

# 디렉토리 서비스 중개 게이트웨이 모형 구축

- 주요 검색포털의 뉴스, 미디어 분야를 중심으로 - \*

## Development of an Intermediary Gateway Prototype System for Directory Services

- Focusing on 'News, Media' Class of Major Internet Directories -

김성원\*\* 김태수\*\*\*

\* 이 논문은 연세대학교 대학원 박사학위논문의 일부를 요약한 것임

\*\* 연세대학교 대학원 문헌정보학 박사졸업(sungwonk@paran.com)

\*\*\* 연세대학교 문헌정보학과 교수(btree@yonsei.ac.kr)

### 목차

1. 서론
  2. 중개 게이트웨이 구축 필요성
  3. 디렉토리 서비스 중개 게이트웨이 모형 구축
    - 3.1 게이트웨이 모형 설계
    - 3.2 조건표 작성
    - 3.3 게이트웨이 모형 구축
    - 3.3 게이트웨이 성능 평가
  4. 결론
- 참고문헌

초록: 인터넷 정보검색과정에서 가장 보편적으로 사용되고 있는 검색방법은 키워드 검색이다. 키워드 검색은 정확률과 재현율의 관점에서 여러가지 단점을 지니고 있다. 이러한 키워드 검색의 단점을 보완해 줄 수 있는 장치로서 다수의 웹 포털에서 디렉토리 검색서비스를 제공하고 있다. 검색포털에서 제공하고 있는 디렉토리 서비스는 포털별로 상이한 분류체계를 사용하는 이유로 이용자에게 불편을 주고 있으며, 이러한 불편의 해소를 위해 디렉토리 서비스간 통합검색을 제공하는 중개 게이트웨이의 구축필요성이 제기되고 있다. 이에 따라 이 연구에서는 네이버, 야후,

엠펙스 등 국내 주요 포털의 디렉토리 서비스를 대상으로 통합검색을 제공하는 중개 게이트웨이의 모형을 구축하고 그 성능을 평가하였다.

키워드: 정보검색, 디렉토리 서비스, 중개 게이트웨이, 분류체계, 조건표, 통합검색

Abstract: The most widely used information searching method in the current internet environment is the keyword-based one, which has certain limitations in terms of precision and recall. Most major internet portals provide directory-based searching as a means to complement these limitations. However, that they adopt different classification schemes brings significant inconvenience to the users, and it consequently suggests a need to develop mapping gateway to provide cross-portal, or cross-directory information searching. In this context, this study attempts to develop a prototype system of intermediary gateway for integrated search, using the directory services of three major portals, Naver, Yahoo and Empas, and test its performance.

Keyword: Information Retrieval, Directory, Gateway, Classification Scheme, Mapping Table, Integrated Retrieval

## 1 서론

개방된 인터넷 정보환경하에서 도처에 산재해 있는 정보자원들은 다수의 상이한 검색서비스에 의해 조직화되어 있다. 오늘날 인터넷을 통한 정보의 유통과정에서 검색의 보편적 도구가 되고 있는 것은 검색포털이다. 누구라도 정보요구가 발생하면 검색포털의 검색기능을 사용하여 원하는 정보의 획득을 시도하게 되었으며, 이 과정에서 키워드 중심의 검색이 일반화되어 있다. 인터넷을 통한 정보의 유통은 편리성을 크게 증가시키고 있으나, 효율성의 측면에서는 인터넷상의 유용정보의 방대함, 정보의 불균질성 및 정보처리의 불완정성으로 인해 여러가지 한계를 지닌다. 이러한 키워드 중심의 정보검색의 불완정성을 보완하기 위해 다수의 검색포털에서는 검증된 정보자원을 체계화된 분류체계의 일종인 디렉토리 구조에 조직화하여 놓고 이용자에게 제공하고 있다. 검색포털에서 제공되고 있는 디렉토리의 분류체계는 문헌분류를 위해 오랫동안 정비되어온 분류체계와는 달리 표준화 되어 있지 않고 디렉토리 서비스를 제공하는 각 검색포털별로 나름의 분류체계를 구축하여 제공되고 있는 이유로 여러가지 문제점을 안고 있다.

개개 디렉토리 서비스는 표준화되어 있지 않은 독자적 분류체계를 구축하여 정보자원을 분류하고 있기 때문에, 이용자들은 각각의 검색포털에서 제공하는 상이한 디렉토리 서비스의 분류체계를 이용하는 과정에서 혼란을 겪고 있다. 또한 통합검색을 제공하는 메타 검색서비스를 통해서도 망라적인 검색을 제공하고는 있으나 검색효율의 측면에서는 지극히 낮은 수준이다. 이에 따라 개별 디렉토리 서비스의 분류체계를 분석하여 작성된 정확한 조건표에 기반하여 디렉토리 서비스간 통합검색을 제공하는 중개 게이트웨이는 대단히 중요한 의미를 갖는다고 할 수 있다. 이 연구에서는 디렉토리 서비스간 통합검색을 제공하는 중개 게이트웨이 구축의 필요성을 살펴보고 그 모형을 설계하여 구현하였다. 또한 구현된 모형시스템의 성능을 평가하여 보았다.

한편 이 연구는 몇 가지 제한점을 가지고 있다. 우선 본 연구에서 구축하는 게이트웨이 모형의 서비스 대상을 네이버(<http://www.naver.com>), 야후 코리아(<http://kr.yahoo.com>), 엠파스(<http://www.empas.com>) 등 세 포털로 제한하였다. 대상을 세 포털의 디렉토리 서비스로 한정 한 이유는 디렉토리 분류체계의 방대함으로 인해 모든 디렉토리 서비스를 포함할 수 없는 점을 반영하였기 때문이며, 대상은 국내의 주요 포털중 검색기능을 중심으로 하는 3대 포털로 한정하였다. 또한 주제분야의 관점에서 실제적인 서비스가 가능한 수준의 게이트웨이 구축은 뉴스, 미

디어 분야를 중심으로 하였고 여타의 주제는 제2 구분인 강항목까지를 대상으로 하였다. 구체적인 게이트웨이 구축 분야로 뉴스, 미디어 분야를 선정한 이유는 연구대상으로 선정한 세 포털에서 공통적으로 전개하고 있는 최상위 항목으로 상이한 디렉토리 분류체계의 비교가 용이하였기 때문이며, 디렉토리 서비스를 통한 검색의 비율이 가장 높은 항목이라는 선행연구 결과를 반영하였다(이준호, 박소연, 권혁성, 2003).

이 연구에서 게이트웨이 구축의 대상을 국내 주요 웹 검색 포털에서 제공하는 세 디렉토리 서비스만으로 제한하였으므로 좀더 많은 포털의 디렉토리 서비스를 포함하는 연구를 수행할 필요가 있다. 또한 주제분야에 있어서도 실제적 구축은 뉴스, 미디어 분야를 중심으로 하였던 바, 더 많은 주제분야를 포함하는 연구를 수행할 필요가 있다.

## 2 중개 게이트웨이 구축 필요성

디렉토리 서비스 관련 선행연구는 크게 문헌정보학 분야에서 개발된 문헌분류체계를 인터넷 정보자원을 대상으로 하는 검색서비스에 적용하는 것과 관련된 연구, 특정 주제분야의 인터넷 정보자원의 조직화를 위한 디렉토리 분류체계의 구축에 관련된 연구를 중심으로 이루어졌다. 또한 이와는 별도로 디렉토리 서비스의 전반적인 구축지침을 제시하거나 실제 디렉토리 서비스 이용행태 분석과 관련된 연구들이 수행되고 있다.

이상적인 정보검색 환경이 구현되려면 이용자들이 검색을 제공하는 서비스의 내부구조에 대해 알지 못하더라도 한번의 검색으로 자신의 정보요구에 적합한 모든 정보자원을 검색할 수 있어야 한다. 이를 위해서는 다수의 상이한 정보검색서비스에 대해 통합검색을 제공하는 중개 게이트웨이(Gateway)의 구축이 필수적이다(Chan and Zeng 2002).

통합검색을 제공하는 중개 게이트웨이의 중요성에 대한 위와 같은 선행연구 결과에 더하여, 다음과 같은 디렉토리 분류체계 분석결과를 고려한다면 중개 게이트웨이 구축은 더욱 중요한 의미를 갖는다.

첫째, 일반적으로 디렉토리 서비스는 다음과 같은 장점을 지닌다. 디렉토리 서비스는 검색과정에서 이용자에게 상하위 항목을 일람할 수 있는 브라우징 기능을 제공하며, 브라우징 결과에 따라 탐색을 확장하거나 특정화시킬 수 있도록 해준다. 또한 검색중인 이용자에게 검색어의 문맥을 파악할 수 있게 해주며, 이용과정을 통해 이용자의 정보요구를 좀더 명확히 파악할 수 있게 해준다는 효용이 있다. 또 다

른 장점으로 디렉토리 서비스가 키워드 검색서비스에 비해 검색효율 측면에서 우수하다는 점이다. 검색서비스별 검색효율을 검토한 결과, 정확률 관점에서 디렉토리 서비스가 키워드 검색에 비해 상대적으로 우수한 검색결과를 제공하는 것을 알 수 있었다. 특히 디렉토리 서비스는 단음절어, 동음이의어, 복합명사의 경우 키워드 검색에 비해 우수한 검색결과를 제공했다. 검색된 적합 정보자원의 관점에서 '소비자경제' '여성운동' '정보통신 신문' 등 일부 검색어의 경우에는 디렉토리 서비스를 통해 검색된 적합자원의 수가 키워드 검색에 비해 많았다. 정보자원의 성격이 키워드만으로 표현될 수 없는 경우에는 인간의 판단에 의해 해당 주제분야로 분류해놓은 디렉토리 서비스를 통한 검색결과가 정확률뿐 아니라 검색된 적합자원수의 측면에서도 우수한 검색결과를 나타냈다.

둘째, 위와 같은 장점에도 불구하고 개별 디렉토리 서비스는 다음과 같은 문제점을 내포하고 있다. 분류체계 평가에서 살펴 보았듯이 디렉토리 서비스간 분류체계가 상이하어 이용자들은 특정 디렉토리 서비스의 분류체계를 활용하여 검색한 경험을 다른 디렉토리 서비스 검색시 활용하기가 쉽지 않다. 이용자들이 특정 디렉토리 서비스를 효율적으로 활용키 위해서는 분류체계를 숙지하는 것이 필요하나 현실적으로 일반 이용자가 모든 디렉토리 서비스의 분류체계에 익숙해지는 것은 불가능하다. 디렉토리 서비스는 코흐(Koch 1997)가 지적하였듯이 상이한 분류체계 및 비논리적인 분류항목의 전개로 인해 이용자의 효율적인 이용을 어렵게 만든다. 이러한 분류체계를 준용한 분류작업은 분류작업이 완벽하더라도 분류체계의 비논리성으로 인해 정보자원을 분산시키기도 한다. 더구나 분류작업상의 오류까지 고려하면 개별 디렉토리 서비스는 많은 문제점들을 내포하게 된다. 또한 선행연구에 의해 밝혀진 바에 따르면 세 포털 디렉토리 서비스에 등록된 자원간 중복도는 참조링크 제외시 0.2565-0.2660, 참조링크 포함시 0.2251-0.2598 이다(배희진 등 2004). 이는 특정한 디렉토리를 통해 이용자들이 찾을 수 있는 정보자원 가운데 다른 디렉토리 서비스에서도 찾을 수 있는 정보자원의 비율이 낮음을 의미한다. 즉 어떤 디렉토리 서비스를 선택하였느냐에 따라 이용자들이 최종적으로 얻게 되는 검색결과인 정보자원에서 커다란 차이가 발생함을 의미한다.

셋째, 위와 같은 이유로 디렉토리 서비스의 장점을 살리면서 단점을 보완할 수 있는 방안의 강구가 필요하다. 위와 같은 선행 연구결과를 반영하여 디렉토리 서비스의 장점을 취하면서 단점을 보완하는 몇 가지 방안을 생각해 볼 수 있다. 하나는 디렉토리 서비스를 위한 표준분류체계를 구축하고 모든 디렉토리 서비스가 표준분류체계를 채택하도록 하는 방안이다. 실제 디렉토리 서비스를 위한 표준분류체계의 구축은 1990년대 후반 한국데이터베이스진흥센터의 참여아래 한국정보통신기술협회 표준화 활동의 일환으로 추진된 바 있다(한국데이터베이스진흥센터 1998). 그

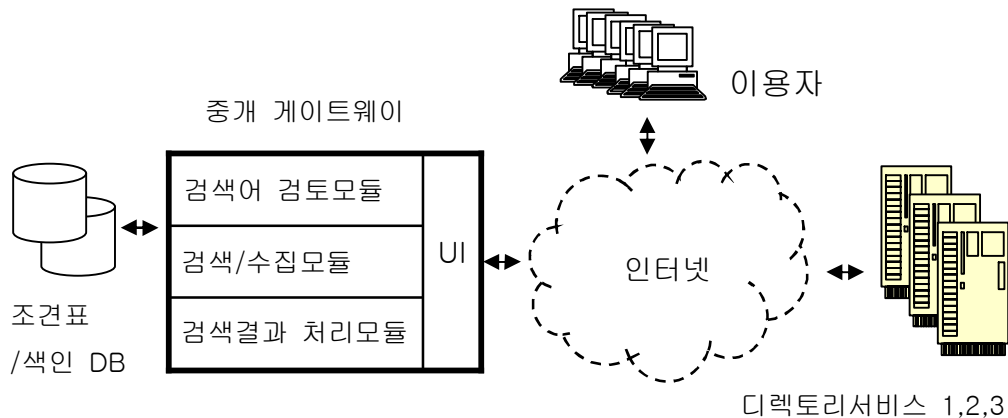
러나 야후의 ‘인문과학’ 항목의 예에서 알 수 있듯이 동일한 디렉토리 서비스 제공사의 분류체계도 서비스 대상지역에 따라 분류체계가 상이함을 볼 수 있다. 또한 디렉토리 서비스 제공사마다 중점을 두고 있는 사업영역이나 이용자 계층이 상이함을 고려하면 표준분류체계의 구축은 쉽지 않을 뿐 아니라 모든 제공사에게 채택하도록 하기도 어렵다. 또 다른 하나의 방안은 디렉토리 서비스를 제공하는 개개 서비스의 분류체계를 인정한 상태에서 상이한 디렉토리 서비스간 상호운용성의 확보 및 통합검색 기능을 제공하는 중개 게이트웨이 구축하는 방안으로서 현실적인 대안이 된다. 통합검색을 제공하는 중개 게이트웨이 구축방안으로는 크게 1) 새로운 단일 분류체계를 구축하는 방안, 2) 기존 분류체계간 매핑을 제공하는 방안, 그리고 3) 단기적으로는 기존 분류체계간의 매핑을 제공하며 장기적으로는 새로운 단일 분류체계를 구축하는 방안의 세가지 접근법을 고려할 수 있다(Chan and Zeng 2002). 이 연구에서는 두 번째 방안인 기존 분류체계간의 매핑을 통한 통합검색 제공 모형 구축을 시도하였다.

요약하면 디렉토리 서비스는 많은 장점을 지니고 있으나, 개별 디렉토리 서비스 분류체계와 분류작업상의 오류와 차이로 인해 완벽한 서비스를 제공하고 있지 못하다. 또한 개별 디렉토리 서비스의 분류체계 및 등록된 정보자원은 디렉토리 서비스에 따라 차이가 크므로 이용자가 어떤 디렉토리 서비스를 이용하느냐에 따라 이용자의 검색결과는 커다란 차이를 보이게 된다. 중개 게이트웨이는 디렉토리 서비스의 기본적인 장점을 살리면서 상이한 디렉토리 서비스 분류항목간의 매핑을 통해 개별 디렉토리 서비스 분류체계의 비논리적 항목전개에 따른 문제점을 해결하고 중복도가 낮은 개별 디렉토리 서비스 등록 정보자원을 통합검색할 수 있게 해주므로 그 구축의 필요성 및 의의가 크다

### 3 디렉토리 서비스 중개 게이트웨이 모형 구축

#### 3.1 게이트웨이 모형 설계

상이한 디렉토리 서비스간 통합검색을 제공하기 위한 중개 게이트웨이 서비스 및 중개 게이트웨이 모형은 다음과 같이 설계하였다.



<그림 1> 중개 게이트웨이 서비스 모형

이용자는 개방망인 인터넷을 통하여 중개 게이트웨이에 접근할 수 있으며, 이용자 인터페이스는 인터넷으로 접속한 이용자의 정보검색 요구를 수용하고 검색결과를 출력해주는 기능을 수행한다.

검색어 검토 모듈은 이용자가 분류항목을 직접 입력하여 검색하였을 경우 이용자 검색어를 조건표의 색인 DB와 비교하여 유효한 검색어인가의 여부를 확인한다. 확인결과 조건표 DB에 등록된 유효 검색어인 경우 조건표에서 각 디렉토리 서비스의 해당 항목 정보를 읽어 들여 해당 분류항목을 이용자에게 제시한다. 이용자의 검색어가 조건표에 등록되지 않은 검색어인 경우 오류 메시지와 DB에 등록된 유사 분류항목을 이용자에게 제공한다. 또한 이용자가 게이트웨이에서 제공하는 탐색기를 통해 분류항목을 선택한 경우 해당 분류항목과 매핑되어 있는 다른 디렉토리 서비스의 관련 분류항목의 정보와 각 분류항목의 위치정보(URL; Uniform Resource Locator)를 검색/수집 모듈로 넘겨준다

조건표 DB는 개별 디렉토리 서비스 분류항목간의 연결관계와 각 디렉토리 서비스의 분류항목 아래 등록된 정보자원에 접근할 수 있는 위치정보인 URL 정보를 저장하고 있으며, 이용자의 직접 검색에 대비하여 색인 DB를 구축하고 있다. 조건표 DB는 이용자가 검색기능을 통해 검색어를 입력하면 조건표 DB에 등록된 검색어인가의 여부를 확인하는 과정에서 참조되며, 이용자가 검색결과로 제시된 분류항목이나 탐색기를 통해 검색을 희망하는 분류항목을 선택한 경우 해당 분류항목과 매핑되는 다른 디렉토리 서비스의 분류항목 정보와 정보자원의 위치정보를 검색/수집 모듈로 넘기는 과정에서 사용된다.

검색/수집모듈은 검색어 검토모듈에서 확인된 이용자 정보요구와 조건표 DB에서 제공된 각 디렉토리 서비스의 해당 분류항목 및 위치정보를 이용하여 해당

URL로 접근하여 정보자원을 검색하고 수집하는 기능을 한다. 검색모듈을 통해 수집된 검색결과는 수집모듈을 통해 임시저장소에 보관되고, 검색 및 수집이 완료되면 검색결과 처리모듈로 수집 데이터를 넘겨준다.

검색결과 처리모듈은 검색/수집모듈이 수집한 데이터를 출력에 적합하도록 전처리한 후, 공통 등록된 정보자원을 파악하기 위한 비교 및 정렬 등의 후처리 절차를 거쳐 시스템에서 설계한 출력양식에 따라 출력할 수 있도록 해준다.

## 3.2 조건표 작성

이 연구에서 구현을 시도한 중개 게이트웨이의 핵심이 되는 부분은 상이한 디렉토리 서비스의 분류항목간 연결을 제공하는 조건표(mapping table)이다. 중개 게이트웨이는 이 조건표를 통하여 이용자의 정보요구를 각 디렉토리 서비스에 적합하게 해석하고 변환하여 개별 디렉토리 서비스를 검색하고 검색결과를 제공할 수 있게 된다.

이 연구에서 제시한 디렉토리 서비스간 중개 게이트웨이 모형의 핵심을 이루는 조건표 작성을 위해 수립한 기본원칙은 다음과 같다.

- 첫째, 기존 분류체계 우선의 원칙
- 둘째, 귀납적 접근원칙 및 상향식 구축의 원칙
- 셋째, 재현을 우선의 원칙
- 넷째, 분류항목 중시 및 등록자원 예상의 원칙
- 다섯째, 정보자원 직접제공의 원칙

디렉토리 서비스간 통합검색을 제공하는 중개 게이트웨이 구축모형 제시를 위한 이 연구에서는 시스템의 실험적 구현을 목표로 하였음과 조건표 작성작업의 방대함을 고려하여 조건표 작성 범위를 제한하였다. 실제 서비스가 가능한 정도의 세분 분류항목을 포함하는 조건표는 최상위 디렉토리 분류 중 <뉴스, 미디어> 류로 작성 대상을 한정하였으며, 전체 주제분야 분류항목간 조건표는 매핑의 기준이 되는 디렉토리 서비스 분류체계의 강항목으로 제한하여 작성하였다.

실제 서비스가 가능한 정도의 세분항목을 포함하는 조건표의 작성 분야로 <뉴스, 미디어> 류를 선정한 이유는 다음과 같다. 첫째, <뉴스, 미디어> 류는 비교대상이 되는 세 검색 포털에 모두 설치되어 있다. 또한 다른 최상위 분류항목과 비교하여 볼 때 세 포털의 <뉴스, 미디어> 류가 포괄하는 분류항목 및 정보자원의 성격과 범위가 가장 유사하여 모형제시 및 실험적 구현을 위한 연구의 목적에 적합하였기 때문이다. 둘째, <뉴스, 미디어> 류의 경우 키워드 검색보다 디렉토리 분류체계를

통한 향해 세션의 비율이 높으며, 전체 향해 세션의 54%를 차지하여 가장 많은 비중을 차지한 분야로서(배희진, 이준호, 박소연, 2004) 중개 게이트웨이 제공을 통한 이용자 편리성의 증진 정도가 클 것으로 기대되었기 때문이다. 셋째, 분류된 정보자원간의 중복도가 낮아 게이트웨이 구축을 통한 이용자 편익증진 효과가 클 것으로 판단되었기 때문이다. 배희진 등(2004)의 연구에 의하면 세 포털 디렉토리 서비스의 뉴스 언론 분야에 등록된 자원간 중복도는 참조링크 제외시 0.2632-0.3993, 참조링크 포함시 0.1099-0.1664 이다. 이는 특정한 디렉토리를 통해 이용자들이 찾을 수 있는 정보자원 가운데 다른 디렉토리 서비스에서도 찾을 수 있는 정보자원의 비율이 낮음을 의미한다. 즉 어떤 검색 포털의 디렉토리 서비스를 이용하느냐에 따라 이용자들이 최종적으로 얻게되는 검색결과인 정보자원은 커다란 차이를 보이며, 통합검색이 제공될 경우 검색되는 적합 정보자원의 수가 크게 증가할 수 있음을 의미한다. 이와 같은 이유에서 디렉토리 서비스의 분류항목 매핑을 통한 중개 게이트웨이의 구현 및 설계방안을 제시해 보고자 하는 이 연구에서는 <뉴스, 미디어> 류를 중심으로 조건표를 작성하였다.

조건표 작성의 기본 원칙으로 제시한 기존 분류체계 우선의 원칙에 따라 조건표는 대상이 되는 세 디렉토리의 분류체계 각각을 기준으로 작성하였다. 별도의 새로운 분류체계를 구축하고 그 분류체계와 연구대상으로 선정한 세 디렉토리 서비스 분류체계를 매핑하였다면 새로운 분류체계를 기준으로 하는 조건표 하나만 작성하면 분류항목간 매핑이 가능할 것이다. 그러나 개개 이용자가 익숙한 분류체계를 사용할 수 있도록 기존 분류체계 우선의 원칙을 채택함에 따라 실제 조건표의 작성은 네이버, 야후, 엠파스 각각의 분류체계를 기준으로 별도 작성되었다. 이렇게 별도로 작성된 조건표는 이용자가 선택한 디렉토리 서비스의 분류체계에 따라 선택적으로 검색을 수행할 수 있게 해준다. 즉 이용자가 네이버의 분류체계를 사용하여 검색을 수행하는 경우 네이버 디렉토리 서비스의 분류항목을 기준으로 작성된 조건표를 기준으로 분류항목간 연결관계가 파악되고 이용자에게 제공되게 된다.

실제 <뉴스, 미디어> 류의 상세 조건표 작성은 다음의 4단계 작업을 통해 수행되었다. 1단계는 조건표 작성을 위한 원시데이터의 수집 과정이다. 2단계는 수집된 원시데이터로부터 불필요한 부분을 제거하고 조건표 작성에 필요한 데이터만을 추출하여 DB로 구축하는 전처리 과정이다. 1단계에서 수집된 데이터는 이 과정을 거쳐 '사이트 명', '등록 정보자원의 URL', 그리고 해당 '디렉토리 서비스의 분류항목'의 데이터 필드로 구성된 레코드 구조로 변환되었다. 3단계는 조건표 작성의 기초로 사용될 분류항목간 연결 데이터 생성의 1차 과정이다. 앞에서 기술한 바와 같은 1, 2단계의 절차를 거쳐 수집된 데이터를 처리하여 정보자원의 URL을 기준값으로 세 디렉토리 서비스 모두에 공통으로 등록된 정보자원을 파악하였다. 공통으로 등

록된 정보자원이 각 디렉토리 서비스에서 등록되어 있는 분류항목을 추출하여 조건표 작성을 위한 기초 연결 데이터를 생성하였다. 즉 정보자원의 고유값인 URL을 기준으로 세 디렉토리 서비스 모두에 등록된 정보자원을 추출하여 각각의 디렉토리 서비스에서 어느 분류항목에 포함되어 있는지를 파악한 후, 파악된 각 디렉토리 서비스의 분류항목을 연결하였다. 4단계는 3단계에서 생성된 연결 데이터를 추가 확보하기 위한 과정으로서 3단계에서 확보된 연결 데이터가 충분치 못했던 이유로 수행하였다. 4단계에서는 2단계에서 수집된 데이터를 기반으로 두개 디렉토리 서비스에 공통으로 등록된 정보자원의 분류항목을 파악한 후, 공통 정보자원이 등록된 해당 분류항목에 포함된 다른 정보자원중 제3의 디렉토리 서비스에 공통으로 등록된 정보자원을 추출하여 연결함으로써 조건표 작성을 위한 기초 연결 데이터를 추가로 생성하였다. 4단계에서 실제 추출된 데이터의 예는 다음과 같다.

<표 1> 4단계 기초 연결 데이터

URL	디렉토리 서비스의 분류항목	
http://www.newspress.co.kr/	네이버	News/Newspapers/Indices/
	엠파스	News_Media/News/Search/Link_Collection
http://anynews.co.kr/	야후	news_and_media/Web_Directories/

실제 조건표의 작성은 위에서 기술한 방법으로 수집한 연결 데이터를 기반으로 수행하였다. 다만 위에서 기술한 기초 연결 데이터 수집작업을 통해서 연결 데이터가 확보되지 않은 분류항목에 대해서는 수작업으로 관련 항목을 배정하였다. 앞에서 파악된 분류항목간 연결 데이터를 분석한 결과 데이터가 모두 일치하는 것은 아니며 불일치가 발생하였는데, 이는 각 디렉토리 서비스 분류체계 및 분류작업의 차이에 기인한다. 이러한 데이터의 불일치에 따라 전체적인 조정과 확인이 필요하였고 수작업으로 전체적인 조건표를 작성하였다. 동일한 분류항목에 대한 연결 데이터가 완전 일치하는 경우 추가적인 조정은 지양하였고, 연결 데이터가 불일치하는 경우에는 기본적으로 다수를 차지하는 연결 데이터를 중심으로 조건표를 구성하였다. 실제 조건표 작성과정에 있어 분류체계간 분류항목의 범위 및 분류계위의 차이로 인해 많은 문제점들이 발생하였으며 조건표의 작성을 위해 처리원칙을 세우고 일관되게 적용할 필요가 있었다.

이렇게 구성된 세 포털 디렉토리 분류항목간 조건표의 일부는 아래 표와 같다. 예시로 제시된 조건표는 실제 서비스가 가능한 세분 분류항목을 포함하는 조건표로서 네이버의 분류체계를 기준으로 작성된 것이다. 실제 검색결과를 제공하는 통합 검색 서비스가 가능하려면 검색/수집 모듈이 구동되어 정보자원을 수집하여야 하고

검색/수집모듈이 구동되려면 검색할 대상이 되는 정보자원의 위치정보(URL; Uniform Resource Locator)가 조건표에 포함되어야 한다. 이에 따라 조건표에 각 디렉토리 서비스 분류항목의 위치정보를 포함시켰고, 분류항목의 위치정보가 포함된 중개 게이트웨이 서비스용 조건표의 예는 다음과 같다.

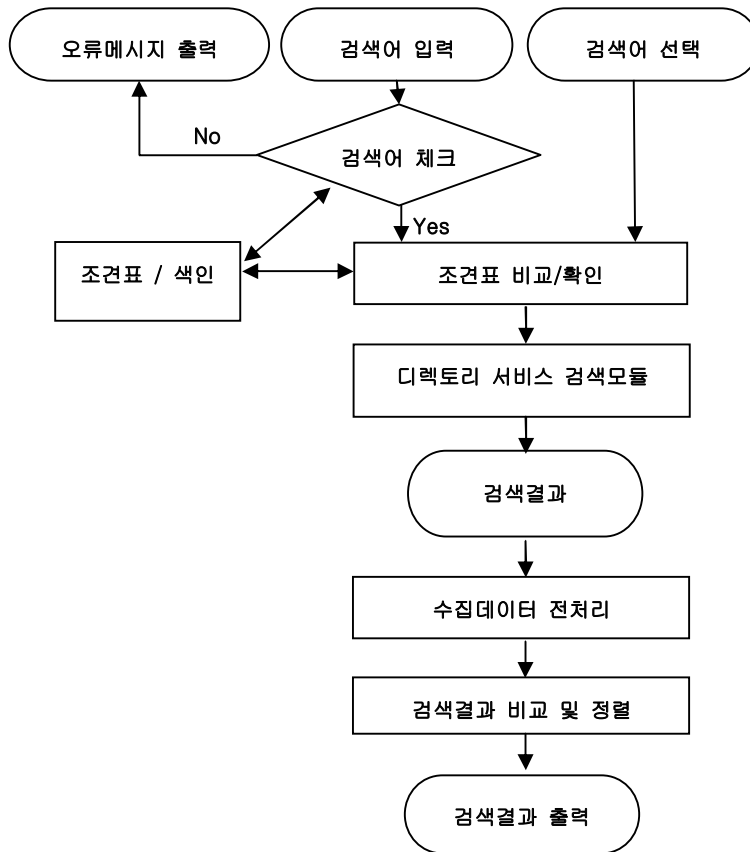
<표 2> 중개 게이트웨이 서비스용 조건표 - 네이버 분류체계 기준

네이버		야후		엠펙스	
분류항목	URL	분류항목	URL	분류항목	URL
뉴스, 미디어 > 신문 > 분야별 신문 > 해상, 해운	http://dir.naver.com/News/Newspapers/Specialty/Maritime	비즈니스와 경제 > 교통, 수송 > 배, 선박 > 뉴스와 미디어	http://kr.dir.yahoo.com/Business_and_Economy/Transportation/Ships_Vassel/News_and_Media/	기업, 쇼핑몰 > 교통, 운송 > 해운, 해상 > 신문 (=뉴스, 미디어 > 신문 > 전문지 > 해운@); 기업, 쇼핑몰 > 교통, 운송 > 수송, 운송 > 신문 (=뉴스, 미디어 > 신문 > 전문지 > 수송, 운송@)	http://dir.empas.com/Company_Shopping/Traffic_Transportation/Maritime/Newspaper/&v=7 ; http://dir.empas.com/Company_Shopping/Traffic_Transportation/Transportation_Shipping/Newspaper/
뉴스, 미디어 > 잡지, 웹진 > 사진	http://dir.naver.com/News/Magazines/Photography/	예술 > 미술 > 시각예술 > 사진 > 잡지 (=뉴스와 미디어 > 잡지 > 사진@)	http://kr.dir.yahoo.com/Arts/Fine_Art/Visual_Arts/Photography/Magazines/	문화예술, 종교 > 예술 > 사진 > 뉴스, 미디어 > 잡지, 웹진 (=뉴스, 미디어 > 잡지, 웹진 > 사진@)	http://dir.empas.com/Arts_Religion/Arts/Photo/News_Media/Newspaper/&v=7

### 3.3 게이트웨이 모형 구축

#### 3.3.1 시스템 구성

구현된 시스템의 구성도는 모형에 제시된 구성과 동일하게 구성하였다. 구체적인 데이터 처리 절차 및 흐름도는 다음과 같다.



<그림 2> 데이터처리 흐름도

### 3.3.2 구현환경

중개 게이트웨이는 구현과 갱신의 용이성을 고려하여 웹 기반으로 구축하였으며 개발환경은 윈도우 프로그래밍을 위한 개발도구인 마이크로소프트의 닷넷(.Net)을 채용하였다.

중개 게이트웨이 개발 및 운용장비의 사양은 다음과 같다.

- 장비명: HP DL380 G3
- CPU: 2 CPU(XEON 2.4GHZ/512KB)
- 메모리: 1 GB
- HDD: 36 GB

중개 게이트웨이 개발 운용을 위한 소프트웨어 및 개발환경은 다음과 같다.

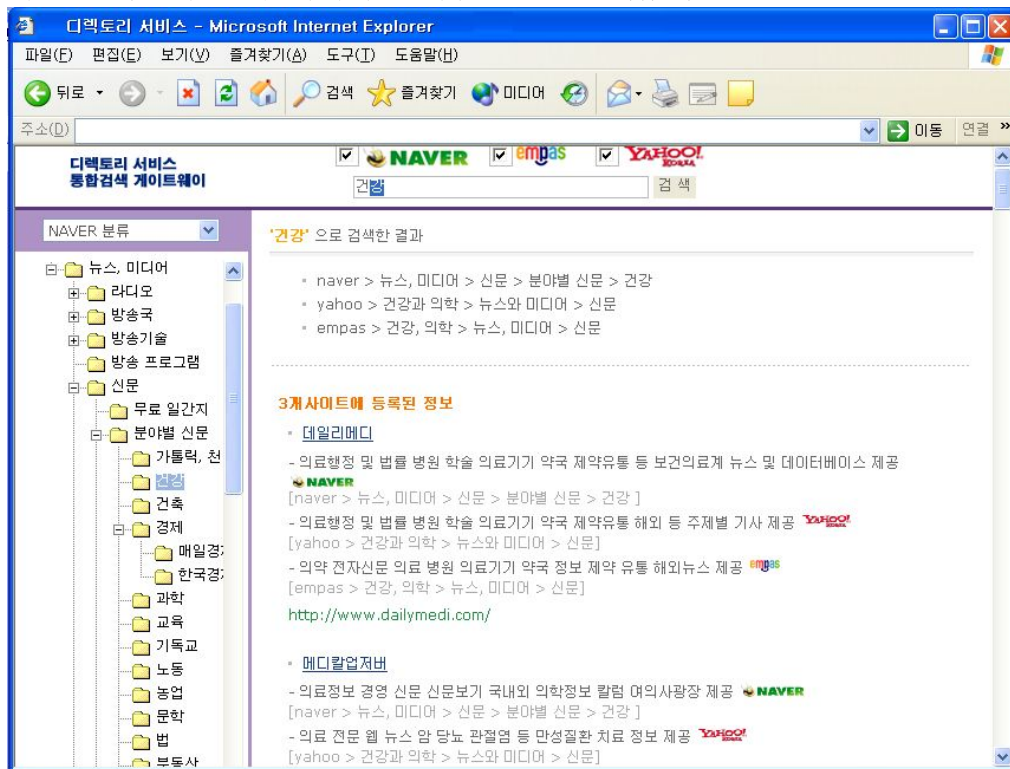
- OS: Windows 2000 Server
- Framework: Microsoft .Net
- Language: C#, ASP.Net, XML, HTC
- DBMS: MS SQL 2000

- 웹서버: IIS(Internet Information Server)
- 구현기술: XMLHTTP

### 3.3.3 인터페이스 설계

#### 3.3.3.1 이용자 인터페이스 설계

중개 게이트웨이를 통한 검색실행 및 검색결과 출력을 위한 이용자 인터페이스 화면은 다음과 같이 세개의 프레임으로 구성하였다



<그림 3> 검색 및 출력화면 설계

상단 프레임에는 검색 대상 디렉토리 서비스를 선택할 수 있는 선택 인터페이스와 디렉토리 검색 인터페이스를 제공한다. 이용자는 디렉토리 서비스 선택 인터페이스를 통해 이용할 디렉토리 서비스를 선택할 수 있으며, 세 디렉토리 서비스 모두 혹은 일부만을 선택할 수 있다. 또한 디렉토리 검색 인터페이스를 통해 디렉토리 서비스의 분류항목을 대상으로 검색을 수행할 수 있다.

좌측 프레임에는 네이버에서 제공되는 것과 같은 '카테고리 탐색기'가 제공된다. 좌측 프레임의 상단에는 이용자가 자주 사용하여 익숙한 디렉토리 서비스의 디렉토리 분류체계를 선택할 수 있는 기능을 제공하였다. 이를 통해 이용자는 세 디렉토리 서비스의 분류체계 가운데 자신이 익숙한 분류체계를 선택할 수 있다. 좌측 프

레이의 하단에는 이용자가 선택한 디렉토리 서비스 분류체계의 카테고리 탐색기가 제공된다. 카테고리 탐색기의 초기화면은 최상위 분류항목만을 나열한다. 나열된 항목 가운데 하위 항목이 포함된 항목의 경우 올림말 앞에 ‘+’ 기호가 표시되며 하위 항목이 없는 경우 ‘-’ 기호가 표시되어 이용자들에게 해당항목의 세분항목 유무를 알려준다. 이용자가 하위 항목이 있는 특정항목의 올림말 앞에 부기된 ‘+’ 기호를 클릭하면 하위 항목이 전개되며 항목의 올림말을 클릭하면 검색을 수행한 후 검색 결과를 우측 프레임에 보여준다.

우측 프레임에는 이용자의 검색수행 결과를 보여주는 공간으로 활용한다. 최초 접속시에는 초기화면으로 게이트웨이에 대한 소개와 사용법을 보여주고 구체적인 검색이 수행되면 검색결과를 보여준다.

제공되는 검색결과는 둘로 나눌 수 있다. 상단에는 해당 검색어가 포함된 디렉토리 분류항목 검색결과가 제공된다. 제공된 분류항목 검색결과는 HTML 링크가 제공되어 이용자가 선택하는 경우 웹 브라우저의 ‘새 창에서 열기’ 기능이 구동되어 해당 디렉토리 서비스의 해당 항목에 포함된 정보자원과 분류항목으로의 접근을 제공한다. 또 다른 하나의 검색결과는 해당 분류항목에 등록된 정보자원으로 화면의 하단에 제공되며, 이용자가 선택한 검색대상 디렉토리 서비스가 복수인 경우 복수 디렉토리 서비스에 공통으로 등록된 정보자원을 우선적으로 제시하고 하나의 디렉토리 서비스에만 등록되어 있는 정보자원을 제시하도록 설계하였다.

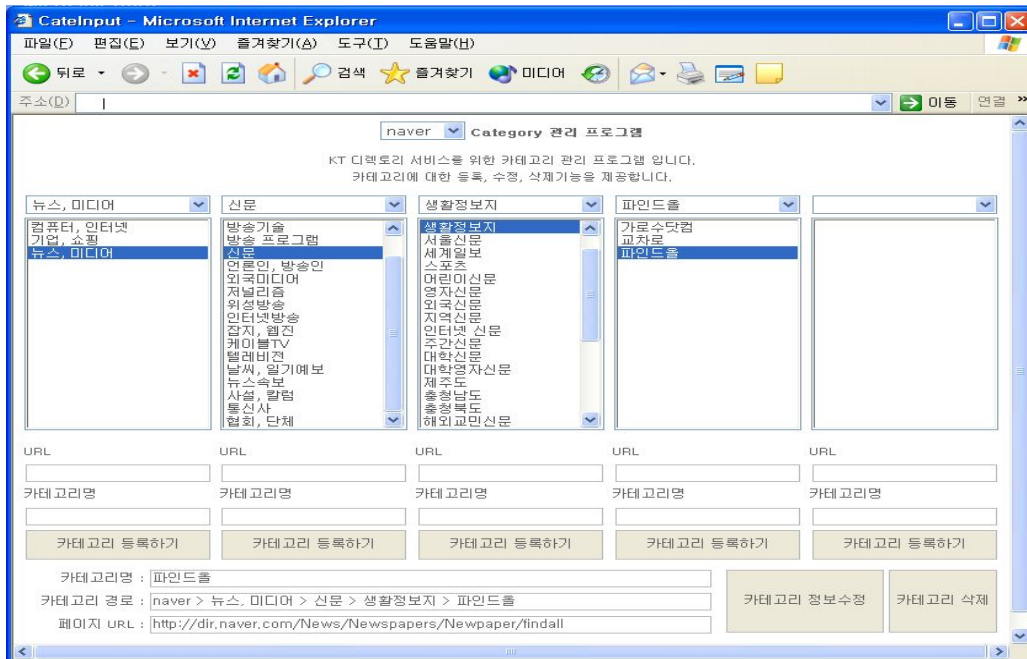
### 3.3.3.2 관리자 인터페이스 설계

중개 게이트웨이 관리자를 위한 신규 분류항목의 등록, 기등록 분류항목의 수정 및 삭제를 위한 관리자 인터페이스 화면은 다음과 같이 구성하였다.

관리자는 상단의 디렉토리 서비스 선택 박스에서 관리하고자 하는 디렉토리 서비스 분류체계를 선택할 수 있다. 관리자가 작업하고자 하는 디렉토리 서비스를 선택하면 시스템은 현재까지 등록되어 있는 해당 디렉토리 서비스의 분류체계를 제공한다.

관리자가 새로운 분류항목의 등록을 원하는 경우, 그 분류항목이 소속된 최상위 항목으로부터 차상위 항목까지를 선택한 후, 새로이 등록을 원하는 분류항목의 카테고리명과 위치정보인 URL을 입력하고 ‘카테고리 등록하기’ 아이콘을 클릭하면 등록이 이루어진다. 관리자가 기등록된 분류항목의 등록내용을 변경하고자 하는 경우 최상위 항목으로부터 해당항목으로 이동하여 해당 분류항목을 선택한 후, 원하는 등록내용을 수정하고 ‘카테고리 정보수정’ 아이콘을 클릭하면 수정된다. 수정절차와 마찬가지로 삭제를 원하는 경우에도 해당항목으로 이동하여 해당 분류항목을

선택한 후 ‘카테고리 삭제’ 아이콘을 통해 불필요한 분류항목을 삭제할 수 있다.



<그림 4> 분류항목 관리화면 설계

### 3.3.4 주요기능

이 연구에서 구현한 중개 게이트웨이 모형의 주요 기능은 다음과 같다.

#### 1) 분류체계 및 조건표 관리 기능

분류체계 및 조건표 관리 기능은 중개 게이트웨이 관리자가 사용하는 관리자용 기능이다. 디렉토리 서비스의 분류체계는 수시로 추가 수정되므로 유용한 서비스를 제공하기 위해서는 분류체계 및 조건표의 현행화가 필요하다. 이러한 요구를 만족시키기 위해 게이트웨이 관리자가 각 디렉토리 서비스의 분류체계의 변동에 따라 등록된 분류항목을 추가하거나 수정할 수 있도록 하는 기능을 구현하였다. 또한 분류체계의 변동을 반영하여 조건표를 수정할 수 있는 기능을 구현하였다. 관리자는 분류체계 관리 기능을 통하여 새로운 분류항목을 등록하거나 등록된 분류항목을 수정, 삭제할 수 있다.

#### 2) 카테고리 브라우징 기능

카테고리 브라우징 기능은 네이버 디렉토리 서비스의 카테고리 탐색기가 제공하는 기능과 동일하다. 이용자는 카테고리 탐색기를 통해 이용경험이 있어 익숙한 디렉토리 서비스의 분류체계를 선택할 수 있고, 그 분류체계에 따라 분류항목을 브라우징할 수 있다. 또한 브라우징 결과를 검토한 후 특정 분류항목을 선택함으로써

해당 분류항목에 포함된 정보자원과 다른 디렉토리 서비스에서 매핑된 분류항목에 등록된 정보자원을 검색할 수 있다.

### 3) 카테고리 검색기능

카테고리 검색기능은 이용자가 선택한 디렉토리 서비스의 분류체계에 포함된 분류항목만을 대상으로 검색하며 분류항목에 등록된 정보자원은 검색대상에서 제외된다. 카테고리 검색은 구축된 조건표의 색인 DB를 통해 조건표 DB 검색을 수행한 후 조건표 DB의 해당 분류항목과 매핑된 다른 디렉토리 서비스의 분류항목 및 해당 항목에 등록된 정보자원을 이용자에게 제공한다. 검색대상을 분류항목만으로 제한한 이유는 정보자원까지를 대상으로 하였을 경우 다른 분류항목에 포함된 정보자원까지 검색되므로 키워드 검색서비스와 같이 검색결과의 정확률을 저하시킬 소지가 있기 때문이다.

### 4) 카테고리 매핑정보 참조기능

이용자가 카테고리 탐색기에서 특정 분류항목을 선택한 경우, 시스템은 조건표 DB에서 이용자가 선택한 개별 디렉토리 서비스의 분류항목과 매핑되는 다른 디렉토리 서비스의 분류항목과 위치정보를 추출한다. 이렇게 추출된 분류항목 매핑정보와 위치정보를 통해 검색모듈은 개별 디렉토리 서비스의 해당 항목에 등록된 정보자원을 수집하고 통합적으로 이용자에게 제공할 수 있다.

### 5) 분류항목 및 정보자원 수집기능

이용자가 검색을 원하는 분류항목을 선택한 경우 시스템은 조건표 DB에서 검색에 필요한 매핑 분류항목과 위치정보를 추출하고 이를 검색/수집모듈로 넘겨준다. 검색/수집모듈은 이를 이용하여 검색을 수행하고 검색된 결과를 수집하게 된다. 검색모듈은 각 디렉토리 서비스의 해당 분류항목에 등록되어 있는 정보자원을 검색한 후, 수집모듈을 통해 수집하고 임시저장장소에 저장한다. 해당 분류항목이 최세분 분류항목이 아닌 경우에는 분류항목에 등록된 정보자원과 더불어 하위 분류항목에 대한 정보도 함께 수집한다.

### 6) 수집 정보자원 처리기능

수집된 정보자원을 이용자에게 제공하기 위해 시스템은 데이터 처리과정을 수행한다. 수집된 데이터는 우선 개별 디렉토리의 인기도 등 불필요한 정보를 제거하는 전처리 과정을 거친 후, 공통으로 등록된 정보자원을 우선적으로 제공하기 위한 비교 및 정렬 등의 후처리 과정을 거친다. 검색결과는 이같은 처리과정을 통해 분류항목, 공통으로 등록된 정보자원, 그리고 개별 디렉토리 서비스에 등록된 정보자원의 순으로 정렬된다.

### 7) 출력기능

전술한 검색, 수집 및 후처리 모듈을 통해 생성된 검색결과를 출력양식에 맞게

변환하여 이용자에게 제공하는 기능을 한다. 시스템은 세 포털에 공통으로 등록된 정보자원을 우선 나열한 다음 두개 포털에 등록된 정보자원을 나열하며 최종적으로 한 개 포털에만 등록된 정보자원을 제공한다. 이용자가 검색결과로 제공된 분류항목을 선택하는 경우 새로운 웹 브라우저 창을 열어 해당 디렉토리 서비스의 분류항목을 제공한다.

### 3.4 게이트웨이 성능 평가

앞 절에서 기술한 바와 같이 개별 디렉토리 서비스 통합검색을 위한 중개 게이트웨이를 구현하였다. 일반 이용자를 대상으로 상용서비스를 제공하기 위해서는 동의어 사전의 구축, 조건표 정비 및 현행화 모듈의 보완, 검색 및 수집 모듈의 효율화, 이용자 인터페이스 개선 등 많은 보완과 추가적인 기능 구현이 필요하겠으나, 중개 게이트웨이 구축의 모형을 제시하고 검색효율의 향상 효과를 확인한 점에서 의미를 지닌다. 중개 게이트웨이 구현의 효과를 확인하기 위해 평가를 수행하였다. 적합한 평가방안의 수립을 위한 사전평가를 수행한 후, 사전평가 결과를 기반으로 평가방법을 설계하고 평가단을 구성하여 평가작업을 수행하였다.

#### 3.4.1 평가 방법

구현된 중개 게이트웨이를 평가하기 위해 연세대학교 대학원 문헌정보학 전공자 등 정보검색분야의 관련자 20여명으로 평가단을 구성하여 평가작업을 수행하였다. 평가단에게는 세 가지 평가를 수행하도록 하였다. 첫째, 구현된 중개 게이트웨이를 사용해보고 중개 게이트웨이 사용 소감 및 개선 필요 사항을 자유로이 기술하도록 하였다. 둘째, 중개 게이트웨이 구현시 시스템의 구비조건과 이용자의 서비스 선택에 영향을 미치는 속성들에 대해 중요도를 평가하고, 각 속성에 따라 중개 게이트웨이를 평가하도록 하였다. 셋째, <뉴스, 미디어> 류를 중심으로 네개 분야로 나누어 평가용 분류항목을 제시하고 분야별로 제시된 분류항목을 사용해본 후, 개별 디렉토리 서비스와 중개 게이트웨이의 검색결과를 정확률, 검색된 적합 정보자원수, 그리고 검색소요시간의 관점에서 평가하도록 하였다.

<표 3> 중개 게이트웨이 평가를 위한 검색 항목

검색 분야	분야별 검색 항목
잡지 분야	1. 미술 잡지, 2. 육아 잡지, 3. 음식, 요리 잡지, 4. 건축잡지
신문 분야	1. 경제신문, 2. 인터넷신문, 3. 대학 영자신문, 4. 부동산 신문, 5. 불교 신문, 6. 환경 신문, 7. 경남지역신문
인터넷 방송 분야	1. 만화 인터넷방송, 2. 건강, 의학 인터넷방송, 3. 청소년 인터넷방송, 4. 여행, 관광 인터넷 방송
기타 분야	1. 포토저널리즘, 2. 원애프방송, 3. 뉴스속보
자체 선정 항목	개인별로 한두개의 분류항목을 자체 선택하여 검색을 수행

평가단에게 중개 게이트웨이의 구비요건 및 서비스 선택에 영향을 미치는 속성별 중요도를 평가한 후, 각 속성을 기준으로 중개 게이트웨이에 대한 평가를 수행하도록 하였다. 이를 위해 선행조사에서 파악된 중요속성들을 중심으로 평가대상 속성을 선정하고, 이들 속성을 포함하여 아래와 같은 평가양식을 평가단에 제시하였다.

<표 4> 검색 서비스 속성별 중요도 및 중개 게이트웨이 평가 양식

검색 서비스 속성	중요도 *	중개 게이트웨이 평가 **
(1) 검색결과의 다양성		
(2) 검색결과의 정확성		
(3) 검색결과의 최신성		
(4) 검색결과의 신뢰성		
(5) 검색의 신속성		
(6) 검색의 편리성		
(7) 검색 결과 분류의 적합성		
(8) 검색 및 결과 화면 디자인		

\* 중요도 평가척도: 각 속성별로 1(전혀 중요치 않다), 2(그다지 중요치 않다), 3(보통), 4(중요), 5(매우중요)의 5단계 나누어 평가

\*\* 중개 게이트웨이 평가척도: 1(매우부족), 2(부족), 3(보통), 4(우수), 5(매우우수)의 5단계 나누어 평가

평가단에게 중개 게이트웨이의 검색효율에 대한 평가를 수행하도록 하였다. 평가단의 평가를 위해 아래 표와 같은 검색어 집합을 제시하였다. 검색어 집합은 구현된 중개 게이트웨이의 검색효율을 검토하기 위해 뉴스, 미디어 분야에 해당하는 검색어를 무작위 추출하여 구성하였다. 뉴스, 미디어 분야로 국한하여 검색어를 선정한 이유는 구현된 중개 게이트웨이의 포괄범위가 뉴스, 미디어 분야를 중심으로 하기 때문이다. 추출된 검색어는 검색효율 비교가 가능한 검색어인가의 여부를 확인하기 위해 다시 조건표에 포함되어 있는 분류항목과 일치하는지 검토한 후, 평가단의 검색효율 평가를 위한 검색어로 선정하였다. 평가단은 이를 중심으로 각 검색 분야에서 개인별로 한 두개의 검색항목을 선택하여 개별 디렉토리 서비스와 중개

게이트웨이를 자유롭게 검색한 후 평가결과를 제시하였다.

평가단에 의한 중개 게이트웨이 검색효율 평가는 위 <표3>에서 제시된 검색 항목을 중심으로 개별 디렉토리 서비스와 중개 게이트웨이를 자유로이 사용하는 과정에서 개별 디렉토리 서비스와 중개 게이트웨이의 검색효율을 검색된 정보자원의 정확률과 적합 정보자원수 그리고 검색소요시간의 척도에 따라 평가하도록 하였다. 평가단에게 제시된 평가양식은 아래와 같다.

<표 5> 중개 게이트웨이 검색효율 평가양식

서비스 / 검색분야	네이버			야후			엠피스			중개 게이트웨이		
	정확률	적합 문헌수	소요 시간 (초)	정확률	적합 문헌수	소요 시간 (초)	정확률	적합 문헌수	소요 시간 (초)	정확률	적합 문헌수	소요 시간 (초)
잡지 분야												
자체 선정 항목												

정확률 및 적합자원수는 검색된 전체 정보자원중 적합 정보자원의 비율 및 수를 개인별로 평가하여 기재하도록 하였다. 이에 따라 동일한 검색항목에 대한 동일한 검색결과에 대해서도 평가자에 따라 정확률과 적합 정보자원의 수가 다르게 평가되는 경우가 발생하였다. 검색된 정보자원 하나하나의 적합성 여부를 따로 평가하도록 하지 않고 검색된 정보자원 전체를 기준으로 정확률과 적합 정보자원수를 평가하도록 하였다. 이는 분류항목에 등록된 정보자원의 수가 많은 경우 개별 정보자원의 적합도에 대한 평가까지 처리하기에 어려움이 있었기 때문으로 이에 따라 재현을 산출의 기초가 되는 적합 정보자원군의 도출이 어려웠다. 또한 제시된 검색항목을 사용한 검색평가에 더하여 검색과정에서 검색효율의 평가에 적합한 검색항목을 발견한 경우 이를 사용하고 평가결과를 제시하도록 하였다.

### 3.4.2 평가 결과

평가단이 중개 게이트웨이를 사용한 후 제시한 일반적인 평가의 분석결과 중개 게이트웨이의 일괄 검색 기능은 효용성이 있으나, 검색결과의 후처리 기능과 인터페이스 부분에서 세심한 기능 개선이 필요한 것으로 파악되었다.

평가단의 평가결과중 중개 게이트웨이 구현시 구비요건과 일반적인 이용자의

서비스 선택에 영향을 미치는 검색서비스의 속성에 대한 평가결과는 다음과 같다.

<표 6> 검색 서비스 구비요건별 중요도 및 중개 게이트웨이 평가

구분		검색의 편리성	검색결과의 정확성	검색의 신속성	검색결과의 다양성
평가단	속성별 중요도평가	4.78	4.67	4.22	4.00
	중개 게이트웨이 평가	3.67	3.78	3.89	4.00
선행조사	속성별 중요도 선행조사 *1	4.56	4.45	4.41	4.60
	특정 검색 사이트 평가 *2	3.72	3.71	3.73	3.76

\*1. 선행조사는 2003년 동서리서치의 조사결과를 5점 만점으로 환산한 결과

\*2. 특정 검색 사이트 평가는 동서리서치의 조사결과를 5점 만점 환산

평가단은 중개 게이트웨이를 포함하는 검색 포털의 주요 구비요건 및 이용시 선택요인으로 검색의 편리성, 검색결과의 정확성, 검색의 신속성, 검색결과의 다양성 등의 순으로 중요 속성을 선택하였다. 이는 검색포털의 구비요건이나 선택요인을 조사한 선행조사의 결과와도 부합한다(한국갤럽 2003; 모드에이전시 2002). 표에서는 평가단의 중개 게이트웨이에 대한 평가결과를 다른 검색 서비스에 대한 일반적인 평가와 비교할 수 있도록 선행조사 결과를 추가하였다. 선행조사는 2003년 동서리서치에서 수행한 검색 서비스의 속성별 중요도 평가결과를 5점 만점으로 환산한 결과이다(동서리서치 2003). 위 표에서 볼 수 있는 바와 같이 평가단이 이들 속성을 기준으로 중개 게이트웨이를 평가한 결과는 특정 검색 사이트를 대상으로 시행된 선행조사 결과와 비교할 때 실험적으로 구현된 시스템임을 고려하면 양호하다고 할 수 있다. 그러나 사용소감 및 개선방안 제시에서 파악된 항목을 중심으로 많은 개선이 필요한 것으로 판단된다.

개별 디렉토리 서비스와 중개 게이트웨이의 검색효율에 대한 평가단의 평가를 종합한 결과는 아래 표와 같다. 응답이 소수에 그친 기타 분야는 통계상의 왜곡을 방지하기 위해 제외하였으며 자체 선정 추가검색항목 역시 응답자가 없어 평가결과에서는 제외하였다.

<표 7> 개별 디렉토리 서비스 및 중개 게이트웨이 검색결과 비교

서비스	네이버	야후	엠피스	중개 게이트웨이
-----	-----	----	-----	----------

검색분야	정확률	적합 자원수	소요 시간 (초)	정확률	적합 자원수	소요 시간 (초)	정확률	적합 자원수	소요 시간 (초)	정확률	적합 자원수	소요 시간 (초)
잡지 분야	0.778	13.3	253	0.649	10.0	273	0.761	14.5	223	0.712	23.4	185
신문 분야	0.890	26.0	208	0.911	17.2	124	0.896	16.2	215	0.827	31.6	232
인터넷방송 분야	0.950	3.6	175	0.900	3.4	186	0.800	3.6	173	0.743	9.4	104
평균	0.873	14.3	212	0.820	10.2	194	0.819	11.4	204	0.761	21.5	174

위 표에서 볼 수 있는 평가단의 중개 게이트웨이 검색효율 평가결과는 다음과 같다. 첫째, 재현된 정보자원수의 관점에서 개별 디렉토리 서비스를 통해 검색을 수행하는 경우 네이버 14.3건, 야후 10.2건, 엠파스 11.4건임에 비해 중개 게이트웨이를 통해 검색되는 정보자원의 수는 21.5건으로 중개 게이트웨이를 통한 검색시 검색된 적합 정보자원수가 많다고 평가하였다. 둘째, 정확률의 관점에서 중개 게이트웨이를 통하여 검색된 검색결과의 정확률은 0.761로서 개별 디렉토리 서비스의 해당 분류항목에 등록된 정보자원의 정확률인 네이버 0.873, 야후 0.820, 엠파스 0.819에 비교하여 낮은 것으로 평가되었다. 셋째, 검색 소요시간의 관점에서 평가자들은 중개 게이트웨이를 통한 검색 소요시간은 174초로 개별 디렉토리 서비스를 사용하여 검색하는 경우의 평균 검색 소요시간 204초보다 소요시간이 단축된다고 평가하였다. 특히 중개 게이트웨이를 통해 검색되는 정보자원 모두를 개별 디렉토리 서비스를 통해 검색하기 위해서는 개별 디렉토리 서비스 모두를 사용해야 하므로 소요시간은 개별 디렉토리 서비스 사용시 소요시간을 합산하여야 한다. 이를 고려하면 중개 게이트웨이 사용에 따른 검색 소요시간 단축효과가 크다.

게이트웨이의 검색효율을 검색된 적합 정보자원수의 관점에서 살펴보면 아래 표와 같다.

<표 8> 재현된 적합 정보자원수 및 증가율 비교

구 분	재현된 적합 정보자원				개별 디렉토리 대비 재현 정보자원 증가율(%)		
	네이버	야후	엠파스	게이트 웨이	네이버	야후	엠파스
잡지 분야	13.3	10.0	14.5	23.4	75.9	134	61.4
신문 분야	26.0	17.2	16.2	31.6	21.5	83.7	95.1
인터넷 방송 분야	3.6	3.4	3.6	9.4	161.1	176.5	161.1
평균	14.3	10.2	11.4	21.5	50.3	110.8	88.6

재현율을 대신하여 검색된 적합 정보자원수로 평가한 이유는 전술한 바와 같이 재현율 산출식의 분모가 되는 적합 정보자원의 집합을 정확히 추출할 수 없었기 때문이다. 이는 정보자원의 적합성 여부가 주관적인 판단에 기반하므로 평가자에 따라 특정 정보자원의 적합성에 대한 판단이 상이하기 때문이다. 평가결과에서 알 수 있듯이 중개 게이트웨이를 통한 검색 수행시 개별 디렉토리 서비스에 비해 재현되는 적합 정보자원의 수가 증가함을 알 수 있다. 이는 중개 게이트웨이의 경우 개별 디렉토리 서비스의 해당항목에 등록된 모든 정보자원을 포함하게 되므로 당연한 결과로 여겨질 수 있으나 한편으로는 조건표가 적절히 작성되었음을 반증하는 것이다. 중개 게이트웨이 사용시 증가하는 적합 정보자원의 증가율을 좀더 구체적으로는 살펴보면 네이버 대비 평균 50.3%, 야후 대비 평균 110.8%, 엠파스 대비 평균 88.6% 가 증가하였다. 또한 개별 디렉토리 서비스를 통해 검색되는 전체 평균 적합 정보자원수 11.9건을 기준으로 하면 79.7%가 증가하였다.

정확률의 관점에서 개별 디렉토리 서비스와 중개 게이트웨이의 검색결과를 비교한 결과는 아래 표와 같다. 중개 게이트웨이를 통한 검색결과와 정확률이 개별 디렉토리 서비스를 사용하는 경우보다 낮게 평가되었다.

<표 9> 개별 디렉토리 서비스 및 중개 게이트웨이 정확률 비교

서비스 검색분야	네이버	야후	엠파스	중개 게이트웨이
잡지 분야	0.778	0.649	0.761	0.712
신문 분야	0.890	0.911	0.896	0.827
인터넷 방송 분야	0.950	0.900	0.800	0.743
평균	0.873	0.820	0.819	0.761

이는 세 디렉토리 서비스에 공통으로 등록된 정보자원의 경우 적합자원일 확률이 높으나 중개 게이트웨이에서는 비교 작업을 통해 한 건으로 제공되는데 반해, 하나의 디렉토리 서비스에만 등록된 정보자원의 경우 공통 등록자원에 비해 부적합 자원일 확률이 높지만 공통 등록자원과 동일하게 각각 한 건으로 제공되며 모두 포함되는 것에 기인한다. 또 한가지 주목할 점은 개별 디렉토리 서비스에 등록된 정보자원의 경우 서비스 제공자가 적합 정보자원만을 선별 등록하여 놓은 것임에도 불구하고 평가자들은 등록자원 모두를 적합한 정보자원으로 평가하지는 않았다는 점이다.

개별 디렉토리 검색 서비스와 중개 게이트웨이를 통한 검색결과를 비교한 위

평가를 통해 볼 수 있듯이, 중개 게이트웨이는 정확률을 크게 저하시킴 없이 검색되는 적합 정보자원수를 증가시킴으로써 검색효율의 개선에 효과가 있음을 알 수 있다. 또한 평가단의 평가내용의 일부로 밝혀진 것과 같이 개별 디렉토리 서비스를 각각 검색하는 경우보다 검색시간의 단축효과가 있음을 파악할 수 있다.

## 4 결론

이 연구를 디렉토리 서비스간 통합검색을 제공할 수 있는 중개 게이트웨이의 모형을 제시하고 구축하였으며, 구축된 중개 게이트웨이의 성능 및 효용을 평가하였다. 이러한 연구는 키워드 검색의 단점을 보완하기 위해 제공되고 있는 디렉토리 서비스 또한 제공 포털마다 상이한 분류체계를 사용하여 이용자에게 불편함을 끼치고 있고 개별 디렉토리를 별도로 반복하여 검색해야 한다는 단점을 보완하기 위한 시도로서 수행되었다.

구현된 중개 게이트웨이를 실제 사용한 후 평가한 결과는 다음과 같다. 첫째, 재현된 정보자원수의 관점에서 중개 게이트웨이를 통한 검색시 검색된 적합 정보자원수가 많아 재현율을 개선시켜준다. 둘째, 정확률의 관점에서 중개 게이트웨이를 통하여 검색된 검색결과와 정확률은 개별 디렉토리 서비스의 해당 분류항목에 등록된 정보자원의 정확률에 비교하여 낮은 것으로 평가되었다. 셋째, 검색 소요시간의 관점에서 중개 게이트웨이를 통한 검색은 소요시간을 단축시켜준다. 특히 중개 게이트웨이를 통해 검색되는 정보자원 모두를 개별 디렉토리 서비스를 통해 검색하기 위해서는 개별 디렉토리 서비스 모두를 사용해야 하므로 소요시간은 개별 디렉토리 서비스 사용시 소요시간을 합산하여야 한다. 이를 고려하면 중개 게이트웨이 사용에 따른 검색 소요시간 단축효과가 크다.

개별 디렉토리 검색 서비스와 중개 게이트웨이를 통한 검색결과를 비교한 평가를 통해 볼 수 있듯이, 시험적으로 구축한 중개 게이트웨이는 정확률을 크게 저하시킴 없이 검색되는 적합 정보자원수를 증가시켜 검색효율의 개선에 효과가 있다. 또한 개별 디렉토리 서비스를 각각 검색하는 경우보다 검색시간의 단축효과가 있음을 파악할 수 있다.

## 참 고 문 헌

- 곽철완. 2002. “전자도서관을 위한 분류체계에 대한 소고”. 국회도서관보, 39(2): 3-14.
- 김태수. 2000. 분류의 이해. 서울: 문헌정보처리연구회.
- 남영준, 최승순. 2002. “한국십진분류체계와 웹문서의 통합 분류체계 개발”. 국회도서관보, 39(1): 25-43.
- 동서리서치. 2003. 인터넷 이용에 대한 조사 보고서. 서울: 동서리서치.
- 모드에이전시. 2002. “인터넷 포털사이트 생각나는 순서는?”. [2005.03.14]. <<http://www.modagency.com/html/support/study/content.asp?idx=10>>
- 박소연, 이준호. 2002. “로그분석을 통한 이용자의 웹 문서 검색 행태에 관한 연구”. 정보관리학회지, 19(3): 111-122.
- 배희진, 이진숙, 이준호, 박소연. 2004. “국내 웹 디렉토리들의 커버리지 및 커버리지 중복성 분석.” 정보관리학회지, 21(1): 173-186.
- 배희진, 이준호, 박소연. 2004. “로그분석을 통한 네이버 이용자의 디렉토리 접근 행태에 관한 연구”. 한국도서관·정보학회지, 35(1): 1-17.
- 신동민. 2001. “인터넷 검색엔진의 디렉토리 구성에 관한 연구”. 정보관리학회지, 18(2): 143-163.
- 이준호, 박소연, 권혁성. 2003. “질의 로그 분석을 통한 네이버 이용자의 검색 행태 연구”. 정보관리학회지, 20(2): 27-41.
- 이진숙. 2003. 국내 웹 디렉토리들의 커버리지 중복성 및 분류일치성. 석사학위논문, 숭실대학교, 컴퓨터학과.
- 한국갤럽. 2003. 인터넷 이용실태 조사 보고서 - 검색, 뉴스 콘텐츠를 중심으로. 서울: 한국갤럽.
- 한국데이터베이스진흥센터. 1998. 데이터베이스의 효율적인 검색을 위한 분류체계 및 초록 작성 지침. 서울: 한국과학기술원 부설 연구개발정보센터.
- Chan, Lois Mai and Zeng, Marcia Lei. 2002. “Ensuring Interoperability among Subject Vocabularies and Knowledge Organization Schemes: a Methodological Analysis”. In *Proceedings of the 68th IFLA Council and General Conference*, 1-7.

- Dahlberg, Ingtraut. 1995. "The Future of Classification in Library and Networks, a Theoretical Point of View". *Cataloging and Classification Quarterly*, 21(2): 23–35.
- Hudon M. 2000. "Innovation and Tradition in Knowledge Organization Schemes on the Internet". In *Dynamism and Stability in Knowledge Organization: Proceedings of the Sixth International Isko Conference*, edited by Beghtol, C., Howarth, L. and Williamson, N.J. Wurzburg, Germany: Ergon Verlag. Quoted in Zins, Chaim. 2002. "Models for Classifying Internet Resources". *Knowledge Organization*, 29(1) : 20–28.
- Koch, Traugott., et al. 1997. "The Role of Classification Schemes in Internet Resource Description and Discovery". [2003.09.02]. <<http://www.lub.lu.se/desire/radar/reports/D3.2.3/>>.
- Lyman, Peter. 2002. "Archiving the World Wide Web". In *Building a National Strategy for Digital Preservation : Issues in Digital Media Archiving*. Washington, D.C. : CLIR. 재인용: 윤희윤. 2003. "디지털 장서관리의 패러독스 분석". *한국비블리아*, 14(1): 5–24.
- Nisonger, T. E. 1997. "The Internet and Collection Management in Academic Libraries : Opportunities and Challenges". In *Collection Management for the 21st Century : A handbook for Librarians*, Edited by G.E. Gorman and Ruth H. Miller. Westport, Conn. : Greenwood Press. 29–57.
- Svenonius, Elaine. 1983. "Use of Classification in Online Retrieval". *Library Resources & Technical Services*, 27(1): 76–80.
- Van der Walt, Martin. 1998. "The Structure of Classification Schemes used in Internet Search Engines". *Advances in Knowledge Organization*, 6: 379–387.
- Vizine-Goetz, Diane. 2002. "Classification Schemes for Internet Resources Revised". *Journal of Internet Cataloging*, 5(4): 5–18.
- Zins, Chaim. 2002. "Models for Classifying Internet Resources". *Knowledge Organization*, 29(1): 20–28.