, which enables the users to link information into a network of logical associations. To understand the usability of the visually browsable knowledge base, 15 test subjects were observed while they use the visual interface and also interviewed afterward. By applying a qualitative test data analysis method, a number of usability problems and further research directions were identified.

키워드: 정보시각화, 정보추출, 이용성평가, 질적평가

Information Visualization, Information Extraction, Usability Evaluation, Qualitative Evaluation

\* 이 논문은 2003년도 연세대학교 학술연구비의 지원에 의하여 이루어진 것임.

\*\* 연세대학교 문헌정보학과 조교수(jlee01@yonsei.ac.kr)

논문접수일자 2005년 5월 15일

게재확정일자 2005년 6월 8일

## 1. 서론

정보시각화 개념의 태동은 1970년대 이전으로 올라가는데, Kline(Bederson and Shneiderman 2003)은 이미 1970년 이전부터 사물이나 행위의 결과를 이해하고자 할 때 도움이 되는 효과적인 방법으로 시각적 방법을 강조하였다. Kline은 시각적 이미지(visual images)나 시각적 이해(visual understanding)가 여러 공식(a lot of formulas)에 의해 표현되고 설명되는 것보다 쉽고 효과적이고 도움이 되는 방법이라고 주장하였다. 시각정보에 대한 중요성은 기원을 Kline보다 훨씬 이전에서 찾을 수 있는데, 서양의 오래된 격언인 '한장의 그림이 천 개의 단어보다 가치가 있다'는 표현으로도 짐작이 가능하다. 그러나 시각적으로 표현되는 이미지정보에 대한 본격적인 연구가 시작된 것은 1980년대 말경부터였으며, Lynch(1991)는 1990년대 이후 이미지정보와 전자적인 이미지 처리기술의 도약을 이룰 것임을 전망하였다. 실제로 1990년대 이후 이미지정보, 시각정보에 대한 연구와 발전이 거듭하였고, 웹자원 혹은 멀티미디어정보가 익숙하고 보편적인 정보자원으로 취급되는 현재에 이르고 있다. 정보시각화도 시각정보에 대한 관심이 높아지고 다양한 시각정보가 데이터베이스화되어 전자적으로 검색이 가능해지면서, 인간-컴퓨터 상호작용, 사용자 인터페이스, 정보검색, 전자공학 분야에서 1990년대 중반 이후 활발히 연구되고 다양한 기법이 소개되었다. 국내에서도 1990년 후반부터 정보시각화에 대한 연구발표물이 나오기 시작하였고 2000년대에 들어서서 문헌정보학 분야에서 서은경(2002)이 정보시각화 모형을 유형별로 분석한

결과를 발표하였다. 정보시각화는 정보검색과 연결하여 이용자가 원하는 정보에 접근하기 위한 탐색과정에서 적용되기도 하고, 검색된 결과를 일련의 체계에 의해 시각적으로 재조직해 화면에 표현하는 부분에 이용되기도 한다. 정보시각화의 개념이 소개되고 다양한 기법이 개발, 분석된 후 최근 몇 년 동안에, 정보시각화 기법을 이용한 사례연구들이 전자공학, 건축분야, 주거환경분야, 전자상거래 분야, 자연과학 분야 등에서 제시되었는데, 아직 모두 시스템의 완성도에서 초기수준에 머무르는 정도이다. 상용화된 정보시각화 시스템 중 다수가 아직 정보시각화 기능의 기본적인 틀만을 제공하고 있으며, 이보다 향상된 기능을 제공하는 상용 프로그램이라 할 수 있는 것도 Iokio사의 Omniscope나 Inxight사의 vizServer 정도로 소수에 지나지 않는다. 더구나 아직도 많은 연구나 개발이 기법이나 기능의 양적 추가에 집중되어 있고, 상호작용 측면이나 이용성을 강조하는 사례는 극히 드문 실정이다.

이 연구는 정보시각화에 대한 관심의 증가와 응용분야의 확대, 그리고 폭발적으로 증가하는 정보자원을 처리, 제공할 수 있는 가능성에 주목하여, 정보시각화 기법을 적용하는 사례를 통하여 질적인 이용성 평가연구를 실시하였다. 정보시각화 기법을 적용하는 사례로서 이 연구는 추출된 정보를 시각적으로 표현해서 이용자가 브라우저를 통하여 원하는 정보에 접근하는 시각화 브라우저를 제안하고, 그 이용가능성을 평가하였다. 정보시각화 기법의 가장 기본적인 목적은 추상적인 데이터(abstract data)를 체계적으로 조직하여 이용자에게 제공하는 것이다. 따라서 이 연구에서는 추상적인 데이터를 정보

추출시스템을 이용하여 체계화시켜 지식베이스를 생성하고, 지식베이스의 정보를 시각적으로 표현, 접근할 수 있는 시각화 브라우저 시스템을 구현하였다. 시각화 브라우저 시스템은 정보시각화 기법 중 쌍곡면형 브라우징 기법을 적용한 비주얼라이저를 포함하고 있다. 이외에 정보추출결과 생성된 개념과 관계들을 시각적으로 표현하는 개념 비주얼라이저와 개념 비주얼라이저의 개념들을 열거한 개념리스트 및 검색된 문장이나 서지사항을 보여주는 뷰어 부분이 시각화 브라우저 시스템을 구성하고 있다.

15명의 이용자를 연구대상자로 선정하여, 구현된 시각화 브라우저를 이용하여 지식베이스에 저장된 북한문제와 관련된 정치, 경제, 외교 분야 영문 신문기사정보 중 이용자가 관심을 가지고 있는 주제에 대해 탐색을 하도록 하였다. 정보추출부터 이용성평가에 이르는 연구의 과정은 다음과 같다.

## 2. 추출된 정보의 시각화 브라우저 이용

이 연구는 자동적으로 추출된 정보를 이용자가 이용하는 방법 중 하나로, 개념-관계-개념(Concept-Relation-Concept)으로 표현되는 네트워크 형태의 시각화 브라우징 방법에 대한 이용가능성을 평가하고자 하는 목적에서 출발하였다. 테스트 데이터를 수집한 후, 일반적인 유형의 정보추출시스템(Paik 2000)을 이용하여 테스트 데이터에서 정보를 추출하였다. 정보추출 과정에서 생성된 개념간의 관계를 토대로 하여

지식베이스가 구축되었다. 지식베이스에 저장된 데이터를 시각적으로 표현하여 이용자가 원하는 정보에 쉽게 접근하도록, 정보시각화기법 중 하나인 쌍곡면형 브라우징 기법을 적용하였다. 쌍곡면형 브라우징은 정보의 표현방법에 있어서 개념들을 직접 또는 간접으로 이어주는 연결선에 의하여 관계를 표현하는 방법으로 (Allen 2002), 개념-관계-개념의 형태로 추출된 정보를 시각적으로 표현하기에 적합한 방법이다. 이용자들은 개념-관계-개념의 네트워크 형태로 제공되는 시각화 브라우저를 통해서 정보를 빠르게 파악할 수 있으며, 자신의 정보요구에 적합한 부분에 집중할 수 있다.

### 2.1 정보추출과정

#### 2.1.1 데이터 준비과정

정보추출결과를 시각화하여 보여주기 위하여 평가에 이용될 데이터를 수집하는 과정이 있었다. 데이터는 북한과 관련된 정치, 경제, 외교 분야 신문 기사를 수집하였는데, 구글([www.google.com](http://www.google.com))이 주로 이용되었다. 수집된 데이터는 연구에 이용된 정보추출시스템(Paik 2000)과 정보시각화시스템인 PersoanlBrain이 처리할 수 있는 영문으로 쓰여진 신문 기사를 대상으로 하였다. 또한 일반적으로 잘 알려진 뉴스기관인 Associated Press나 New York Times, 혹은 Washington Post와 같은 곳에서 작성된 기사를 선정하여 이용성평가 자료로 이용하였다. 시각화 브라우저 평가를 위하여 60개의 뉴스 기사를 선정하였고, 이 기사들을 정보추출시스템으로 처리하여 지식베이스를 생성하였다.

### 2. 1. 2 정보추출방법

수집한 테스트 데이터에서 정보를 추출하기 위하여 Paik(2000)의 연구에서 구현된 정보추출시스템을 이용하였다. 이 정보추출시스템은 특정분야를 대상으로 구현된 시스템이 아니라 일반적으로 여러 분야에서 이용할 수 있는(domain-independent) 시스템으로, 자동적으로 해당주제분야의 지식베이스를 구축해 준다. 이 연구에서는 뉴스기사에서 정보를 자동적으로 추출하였으며, 이 시스템을 이용하여 법률정보를 성공적으로 추출한 연구도 발표된 바 있다(Paik & Lee 2004). 이 연구에서 이용한 정보추출시스템은 새로운 텍스트 데이터가 입력되었을 때, 데이터를 수작업이나 시간이 소요되는 사전의 수정(modification)작업 없이 신속하게 데이터를 처리해서 결과를 제공한다. 정보추출 결과로 지식베이스에 입력되는 새로운 지식은 새로 추출된 이름(names), 사건(events), 개념(concepts), 관계(relations) 등이다. 이 연구에서 제시하는 시각화 브라우저의 평가를 위하여, 정보추출시스템을 이용하여 뉴스기사에서 개념-관계-개념 형식의 표현들을 추출하여 지식베이스를 구축하였다.

정보추출시스템은 텍스트 내에서 하나의 개념과 다른 개념을 이어주는 관계를 추출하였는데 이때 개념간의 인접성(close proximity)을 이용하여 테스트 데이터에서 정보를 추출하고 조직하였다. 정보추출시스템은 또한 시간적인 순서대로 테스트 데이터를 조직하는 기능을 가졌는데, 예를 들어 George Bush와 같은 사람을 지칭하는 고유명사에 대한 연대순 프로파일 을 만드는 것이 하나의 예이다.

정보추출시스템은 여러 개의 모듈로 구성되

었으며, 다양한 주제 분야에 속하는 텍스트를 컴퓨터 처리가 가능한 지식표현으로 변환할 수 있다. 정보추출시스템을 구성하는 모듈은 다음과 같다. 의미추출모듈(semantic extraction module)은 두 개념간의 의미관계를 지정하기 위하여 사전처리모듈(preprocessing module)의 결과를 이용한다. 문장영역파악모듈(sentence boundary identification module)은 구두점을 이용하여 각 문장의 시작과 끝을 결정한다. 품사태깅모듈(part-of-speech tagging module)은 각 단어에 적절한 품사태그를 지정한다. Pennsylvania 대학에서 개발된 품사태그리스트(Santorini 1990)에 근거하는 확률적 품사태그를 사용하였다. 형태소분석모듈(morphological analysis module)은 단어의 기본형을 결정한다. 예를 들어, 과거동사인 'told'의 현재형인 'tell'로 표준화시키는 경우에 해당한다. 복수단어로 된 개념파악모듈(multiword concept identification module)은 복수단어로 이루어진 개념의 시작과 끝을 찾는다. 예를 들자면, 'North Korea'는 하나의 고유명사개념으로 처리하며, 'nuclear material'과 같은 명사구는 복수단어로 된 복합개념으로 처리한다.

개념의 범주화와 표준화 모듈(concept categorization and normalization module)은 고유명사와 숫자를 미리 만들어진 개념온톨로지에 따라 의미유형별로 분류한다. 온톨로지는 64개의 복수계층 의미유형(multi-level semantic type information)으로 구성되었다. 'Bush'는 인명에 포함되며 이 모듈 안에 있는 조희용 데이터베이스를 통해 'George Bush'로 표준화된다. 날짜와 같은 숫자 개념도 분류되고 표준화되는데, 예를 들어 'early this year'라는 구가

있을 때 날짜는 2005년 1월 1일부터 2005년 3월 31일과 같은 기간으로 표준화된다. 날짜의 표준화를 위해서 출처신문기사의 보도일자와 조회 데이터베이스가 이용된다. 이 모듈의 처리 결과는 쌍곡면형 비주얼라이저의 콘텐츠를 구성한다. 정보추출결과로 구축된 지식베이스를 이용하기 위해서 이용자는 쌍곡면형 비주얼라이저를 보면서 원하는 개념을 찾는다. 이때 쌍곡면형 비주얼라이저의 구조는 고유명사를 범주화한 온톨로지와 동일하게 된다. 고유명사 범주의 말단노드가 실제 인명이 되는 것도 이 때문이다.

의미관계는 개념들이 어떻게 연결되는가를 정의하고 있다. Washington Post의 2005년 3월 20일자 기사 내용 중 'Pyongyang had exported nuclear material to Libya'의 문장의 경우 'seller' 의미관계가 도시명인 'Pyongyang'과 동사개념인 'exported' 사이에 존재한다. 그리고 'buyer' 의미관계가 국가명인 'Libya'와 'exported' 사이에 존재한다. 마지막으로 'object'라는 관계가 'nuclear materials'와 'exported' 사이에 생기게 된다.

대부분의 의미관계가 양쪽에 관련개념을 연결하는 구조로 되어 있으며, 개념-관계-개념의 형태(triples)를 이룬다. 하지만 'necessary,' 'negative,' 'past,' 'possibility'와 같은 의미관계는 하나의 개념에 연결되어 개념-관계의 형태를 이룬다. 의미관계는 Conceptual Dependency Theory(Schank 1972)와 Conceptual Graph Theory(Sowa 1984)의 관계 사례를 참조, 채택하여 사용하였다. 이 연구에서 사용된 정보추출시스템은 'accompany,' 'affiliation,' 'agent,' 'argument,' 'association,' 'cause' 등의 37개

의미관계를 추출하였다.

개념-관계-개념 형태의 정보를 추출하기 위하여 몇 가지 알고리즘이 적용되었다. 공통참조 개념추출 알고리즘(Co-referential concept extraction algorithm)은 주어개념(subject concept)과 동격 사이의 관계를 파악하는 방법이다. 관계표시공식(Relation revealing formula)은 주개념과 관계사절 구성요소 사이의 관계를 밝혀주는 방법이다. 특정의미관계 알고리즘(Special semantic relation algorithm)은 전치사, 구두점, 명사구, 동사구 등으로 구성된 특정유형의 개념과 단서를 이용하여 의미관계를 추출한다. 마지막으로, 구문관계에서 의미관계로의 매핑 알고리즘(Syntactic relation to semantic relation mapping algorithm)에서는 타동사의 주어와 같은 구문관계가 행위의 주체(agent)라는 의미로 변환될 수 있다는 개념과 이와 유사한 규칙들을 이용하여 의미관계를 추출한다. 이것은 지금까지 많은 정보추출시스템들이 기본적으로 사용해 온 방법이다.

## 2. 2 시각화 브라우저 환경 구현

추출된 정보를 시각적으로 표현하기 위한 시각화 브라우저는 The Brain Technologies Corp (<http://www.thebrain.com/>)사의 Personal-Brain이라는 상용 시스템을 비주얼라이저에 부분적으로 이용하여 구현되었다. 완성도 있는 상용 소프트웨어를 이용함으로써, 연구에 이용될 시각화 브라우저가 질적으로 우수한 인터페이스를 제공할 수 있는 이점을 고려하였다. 이미 상용화된 시각화 시스템을 이용함으로써, 이 연구에서 구현하는 시각화 브라우저가 일반 인터

페이스 상의 이용성 문제를 다루기보다는 정보 추출결과의 시각화를 평가하는데 초점을 맞출 수 있었다.

연구에서 이용된 PersonalBrain은 정보시각화 시스템으로 PersonalBrain 이용자는 정보단위를 서로 연결시켜 논리적인 관계로 형성된 네트워크를 생성할 수 있다. PersonalBrain에서 사용하는 용어 중에서 'Thoughts'는 이 연구에서 말하는 개념에 해당하며 'Links'는 관계를 의미한다. PersonalBrain은 이용자가 생각하는 방식으로 정보를 구조화하여 표현할 수 있도록 지원한다는 점에서 이 연구에 이용하기에 적합한 것으로 판단되었다. PersonalBrain은 이 연구에서 제안하는 시각화 브라우저의 모든 비주얼라이저를 구현할 수 있는 기능을 지원하였다.

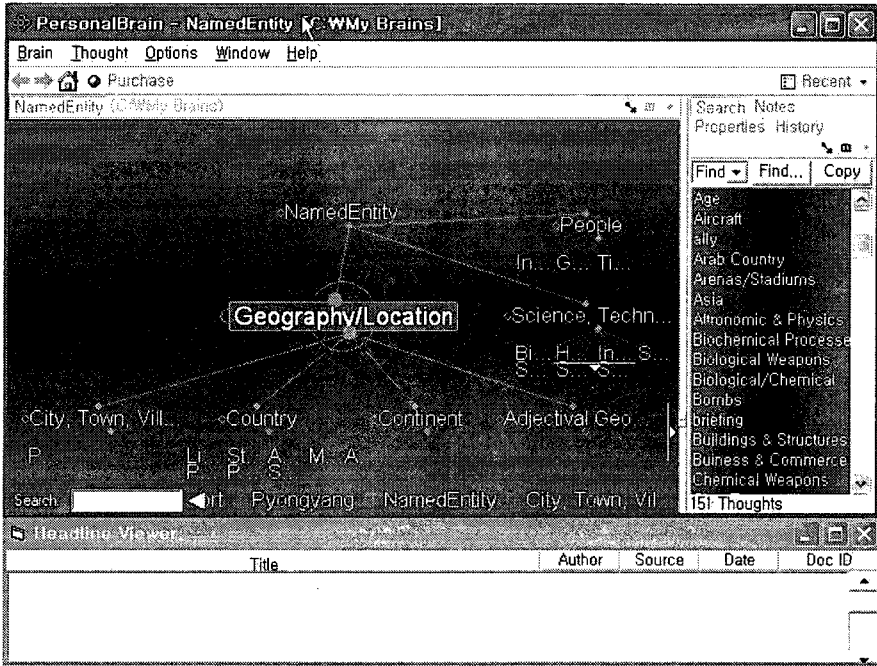
시각화 브라우저는 크게 네 부분으로 구성되어 있으며, 각각은 쌍곡면형 비주얼라이저, 개념 비주얼라이저(Concept Visualizer), 개념 열거표(Concept List), 기사내용뷰어(Headline Viewer)이다. 그림 1은 시각화 브라우저의 초기화면으로 위의 네 부분 중 세 부분을 보여주고 있다. 그림에서 보면 왼쪽 상단에 위치하는 것이 쌍곡면형 비주얼라이저이고, 그 옆의 오른쪽 상단이 개념들을 열거해서 보여주는 개념 리스트이다. 기사내용뷰어는 시각화 브라우저 화면의 하단에 위치하며 적합 문장의 서지사항을 제공한다.

첫 부분인 쌍곡면형 비주얼라이저는 이용자가 원하는 개념이나 주제에 빠르고 쉽게 접근할 수 있도록 주제 카테고리들을 트리 형식으로 표현하고 있다. 테스트 데이터에서 추출된 모든 고유명사를 계층구조로 제시하는데 <그림 1>은 Geography/Location을 선택한 상태이다. 화면

은 Geography/Location의 하위개념으로 City, Town, Village; Country; Continent; Adjectival Geographical Names가 연결되고 있음을 보여주고 있다. 이러한 하위개념에는 실제 지역명을 나타내는 평양 등의 고유명사가 보여지게 된다. Geography/Location의 자매노드인 People이나 Science, Technology, industry도 화면에서 같이 보여주고 있다. 개념 비주얼라이저는 정보추출의 결과인 개념-관계-개념의 세 부분으로 표현되는 항목들을 제시하는 부분으로 쌍곡면형 비주얼라이저를 통해 선택된 범주가 있을 때 이에 해당하는 개념과 관계를 표시하는 개념 비주얼라이저가 쌍곡면형 비주얼라이저 부분에 나타난다. 이용자는 기사내용뷰어의 정보를 통해서 적합한 뉴스기사, 더 나아가서 적합한 문장에 접근할 수 있다. 시각화 브라우저의 각 부분에 대한 자세한 설명은 다음과 같다.

### 2. 2. 1 쌍곡면형 비주얼라이저

이 연구에서 구현된 시각화 브라우저는 정보 시각화 기법 중에서 쌍곡면형 브라우징 방법을 이용하여 데이터베이스 내의 뉴스기사 및 문장으로 접근하도록 하는데, 이 때의 접근점은 트리구조로 구성된 고유명사이다. 쌍곡면형 브라우징은 알려진 바와 같이, 가장 중심이 되는 뿌리노드(root node)에서 시작하여 그 하위에 카테고리들이 연결되어지고 하위 카테고리에서 또 연결되는 세부 하위개념들이 존재하는 계층형 시각화 표현방법이다. 따라서 뿌리노드에 가까울수록 데이터베이스 내에서 상위개념에 위치한다. 그리고 비록 비주얼라이저에서는 이용자가 관심을 가지고 선택한 부분만을 한정적으로 보여준다 할지라도, 이미 전체 개념과 그 계



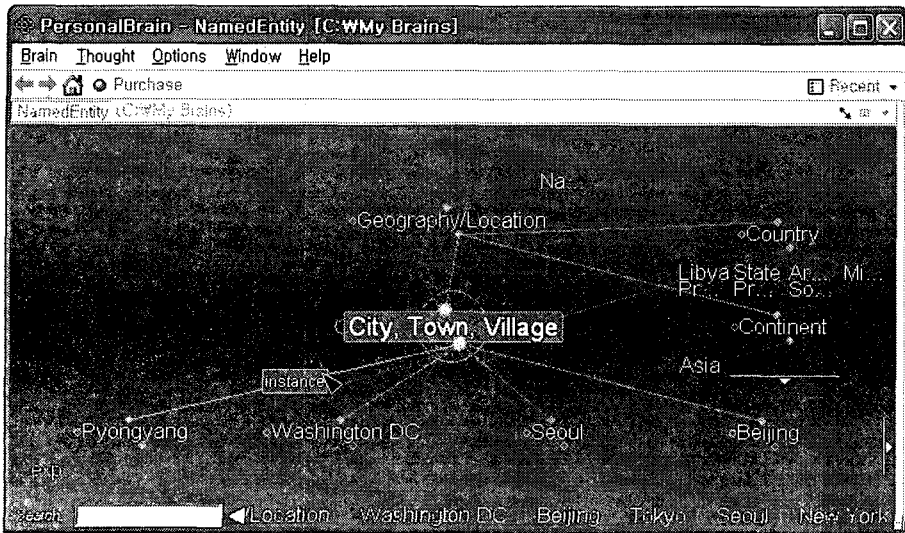
〈그림 1〉 시각화 브라우저

위는 이미 정해져 있는 상태이다. 이 연구에서의 개념과 계위에 따르면, 말단 노드(terminal node)는 가변적인데 그 이유는 접근점이 고유 대명사들의 표현단계에 따라 다르게 결정되기 때문이다.

쌍곡면형 브라우저는 데이터베이스 내의 다량의 데이터를 일련의 체계에 따라 용이하게 볼 수 있도록 지원한다. 이용자는 관심이 있는 개념을 마우스를 이용하여 포인터를 끌어서(dragging) 스크린을 움직이고 관심이 있는 개념과 연결된 개념들을 보게 되거나, 관심이 있는 개념을 클릭하여 선택하고 이 때 쌍곡면형 브라우저는 선택된 개념을 중앙으로 이동시키고 관련 개념들을 주변에 배치시키게 된다. 이 연구에서는 후자의 방법을 이용하였다.

〈그림 2〉는 이 연구의 이용자가 쌍곡면형 비

주얼라이저를 이용한 예를 보여주고 있다. 이 그림은 〈그림 1〉의 초기화면에서 다른 기능을 제외하고 쌍곡면형 비주얼라이저만 따로 보여주고 있는데 고유명사 트리 중 'Geography/Location'의 하위개념 중에서 'City, Town, Village'를 선택했을 때의 결과이다. 'City, Town, Village' 아래의 다른 하위개념들은 'Pyongyang,' 'Washington DC,' 'Seoul,' 'Beijing' 등이다. 이용자가 화면의 노드를 선택하거나 말단노드에 도착하게 되면 그 해당노드의 개념을 포함하는 개념-관계-개념을 보여주는 개념 비주얼라이저 내용이 형성되어 개념 비주얼라이저로 이동할 수 있다. 화면 하단부의 탐색어 입력필드는 노드로 표현된 개념의 계층트리를 브라우즈하는 대신 키워드를 직접 입력해서 원하는 노드에 접근할 수 있도록 해준다.



<그림 2> 쌍곡면형 비주얼라이저

### 2. 2. 2 개념 비주얼라이저

위에서 언급하였듯이, 쌍곡면형 비주얼라이저를 이용하여 이용자가 말단노드에 도착하고 하나의 노드를 선택하게 되면, 개념 비주얼라이저는 쌍곡면형 비주얼라이저에서 선택된 노드를 포함하는 개념-관계-개념의 항목들을 제시하게 된다. 개념들은 서로 직선으로 연결되어 표시되고, 이때 관계는 개념들을 연결하는 직선상에 존재하게 된다. 개념과 관계는 색채에 의해서 시각적으로 쉽고 빠르게 구별되도록 하였다. 그림 3은 개념 비주얼라이저의 예시 화면이다.

개념 비주얼라이저에서 보여주는 내용은 Washington Post에 실린 내용으로 다음의 문장에서 추출한 정보를 개념과 관계로 표현하고 있다.

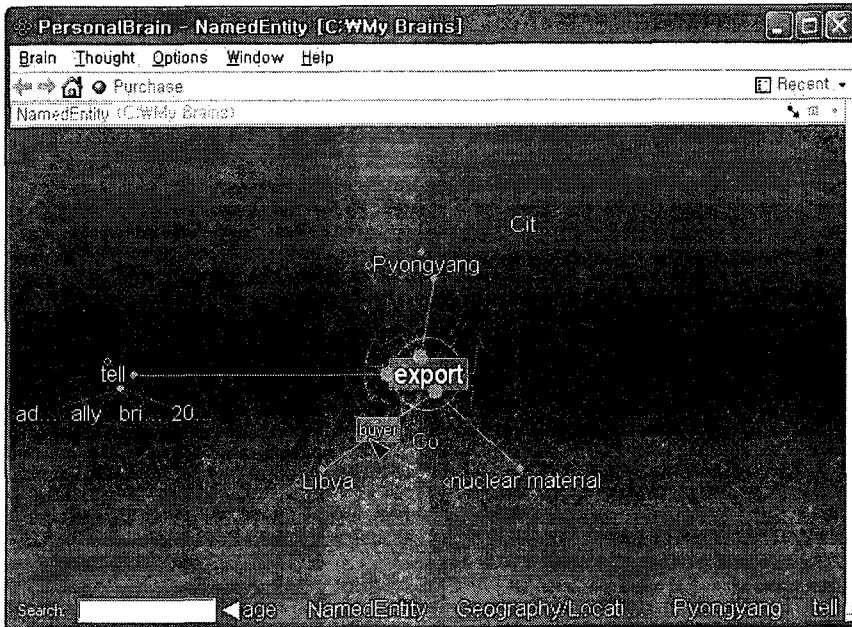
*In an effort to increase pressure on North Korea, the Bush administration told its Asian allies in briefings earlier this year that Pyongyang had exported nuclear ma-*

*terial to Libya (by Dafna Linzer on March 20, 2005).*

위의 문장과 그림 3의 화면내용을 비교해 보면 알 수 있듯이 모든 정보가 다 추출되어 개념 비주얼라이저에 보여지는 것은 아니다. 화면내용은 'Pyongyang had exported nuclear material to Libya'라는 부분에서만 정보를 추출해서 보여주고 있다. 다른 추출된 정보는 옆에 연결된 'tell'과 같은 개념을 더블클릭하면 볼 수 있다.

개념 비주얼라이저는 개념-관계-개념의 네트워크를 표현하고 있을 뿐만 아니라 이 네트워크를 이용하는 몇 가지 기능들을 같이 제공하고 있다. 첫째, 쌍곡면형 비주얼라이저와 마찬가지로 키워드탐색 기능은 이용자들이 비주얼라이저 이용을 생략하고 바로 관심대상인 개념을 입력할 수 있도록 하는 방법이다. 둘째, 앞에서 언급하였듯이 뉴스기사의 같은 부분에서 추출, 형성된 개념-관계-개념을 한 화면에서 처리할 때





<그림 3> 개념 비주얼라이저

그 양을 조절하고 더블클릭하기를 이용하여 생략된 부분을 볼 수 있도록 하였다. 하지만 많은 수의 개념-관계-개념이 형성된 경우, 특히 수백 개에 달하는 개념과 관계가 생기는 경우도 있었는데, 이용자가 탐색에 어려움을 겪는 대표적인 예가 되었다. 개념 비주얼라이저에서 개념과 관계가 다수 형성되었을 때 로드시간이 길어지게 되고, 상대적으로 생략되는 개념은 많아지며 축약된 이름으로 표현되는 개념도 많아지게 된다. 이에 대한 이용자의 반응 및 평가는 2.3.3 이용자 평가 부분에서 자세히 언급하였다.

### 2. 2. 3 기사내용뷰어

시각화 브라우저의 하단부에 위치하는 기사내용뷰어(Headline Viewer)는 이용자들로 하여금 개념 비주얼라이저에 표시된 개념-관계-개념을 포함하는 문장이나 기사를 볼 수 있도록

한다. 이용자는 기사내용뷰어를 이용하여 검색된 적합한 문장이나 기사의 전문을 선택하여 볼 수 있다.

### 2. 3 시각화 브라우저 이용 연구

이용자들이 추출된 정보에 접근하는 방식으로 이 연구는 2.2에서 설명한 시각화 브라우저를 이용하는 방법을 제안하고 이에 대한 이용성을 평가하고자 하였다. 따라서 이용자들이 작업을 수행하거나 정보추구를 하는 과정에 있어서 쌍곡면형 비주얼라이저나 개념 비주얼라이저를 이용하는 것이 원하는 정보에 접근함에 있어서 얼마나 효과가 있으며, 이용자의 목표달성에 어떤 역할을 수행하는지에 대하여 조사하는 것이 이 연구의 핵심이라 하겠다. 이 연구는 세 부류의 이용자를 대상으로 시각화 브라우저 이용 중

의 의견표현(verbal protocols)과 이용 후 개별 인터뷰를 통해 이용자 평가에 대한 자료를 수집하였다. 이 연구는 질적 평가를 수행하도록 설계되었으며, 수집된 내용도 질적 연구에서 수행하는 내용분석방법에 의하여 분석, 정리되었다.

2. 3. 1 연구대상자 선정

이용자로 선정된 연구대상자는 모두 15명으로, 연구에서 제시된 시각화 브라우저를 통해 자신들의 정보요구에 적합한 문서를 찾는 작업을 수행하였다. 연구에 참여한 15명의 선정은 판단샘플링 방식을 통하였고, 영문 기사를 검색하고 이해할 수 있는 수준의 대상자를 선정하였다. 15명의 연구대상자는 이용자의 유형별로 나누어 본다면, 세 가지 유형으로 분리될 수 있는

데, 한 부류는 '숙련된 이용자(Power Users)'이고, 다른 한 부류는 '정보전문가(Information Professionals)'이며, 나머지 한 부류는 '비즈니스 이용자(Business Users)'이다. 이 세 그룹은 손쉽게 빠르게 원하는 정보에 접근하고 필요한 정보를 공유하는 능력을 가지고 있다는 공통점이 있는 반면에, 그룹이 추구하는 정보의 유형과 이유는 각기 다르다는 특징을 가지고 있다. 연구대상자의 정보이용능력과 지식을 고려하고, 제공되는 시각화 브라우저에 대한 다양한 의견과 평가가 이루어지도록 연구대상자를 선정하였다. <표 1>은 연구대상자의 특징분석표이다.

<표 1>에서 보여주는 연구대상자 개인별 프로파일을 종합해 볼 때, 연구대상자는 유형별로 다음과 같은 특징을 나타내고 있음을 알 수 있다.

<표 1> 연구대상자 프로파일

연구대상자 (분류*)	특징		
	학문 및 지식배경	주된 이용정보	정보이용패턴
1 (PU)	문헌정보, 어학	학술정보	브라우저 + 정보포탈 > 키워드탐색
2 (PU)	문헌정보, 프로그래밍	컴퓨터업계뉴스	웹(상당량의 즐겨찾기항목 보유)
3 (PU)	문헌정보	학술정보	웹, 데이터베이스
4 (PU)	문헌정보, 심리학	학술, 미디어	데이터베이스, 정보포탈
5 (PU)	문헌정보, 어학	학술, 교양	웹, 데이터베이스
6 (IP)	전산	컴퓨터관련기술	웹검색엔진, 키워드탐색
7 (IP)	커뮤니케이션	학술정보	대학도서관시스템, 웹
8 (IP)	경제	학술정보	데이터베이스
9 (IP)	문헌정보	학술정보	웹, 데이터베이스
10 (IP)	물리학	최근연구	웹, 동료, 기타전자자원
11 (BU)	마케팅	시장동향	키워드탐색 선호
12 (BU)	경영	광고물, 미디어	인쇄물을 포함한 각종 광고매체
13 (BU)	어학	정보업계동향	정보포탈, 업무보고서
14 (BU)	정치학, 경제학	투자정보	데이터베이스
15 (BU)	문헌정보	IT분야 마케팅	웹

\* PU = Power User, IP = Information Professional, BU = Business User

숙련된 이용자(Power users): 이 그룹의 이용자는 정보조직, 검색 엔진, 데이터베이스, 검색어 선정 등에 대한 지식을 바탕으로 정보처리나 정보기술의 전문지식과 능력을 지닌 집단이다. 이들은 보통 여러 데이터베이스를 복합적으로 이용하며, 고급탐색능력을 지니고 있다. 이 집단이 관심을 가지는 정보는 디지털 도서관 구축, 정보검색기술의 발전, 검색결과의 분석 및 이해, 정보시각화 기법 등이다.

정보전문가(Information Professionals): 정보전문가들은 교육기관에서 교육, 연구를 행하는 집단으로, 이들은 숙련된 이용자들이 보이는 특성을 지닌 동시에, 정보 배포에 관심을 가지고 있다. 정보전문가들이 주된 작업을 열거한다면, 1) 정보관리, 검색을 위한 최근발전동향조사, 2) 수집된 최신정보의 조직, 관리, 이용, 유지, 3) 정보의 분석, 종합, 합성, 배포 등이다.

비즈니스 이용자(Business Users): 비즈니스 이용자는 자신이 소속된 기관의 내·외부에서 필요한 정보를 수집하고자 한다. 이들이 관심을 두고 있는 정보의 유형으로는 1) 현재 진행하고 있는 업무와 관련된 정보, 예를 들면 연수, 훈련, 이전의 사업 계획서, 재정분석자료, 2) 관련업계의 시장정보, 예를 들면 소속회사, 고객, 업무향상 등과 관련된 추세정보 및 뉴스 등이다. 비즈니스 이용자들은 투자, 판매전략, 시장성 파악, 무역, 컴퓨터 관련업 등 다양한 분야에 속한 집단이다.

### 2. 3. 2 관찰 및 인터뷰

15명의 이용자는 각각 자신의 정보요구를 가지고 시각화 브라우저를 이용하여 정보를 검색하였다. 검색에 앞서 각 이용자는 시각화 브라우저의 이용법과 설명에 대한 개별지도를 받았다. 이용법과 더불어 시각화 브라우저 설명은 정보추출과정과 그 결과 생성된 개념과 관계, 그리고 데이터베이스 생성과정과 방법 및 결과 표현 등을 포함하였다. 개별지도를 마친 후, 이용자는 자신의 정보요구에 맞는 정보를 탐색하도록 요구되었는데, 이 때 시각화 브라우저를 이용하였다. 탐색주제를 인위적으로 주기보다는 이용자 자신에게 관심이 있는 탐색주제를 선정하도록 함으로써, 탐색 환경에 있어 최대한 자연적인(Naturalistic) 설정이 되도록 설계하였다. 단, 연구자는 이용자가 자신의 탐색주제를 선정함에 있어서 사용하는 데이터베이스의 성격에 맞는 주제를 이용하도록 방향을 잡아주는 역할을 하였다.

이용자들은 시각화 브라우저를 이용하는 동안 소리내어 생각하기(think-aloud)를 함으로써 탐색과 관련된 사고과정을 설명하였다. 예를 들어, 시각화 브라우저를 사용함에 있어서 탐색에 도움이 되는 부분, 혼돈스럽거나 문제를 일으키는 부분, 수정·보완이 필요한 부분 등을 말로 표현하였다. 다음은 이용자에게 설명하도록 요청하였던 일반적인 사항들이다.

- 시각화 브라우저에 대한 전반적인 느낌은 무엇인가?
- 시각화 브라우저가 구성된 방식이 이해하고 사용하기 쉬웠는가?
- 시각화 브라우저의 특징이나 기능 중 이용

하기 쉬웠던 특징과 특별히 이용이 어려웠던 특징은 무엇인가? 시각화 브라우저가 개념 및 정보를 표현하는 방식에 대해 어떻게 생각하는가? 시각화 브라우저를 이용한 정보표현이 이용자들의 정보추구과정에 도움이 되었는가? 도움이 되었다면 어떤 방식으로 도움이 되었는가?

- 시각화 브라우저 시스템으로부터 검색된 정보는 이용자의 정보요구에 얼마나 적합한 결과를 제공하는가?
- 시각화 브라우저와 이용자가 최근에 이용한 다른 시스템과 구조상의 차이점은 무엇인가?
- 시각화 브라우저는 이용자의 정보추구과정에 적합한 방식인가?
- 시각화 브라우저를 이용하여 탐색할 수 있는 정보의 유형은 어떤 것들인가?

Marchionini(1987)은 정보추구자들이 브라우징 시스템을 이용하는 이유로 다음의 세 가지를 설명하고 있다. 첫째, 브라우징은 시스템 구조면에서 그리고 명령체계나 탐색진행과정면에서 이용하기 편리하고, 둘째, 탐색목적이나 탐색주제가 명확하게 파악되지 않은 상황에서도 브라우징은 사용될 수 있으며, 셋째, 세밀하게 탐색계획을 세우고 분석을 해야 하는 키워드탐색에 비해 브라우징은 이용자로서 하여금 이에 대한 부담을 덜 느끼도록 한다는 점이다. 실제로 이 연구에 참여한 15명의 이용자들은 시각화 브라우저를 이용하여 탐색을 진행하는 동안에 Marchionini의 첫 번째와 두 번째 이유에 해당되는 의견을 자주 언급하였다. 이 연구에서 Marchionini가 지적한 세 번째 이유인 탐색에 관련된 인지적 부담감에 대한

언급이 적었던 이유는 연구의 성격이 시각화 브라우저와 키워드탐색을 비교하기보다는 시각화 브라우저의 사용에 집중되었기 때문으로 보인다. 실제로 저자가 행했던 이전의 연구(이지연 2001)에서는 이미지검색에 있어 키워드탐색보다 비주얼 브라우징을 선호한 경우가 많이 있었으며 연구대상자들은 그 이유로 키워드선정의 어려움이나 부담을 지적하였다.

다음 장은 이용자들이 시각화 브라우저를 평가한 내용을 분석한 부분이다. 이 때 질적평가는 내용분석을 통하여 이루어졌으며, 내용분석에 있어서 Marchionini의 첫 번째, 두 번째 항목을 분석의 초기틀로 이용하였다.

### 2. 3. 3 이용자 평가

이용자 평가는 이용자들의 탐색행위에 대한 연구자의 관찰 및 이용자들이 소리내어 생각하기와 인터뷰를 통하여 언급한 시각화 브라우저에 대한 의견을 종합하여 분석하였는데, <표 2>는 이용자 평가 분석의 전체적인 틀을 보여주고 있다.

<표 2>에서 제시하고 있는 분석의 범주는 크게 세 부분으로 나눌 수 있는데, 첫 번째는 시각화 브라우저의 구조적인 측면을 다루고 있고, 두 번째는 이용자의 탐색목적에 관련된 것이며, 마지막 부분은 시스템의 기타 기능적 특징에 대한 것이다. 다음은 표 2에 언급된 평가내용의 구체적 설명이다.

#### 2. 3. 3. 1 시각화 브라우저 시스템의 구조적 특성 평가

15명의 이용자 모두는 사용할 시각화 브라우저가 뉴스기사에서 개념과 개념간의 관계를 추

〈표 2〉 이용자 평가내용의 항목별 분석

범주	내용
1. 시각화 브라우저의 구조적 특성 평가	<ul style="list-style-type: none"> <li>• 시각적 표현방식에 대한 평가</li> <li>• 데이터베이스 이해에 대한 평가</li> <li>• 표현되는 정보량에 대한 평가</li> </ul>
2. 탐색목적관련 평가	<ul style="list-style-type: none"> <li>• 탐색목적의 명확성과 관련된 평가</li> <li>• 학습효과에 대한 평가</li> <li>• 탐색 시 문제점 파악 및 대처방식에 대한 평가</li> <li>• 능동적이고 적극적인 탐색태도에 대한 평가</li> </ul>
3. 기타 기능관련 평가	<ul style="list-style-type: none"> <li>• 개념간의 조직에 대한 평가</li> <li>• 적합한 개념 발견 가능성에 대한 평가</li> <li>• 정보접근의 용이성에 대한 평가</li> <li>• 이용자별 맞춤성에 대한 평가</li> <li>• 개념과 범주 조절기능, 관련성 추가기능 등 기능성에 대한 평가</li> </ul>

출해서, 기존의 메뉴 등의 형식을 이용하는 것이 아니라 시각적으로 보여준다는 아이디어에 상당한 관심을 보이며 흥미로워하였다. 시각화 브라우저를 표현할 수 있는 유사한 정보원에 대하여 이용자들은 시소러스, 사전, 색인, 백과사전 등을 언급하였다. 이러한 정보원들은 개념들을 일련의 체계에 의해 집합시켜 놓았다는 점에서 시각화 브라우저와 유사하다는 의견이었다.

이용자들은 시각화 브라우저의 쌍곡면형 비주얼라이저 부분이 개념과 하위개념 간의 연결들을 통해서 전체 데이터베이스의 크기를 짐작할 수 있고, 각 개념에 해당하는 정보의 폭이 얼마나 광범위한지 알 수 있는 구조를 이룬다고 평가하였다. 그러나 이용자들은 쌍곡면형 비주얼라이저에서 개념 비주얼라이저로 이동해서 그 안에서 개념-관계-개념을 통한 브라우징을 하고 나면, 쌍곡면형 브라우저를 거의 사용하지 않고 개념 비주얼라이저에서 개념과 관계들을 계속 집중해서 탐색하는 경향을 나타냈다. 그리고 개념 비주얼라이저에 너무 많은 수의 관련 개념들이 제시되는 경우에 이용자들은 제공되

는 정보량에 압도당하며 자신의 탐색을 종료하기 시작했다. 이용자들이 보여준 이러한 탐색행태는 쌍곡면형 비주얼라이저, 그리고 특히 개념 비주얼라이저의 경우에, 대규모의 데이터베이스에서 개념들을 표현할 때 스크린 크기에 맞추어 체계적으로 조직해서 보여줄 수 있는 범위에 한계가 있다는 점을 시사하고 있다.

개념 과부하를 경험한 이용자들은 개념들이 어떻게 군집화되고 표기되었는지를 살피고자 하였다. 대부분의 이용자들이 개념 비주얼라이저에서 500개 이상의 개념들이 제시되는 상황을 경험하였는데, 이런 경우 이용자들은 차라리 기사내용뷰어에서 기사제목들을 차례로 읽어 나가려고 하였다. 이용자들은 “be,” “he,” “saw,” “several,” “you”와 같은 개념들은 제외시킬 것을 제안하였다.

시각화 브라우저와 관련된 다른 문제점은 쌍곡면형 비주얼라이저의 범주화와 관련된 부분이다. 모든 이용자들은 탐색경험이 풍부하고 숙련된 정보이용자로서 범주나 개념을 주제별로 조직하는 방식에 있어 나름대로 기대하는 방향

이 있었다. 이때 자신들이 기대한 방식과 비주얼라이저에서 보여주는 주제표현방식이 다를 경우, 탐색에 혼돈을 느끼는 것을 알 수 있었다. 이점은 더 넓은 의미해서 해석하자면, 텍스트로 주제나 정보를 표현하는 것의 한계를 드러내는 것과 일맥상통하는 것이라 할 수 있다. 정보조직이나 정보표현에 있어서 완벽하게 객관적이고 만장일치적인 결과는 기대하기 힘들다. 이 연구에서 쌍곡면형 비주얼라이저에서 개념들을 범주화한 방식과 15명 이용자 각각이 기대하는 범주화방식은 차이가 있는 것으로 나타났고, 이러한 차이는 쌍곡면형 비주얼라이저 이용 상의 문제점으로 이어지는 평가결과를 낳았다.

시각화 비주얼라이저의 시스템적 구조부분에 대한 이용자의 평가를 요약한다면, 추출된 정보를 조직, 표현함에 있어서 시각적인 방법을 이용하는 것은 이용자들의 관심을 유도하고 새로운 방식의 정보접근을 가능하게 한다는 긍정적인 측면을 보여 주었다. 이용자들은 시각화 비주얼라이저의 구조와 정보표현방식, 이용방법에 관심을 보이며 긍정적인 반응을 나타내었다. 그러나 관찰과 인터뷰 결과, 시각화 브라우저에서 제공되는 정보량이 소화할 수 있는 양을 넘었을 때 이용자들은 쉽게 탐색주제에서 벗어나 길을 잃게 되거나 탐색을 종료하고자 하였다. 이점은 하이퍼텍스트가 소개되던 초기에도 언급되었던 유사한 비판이다. 여러 연구자들이 하이퍼텍스트와 관련하여 “하이퍼스페이스에서 길을 잃어버리는 네비게이션 문제”를 지적하였고, 이에 대한 보완책으로 맵이나 세심하게 구성된 색인 등을 이용할 것을 제안하였다(Marchionini 1988). 시각화 비주얼라이저도 개념으로 표현되는 정보량과 정보조직에 있어서 세밀하게 계획, 보완되어

야 함을 알 수 있었다.

### 2. 3. 3. 2 탐색목적과 관련된 평가

15명의 이용자들은 모두 자신의 탐색주제를 가지고 60개 기사에서 추출된 정보를 표현한 시각화 브라우저를 이용하였다. 그러나 엄밀한 의미에서 말한다면, 이들은 업무상 혹은 연구 목적으로 생성된 정보요구를 가지고 시각화 브라우저 시스템을 이용했다기보다는 시각화 브라우저 시스템을 평가하기 위해서 자신이 관심있는 탐색주제를 선정했다고 보는 것이 정확한 것이다. 즉, 자연적으로 생긴 정보요구를 해결하기 위해 탐색행위를 한 것이 아니라, 시각화 브라우저를 평가하기 위해서 탐색행위를 한 것이다. 이점을 미루어 본다면, 15명의 이용자들의 탐색목적은 자신들의 정보요구를 해결하기 위한 것이 아니므로 명확한 탐색목적이 있는 경우라고 보기 힘들다. 15명의 이용자 중 비즈니스 이용자들은 숙련된 이용자나 정보전문가에 비해서, 남북 경제협력, 개성공단에서 유망한 사업, 북한에 투자한 외국기업 등 비교적 자신의 업무와 관련된 주제를 탐색하고자 하였으므로 어느 정도 틀이 잡힌 탐색 목적을 가졌다고 볼 수도 있으나, 여전히 정보요구와 탐색목적이 선행된 상태에서 탐색행위가 이루어졌다고 보기는 힘들다. 이용자들은 탐색목적이 명확하게 정의되지 않은 경우에 시각화 브라우저의 가치를 더 높게 평가하였다. 이러한 평가는 다음 부분에 설명되는 관찰결과에 의해서 더 확실해지고 있다.

시각화 브라우저를 통해서 추출된 정보를 시각적으로 접근함에 있어서 높은 학습효과를 관찰할 수 있었다. 15명의 이용자는 정보추구에 있어서는 모두 능숙한 정보이용자였지만 이 연

구에서 제시한 시각화 브라우저와 유사한 시스템은 사용한 경험이 없는 초보자였다. 그러나 이용자들은 시각화 브라우저 시스템을 이용하기 시작하고 얼마 되지 않는 시간에 - 대부분 30분 이내에 - 시스템 기능, 인터페이스, 검색된 결과를 해석하는데 모두 익숙하고 편안하게 보였다. 이용자들은 곧 쌍곡면형 비주얼라이저가 보여주는 개념간의 관계를 이해하였고, 심지어 어떤 이용자는 쌍곡면형 비주얼라이저를 “전체정보(universe)”라고 부르기도 하였다. 또한 이용자들은 개념 비주얼라이저에서 보여주는 관계가 일반적인 관계부터 아주 구체적인 관계까지 다 포괄하고 있다는 것을 이해하고 있었다. 전반적으로, 숙련된 사용자(Power users)가 시각화 브라우저를 가장 열심히(enthusiastic) 이용하였고, 그 다음이 정보전문가(Information professionals)였고 비즈니스 사용자(Business users)가 상대적으로 가장 덜 열성을 보였다. 이 점은 분명한 탐색목적이 정의되지 않은 숙련된 사용자나 정보전문가가 상대적으로 구체적인 탐색목적을 가졌던 비즈니스 이용자에 비해서 시각화 브라우저를 이해하고 탐색을 수행하는데 뒤지지 않았다는 것을 보여주고 있음이다. 오히려 숙련된 사용자나 정보전문가는 어떤 면에서는 더 우수한 탐색능력을 보여주었는데, 이는 정보검색기술과 관련된 이해력이나 문제가 발생했을 때 문제점을 파악하는 능력, 새로운 방식으로 표현되는 정보자원에 대한 실험정신은 숙련된 사용자나 정보전문가가 더 높기 때문인 것으로 이해할 수 있다.

### 2. 3. 3. 3 기타 전반적인 평가

Marchionini가 브라우저에 대해서 언급했던

특징 이외에, 이용자들이 시각화 브라우저를 이용하고 평가한 내용은 다음과 같다. 우선, 세 부류의 이용자가 집중해서 평가한 시각화 브라우저의 특성이 조금씩 달랐는데, 숙련된 이용자는 쌍곡면형 비주얼라이저에서 나타난 개념간의 계층형 조직에 대해 잘 이해하면서도 더 직관적인(intruitive) 방식으로 개념이 조직되어야 한다는 점을 강조하였다. 정보전문가는 쌍곡면형 비주얼라이저의 다른 특징을 지적하였다. 이들은 쌍곡면형 비주얼라이저를 이용하면 개념 비주얼라이저를 이용할 때보다 적합한 개념을 우연히(serendipitously) 발견할 가능성이 높고 이러한 특징은 탐색에 도움이 된다고 평가하였다. 비즈니스 이용자는 자신이 이용하는 다른 데이터베이스의 개념과 범주화방식이 유사하기 때문에 시각화 브라우저를 통해서 쉽게 원하는 정보를 찾을 수 있다고 보고하였다.

이용자들은 이 연구에서 제시한 실제 정보검색 환경에서 시각화 브라우저의 이용가능성에 대해서 긍정적인 반응을 나타냈다. 대부분의 이용자들이 문서편집 프로그램과 같이 맞춤형이나 사용자별로 자신의 연구나 업무영역에 적합한 어휘들을 사전에 추가시켜서, 개념이나 범주를 시각화 브라우저를 통하여 볼 때, 이용자들의 전문지식이 이미 데이터베이스에 있는 개념이나 관계와 더불어 지식을 증가시키고 맞춤형하여 이용할 수 있다면 좋겠다고 응답하였다. 모든 이용자들은, 특히 비즈니스 이용자들은 시각화 브라우저가 개념들을 생성하는 과정에서 이용자의 요구사항을 반영하는 기능이 추가되기를 원하였다. 예를 들어, 개념과 범주들을 불러인 연산자에 의하여 조절할 수 있기를 언급하기도 하였고, 세세한 탐색을 광범위하게 한 후

에는 개념들을 이동할 때 건너뛰기(skip)를 하는 기능을 원하기도 하였다. 불리언 연산자의 사용을 원한 경우는, 시각화 브라우저를 이용할 때에도 이용자들은 키워드탐색 기능을 같이 이용하기를 원한다는 것을 보여주는 예이다.

세 부류의 이용자들이 시각화 브라우저를 항상 시키는 방안으로 언급했던 또 다른 부분은 정보추출방식에서 특정 이용자 집단에 더 유용하고 적합한 정보를 제공할 수 있도록 관련성(association)을 추가시켰으면 하는 것이었다. 예를 들어, 업무상 거래와 관련이라던지 혹은 재정 관련과 같이 특정분야나 특정 이용자에게 더 관련성이 높을 것으로 예상되는 정보를 모아주는 방식을 원하였다. 이렇게 할 때, 교육분야나 기타 여러 분야에서 이용가치나 가능성이 높은 시스템으로 인정받을 수 있을 것으로 기대된다고 평가하였다.

이용자 평가를 요약한다면, 이용자들은 시각화 브라우저를 이용하여 추출된 정보를 시각적으로 접근하는 점에 대하여 Marchionini의 처음 두 가지 설명과 흡사한 평가를 내렸다. 시각화 브라우저는 시스템 구조상 추출된 정보에 시각적으로 접근함에 있어, 키워드탐색 시에 요구되는 엄격한 규칙과도 같은 명령문이나 요구되는 기능을 사용하지 않아도 되는 장점이 있었고, 탐색목적이 명확하게 정의되지 않은 상황에서도 주어진 정보를 파악하고 원하는 정보에 접근해갈 수 있다는 긍정적인 특징이 있음을 발견하였다. 시각화 브라우저는 추출된 정보를 네트워크 형태로 시각화하여 표현하고 있는데, 이는 이용자들의 브라우징 욕구를 증가시켜서 정보탐색행위의 촉진제 역할을 하는 것을 알 수 있었다. 반면에 시각화 브라우저는, 특히 개념 비주얼라이저 경우에 너무 많은 개념-관계-개념

의 항목들이 제시되었을 때 이용자들은 탐색방향을 잘못 잡고 주제에서 이탈하는 문제점이 있음이 발견되었다.

### 3. 결론

이 연구는 이용자가 자동추출된 정보를 접근함에 있어서 의미있는 연결성(semanticly meaningful relationship)과 시각적인 표현(visual representation)이라는 두 가지 특징을 제공하는 시각화 브라우저 시스템을 이용할 때의 반응 및 이용가능성을 평가하였다. 연구에 이용된 시각화 브라우저는 상용 정보시각화 소프트웨어인 The Brain Technologies Corp사의 Personal-Brain을 이용하여 구축되었다. 추출된 정보를 개념-관계-개념으로 구성하여 이용자가 개념-관계-개념의 네트워크 형태로 표현된 데이터베이스 정보에 접근할 때 어떤 반응을 보이는지 관찰과 인터뷰를 통하여 질적인 평가를 수행하였다. 이용자 평가 결과 정보추출결과의 시각화표현에 대해서 다음과 같은 고려사항들이 발견되었다.

첫째, 쌍곡면형 비주얼라이저에서 보여주는 개념간의 계층구조의 객관성 및 명확성이 중요한 것으로 나타났다. 또한 이용자가 원하면 개념간의 연결을 일부 수정하거나 새로운 개념을 추가시키는 기능을 포함시켜 이용자별 맞춤성(customization)을 제공할 수 있는 방향에 대한 제안이 있었다. 개념 비주얼라이저는 정보추출을 통해 파악된 개념-관계-개념이라는 형식을 이용하여 이용자의 정보요구에 적합한 것으로 파악되는 연결고리들을 보여주는 것인데, 이때도 역시 개념과 관계 간의 의미있는 결합이



무엇보다 중요한 부분임이 드러났다. 그리고 한 화면에 보여지는 개념-관계-개념의 결합이 적정수준을 유지할 때 유용한 탐색기능을 수행할 수 있었다.

둘째, 이용자들은 위에서 언급한 개념 비주얼라이저에서 많은 수의 결합들이 검색되고, 모든 결합들을 다 시스템에 불러오는 시간문제를 개선해야 하는 부분으로 지적하였다. 해결방안은 두 가지인데, 하나는 로드시간을 단축하는 것이고 다른 하나는 멈춤(stop)기능을 추가시켜서 지나치게 많은 개념과 관계의 조합들이 검색되고 화면에 제시되어야 하는 경우에 이용자가 원하는 시점에서 현재 작업을 끝낼 수 있도록 하는 것이다. 실제로 몇몇 이용자는 멈춤기능이 있었으면 하는 바램을 표현하였다.

셋째는 두 번째 항목에 관련된 부분으로 시각화 브라우저에서 처리하는 기능들을 이용자가 알 수 있도록 피드백을 제공하는 것이 중요하다는 점이다. 처리수준, 대기시간 등을 표시해야 하는 것은 물론이고, 이용자가 각 시점까지 수행한 탐색경로(search history)를 알 수 있으면, 브라우저의 대표적 단점인 탐색주제에서 벗어나는 경우를 어느 정도 줄일 수 있을 것으로 기대된다.

넷째, 시각화 브라우저를 포함한 모든 브라우저에 관련되는 사항으로, 이용자들은 마우스 조작 중 끌기(dragging)와 클릭하기에 혼돈을 일으키는 경우가 있었다. 쌍곡면형 브라우저를 이용할 때 대부분 끌기와 클릭하기를 이용해서 개념들을 이동시키고 선택하게 된다. 그러나 마우스 조작에 대한 표준화가 이루어지지 않아 어떤 시스템들은 끌기를 통해 화면에서 개념의 위치를 이동시키는 반면에 다른 시스템들은 클릭

을 통해 선택된 개념을 화면의 중앙으로 이동시킨다. 이 연구에서 이용한 PersonalBrain은 후자에 속한다. 끌기와 클릭하기는 하나의 예에 불과한 것으로, 정보시스템과 관련된 여러 분야의 표준화 노력이 이루어지고 있음을 고려할 때, 브라우저이나 시각화기법과 관련된 표준화 노력이 같이 병행되어야 함을 알 수 있었다.

마지막으로, 이 연구는 정치, 경제, 외교 분야의 뉴스기사에 한정된 데이터를 대상으로 시각화 브라우저의 이용가능성을 평가하였다. 그리고 정보시각화의 여러 방법들 중 개념과 관계라는 구조를 표현하기에 적절하다고 여겨지는 쌍곡면형 브라우저 방법을 채택하였다. 그리고 평가의 구체성과 깊이있는 답변을 수집하기 위하여 이용자로 선정된 연구대상자 그룹도 모두 고급이용자로 제한하였다. 검색된 정보를 시각적으로 표현하고 접근할 수 있는 기법은 쌍곡면형 브라우저 이외에서 1차원, 2차원, 3차원 이상의 기법 등 다양한 방법이 현재까지 소개되었다(Ware 2004; Spence 2001). 그러나 아직도 대다수 시각화기법들이 소개 수준에 그치거나 기본적 수준정도만 상용 소프트웨어에 적용되고 있는 실정이다. 그리고 정보시각화 시스템의 이용성에 대한 연구는 미미한 수준에 머물러 있다. 이 연구는 위에서 언급했듯이 데이터의 범위, 대상이용자, 적용된 시각화기법이 매우 제한적이지만, 정보추출과 정보시각화를 연결하여 이용성평가의 범위까지 시도했다는 점에서 현재 정보시각화 연구와 응용의 범위를 넓혔다는 점에서 의의를 찾을 수 있다. 동시에 이 연구에서 이용자들이 지적했던 추출된 정보의 시각화 표현에 대한 긍정적인 반응과 개선사항으로 지적했던 부분들을 반영한 후속연구를 기대할 수 있을 것이다.

## 참 고 문 헌

- 서은경. 2002. 정보시각화에 대한 스킴모형별 비교 분석. *한국문헌정보학회지*, 36(4): 176-205.
- 이지연. 2001. 키워드탐색과 비주얼 브라우징을 이용한 이미지 검색 시스템. 『*한국정보관리학회지*』, 18(4): 183-200.
- Allen, M. 2002. "The Hype over Hyperbolic Browsers." *Online*, 26(3): 20-28.
- Bederson, B. and Shneiderman, B. 2003. *The Craft of Information Visualization: Readings and Reflections*. San Francisco: Morgan Kaufmann.
- Lynch, C. 1991. "The Technologies of Electronic Imaging." *Journal of American Society for Information Science*, 42(8): 578-585.
- Marchionini, G. 1987. "An Invitation to Browse: Designing Full-Text Systems for Novice Users." *The Canadian Journal for Information Science*, 12(3-4): 125-138.
- Marchionini, G. 1988. "Hypermedia and Learning: Freedom and Chaos." *Educational Technology*, 28(11): 8-12.
- Paik, W. 2000. CHronological information Extraction SyStem (CHESS). PhD Dissertation, Syracuse University, Syracuse, NY.
- Paik, W. and Lee, J. 2004. "Extracting Legal Propositions from Appellate Decisions with Text Discourse Analysis Methods." *Lecture Notes in Computer Science (LNCS)*, Vol. 3292, Springer-Verlag: 621-633.
- Santorini, B. 1990. "Part-of-speech Tagging Guidelines for the Penn Treebank Project." Technical report, Department of Computer & Information Science, U. of Penn.
- Schank, R. 1972. "Conceptual Dependency: A Theory of Natural Language Understanding." *Cognitive Psychology*, 3(4): 552-631.
- Sowa, J. 1984. *Conceptual Structures: Information Processing in Mind and Machine, Reading*. MA: Addison-Wesley.
- Sager, N., Friedman, C., & Lyman, M.S. 1987. *Medical Language Processing: Computer Management of Narrative Data, Reading*. MA: Addison-Wesley.
- Spence, R. 2001. *Information Visualization*. New York: ACM Press.
- Ware, C. 2004. *Information Visualization: Perception for Design*. San Francisco: Morgan Kaufmann.