

# 차세대 검색서비스의 속성에 관한 연구\*

## A Study on the Features of the Next Generation Search Services

이수상(Soo-Sang Lee)\*\*

이순영(Soon-Young Lee)\*\*\*

### 초 록

최근 정보검색 환경은 검색 2.0으로 대표되는 차세대 검색서비스에 대한 논의들이 활발해지고 있다. 따라서 이 연구에서는 정보검색의 발전과 진화에 대한 다양한 논의들을 토대로 정보검색의 발전 과정을 구분하였다. 그리고 현재 거론되고 있는 차세대 검색서비스의 등장 배경, 주요 개념, 그리고 관련 사례와 속성을 파악하였으며, 이러한 속성과 사례에 대한 데이터를 통해 차세대 검색서비스를 설명하는 핵심적인 키워드를 확인하기 위한 군집 분석을 수행하였다. 군집 분석의 결과 차세대 검색서비스를 대표하는 주요 키워드는 소셜 검색, 지능형 의미 검색, 그리고 관계기반 검색 등으로 나타났다.

### ABSTRACT

Recently in the area of the information environment, there are lively discussions about search 2.0 which is representative of the next generation search services. In this study, we divide information search model into matching and linking models according the developmental stages. Therefore, on the one hand, we analyze the background, main concepts, related attributes and cases of the next generation search services and the other, we identify the representative keywords by the group analysis of various attributes and cases of it. The result shows that the main keywords such as social search, artificial intelligence and semantic search, and relation/network based search are representative of the search 2.0.

키워드: 차세대 검색서비스, 검색 2.0, 소셜 검색, 지능형 의미 검색, 관계기반 검색  
next generation search services, search 2.0, social search,  
artificial intelligence and semantic search, relation/network based search

---

\* 본 연구는 한국과학기술연구원의 "디지털 콘텐츠 연계융합 미래형 모델 개발 연구" 과제로 연구된 내용을 수정·보완한 것임.

\*\* 부산대학교 문헌정보학과 부교수(sslee@pusan.ac.kr) (제1저자)

\*\*\* 부산대학교 대학원 문헌정보학과 박사과정(libry@nate.com) (공동저자)

■ 논문접수일자: 2009년 11월 10일 ■ 최초심사일자: 2009년 11월 16일 ■ 게재확정일자: 2009년 11월 23일  
■ 정보관리학회지, 26(4): 93-112, 2009. [DOI:10.3743/KOSIM.2009.26.4.093]

## 1. 서론

인터넷이 널리 보급되고 참여, 공유, 개방을 모토로 하는 웹 2.0 시대에 있어 정보서비스의 이용자는 프로슈머(prosumer)라는 새로운 개념으로 인식되고 있다. 현재 이 개념은 구체적으로 다음의 두 가지의 의미로 사용된다. 첫째, 마케팅 또는 경제 영역에서 유래하였으며, 생산자(producer)와 소비자(consumer) 행위를 동시에 수행하는 생산소비자 또는 '생비자'를 의미한다. 이 용어는 소비자들이 과거의 수동적인 역할에서 벗어나 제품의 생산과 유통과정에 적극적으로 참여하는 방향으로 기능을 확장하고 있다는 의미가 포함되어 있다. 특히 정보사회에서 정보를 일반적으로 수용하는 것이 아니라 정보의 생산과 소비를 함께 수행하는 개인을 말한다. 둘째, 프로슈머는 또한 전문가(professional)이자 아마추어(amateur)인 개인으로 프로투어(protour)라고 불리기도 한다. 이전에는 전문가와 아마추어 간에는 분명 구분이 있었지만, 현대에는 아마추어의 수준이 전문가 수준까지 도달한 경우가 많으며, 전문가 같은 아마추어의 활동이 많아졌다. 아마추어이지만, 웹상에서 전문가 수준의 디지털 카메라나 콘텐츠 편집기술을 사용하여 UCC(User Created Content)를 제작하여 공유하는 작업을 수행하는 개인을 한 예로 들 수 있다.

정보서비스의 관점에서 보면, 프로슈머는 과거 사람들이 권위 있는 기관이나 개인이 제공하는 정보의 이용에만 제한되었던 수용자나 정보의 수동적인 소비자에 그치지 않고 정보통신기술의 급격한 진전과 보급, 즉 미니 홈피, 인터넷 댓글, UCC, 메신저 등을 통해 능동적이고

적극적인 정보 생산자와 유통자 역할까지 아우르는 정보 프로슈머라고 할 수 있다. 또한 프로슈머들의 정보생산 수준도 단지 아마추어의 취미활동 수준에서 그치는 것이 아니라 집단지성의 발현으로 전문가 수준에 비교해도 전혀 손색이 없는 경지에까지 이르고 있다. 이처럼 정보 프로슈머가 급증하게 된 원인은 바로 지금의 청년층과 청소년층이 어린 시절부터 자신들의 생활 전반에서 정보통신기술과 기기들을 체화하고 있고, 새로운 정보통신기술과 기기의 등장을 열광하고 환호하는 이른바 N세대, M세대, 엄지족 등으로 표현되는 디지털 세대이기 때문일 것이다.

그렇다면, 이러한 정보 프로슈머들을 위해서 정보환경은 어떻게 변화하여야 하는가? 그들은 너무 많은 정보가 노출되는 것을 꺼려하면서 동시에 자신들의 생각이 반영되는 정보환경을 원한다. 그들은 또한 다양한 유형의 콘텐츠들을 언제 어디서든 편리하게 검색할 수 있기를 원한다. 특히 정보검색 환경에 대해서는 같은 키워드를 다른 이용자가 입력하면 매번 다른 검색 결과가 나타나야 한다고 기대하며, 이용자의 유동적 상황을 반영하는 검색서비스를 원한다. 콘텐츠 내에 내재되어 있는 정보적 가치를 잘 해석하여 자신들의 정보요구에 적합한 검색결과가 나타나기를 원한다. 일반적으로 정보 콘텐츠를 검색하고 제공하는 환경에서 벗어나 콘텐츠의 연계와 융합적 처리가 가능한 정보환경의 이용을 원한다. 정보의 최신성과 신뢰성뿐만 아니라, 이용자 정보요구의 만족도를 극대화하고, 정보의 생산과 소비를 편리하게 하는 새로운 정보환경을 원한다. 그들이 원하는 이러한 것들은 현재의 정보환경과는 다른 차세대의 정보환경

이다. 정보환경(information environment)은 양질의 정보자원에 대한 편리하고 효율적인 접근을 보장하여 학습, 교육, 연구와 같은 지식의 생산, 유통, 보존, 소비 등의 활동이 원활하게 수행될 수 있도록 하는 정보 인프라 또는 정보 서비스를 의미한다. 물론 이러한 역할 중에서 가장 핵심이 되는 것은 자원탐색을 위한 도구, 즉 정보검색서비스이다.

따라서 본 논문은 현재 거론되고 있는 다양한 정보검색서비스의 속성을 분석하고, 차세대 정보환경이 지향하고 있는 방향성을 모색하고자 한다. 이를 위하여 먼저 정보검색의 발전과 진화에 대한 논의에서 시작하여 이의 발전 과정을 구분·정리하는 작업을 수행한다. 그리고 다양한 유형의 차세대 검색서비스의 사례들을 수집하여 내용을 구체적으로 분석, 주요한 개념과 속성들을 도출하게 된다. 그러한 사례와 속성의 관계를 행렬로 구성하여 군집분석(clustering analysis) 기법을 통하여 유사한 속성의 군집들을 구분하고, 군집별로 차세대 검색서비스가 지향하는 키워드가 무엇인지를 확인하고자 한다.

## 2. 검색 모델의 구분

검색 모델은 일반적으로 이용자가 입력한 키워드를 최대한 만족하는 콘텐츠를 보여주도록 설계된 매칭검색 모델과 콘텐츠의 외형(서지사항)과 내용에서 관계 정보를 추출, 이를 두 콘텐츠 간 연계 정보로 활용하도록 설계된 연계검색 모델로 구분이 가능하다. 매칭검색 모델은 단일의 DB를 대상으로 한 검색에서 벗어나 복수의 DB를 대상으로 하는 통합검색, 개인

의 특성을 반영하는 개인화검색 등과 같은 기법들이 개발되었고 이뿐 아니라 정확률과 재현율을 높이기 위한 다양한 알고리즘도 개발되어 왔다. 그리고 연계검색은 방대한 정보자원들을 특정한 분류체계에 따라 구분하여 디렉토리 검색을 제공하는 군집연계와 대상 정보자원들을 직접 찾아서 링크해 주는 직접연계 등의 기법들로 나누어져 개발되어 왔다. 특히 군집연계의 경우, 방대한 정보자원을 자동으로 군집하는 기법(clustering)들도 활용되었다. 물론 매칭검색과 연계검색은 별도로 개발되어 왔지만, 하나의 검색시스템에 혼합된 기능으로 구현되어 있다.

### 2.1 매칭검색 모델

매칭검색(matching search)은 검색 모델의 기초이며, 정보콘텐츠의 내용 표현과 검색을 위한 검색질의의 표현 간의 유사도(similarity)를 평가하여 일정한 기준의 유사도에 따라 적합성을 판단하고 검색결과를 도출하는 방식으로 처리되는 정보검색 모델을 말한다. 매칭검색 모델에서 가장 중요한 현안은 유사도를 어떻게 측정할 것인가 하는 방법론의 문제와 검색성능을 향상시키기 위한 효과적인 방법이 무엇인가 하는 문제이다. 전자의 경우 매칭검색의 세부 모형으로 설명될 정도로 중요한 것이며, 방법에 따라 불리언 모델, 벡터모델, 확률모델 등으로 구분된다. 특히 불리언 모델은 가장 대중적인 매칭검색 모델이며, 대용량의 상업적 검색시스템에서 가장 많이 채택되었다. 유사도 측정 방법이 매칭의 결과를 만들어 내는 문제라면, 후자의 검색성능 향상 방법은 효과적인

검색질의의 작성, 검색결과를 토대로 하는 추가적인 검색, 질의확장 등의 기법을 의미한다. 여기서 검색성능은 검색의 속도를 나타내는 검색의 효율성과 검색결과와 정확성을 의미하는 검색의 효과성을 포함한다(정영미 2005).

최근의 매칭검색 모델에서의 주요 현안은 분산 환경에 존재하며, 다양한 유형의 복수 DB들에 대한 통합검색 방법의 문제에 있다. 통합검색(integration search)은 검색 대상 DB로부터 메타데이터를 통합하여 검색할 것인지, 검색기능을 분산 처리할 것인지 하는 기법에 따라 메타통합검색과 메타검색으로 구분할 수 있다. 메타통합검색은 통합검색 대상 DB로부터 개별 메타데이터를 수집하여 통합 메타데이터로 구축한 후 검색을 하는 경우를 말한다. 대상 DB로부터 메타데이터를 수집하는 대표적인 기법은 OAI-PMH라는 메타데이터 수확 프로토콜을 사용하는 경우이다.

한편, 메타검색은 분산검색이라고도 하는데, 이는 이용자가 요청한 질의어를 디지털장서를 소유하고 있는 내외부의 장서 시스템에 복수의 검색질의어를 전달하여 해당 시스템에서 수행된 검색 결과를 반환받아 조합하여 이용자에게 제공한다. 이 방식은 검색 질의어를 전달하고 결과를 수집하는 구체적인 방식에 따라 표준기반 메타검색과 비표준 방식의 메타검색과 같이 두 가지 유형으로 구분이 가능하다. 표준기반 검색은 검색 질의어 요청과 결과 회신에 표준적인 프로토콜이나 요소기술을 사용하는 방식이다. 대표적인 표준이나 요소기술은 Z39.50, SRU, Web Services(SOAP, UDDI, WSDL), Web/CGI, OpenAPI 등이 해당된다. 한편, 비표준 검색은 표준기반 검색 방식과 달리 대상

시스템의 고유한 검색방식에 따라 질의 요청과 응답 수신 방식이 구현되며, 주로 전용 S/W를 사용한다(이수상 2006).

방대한 정보자원들을 대상으로 하는 통합검색은 개인의 특수한 관심에 부합되는 검색 결과를 요구할 시에는 불만이 야기될 수 있으므로, 개인화검색을 수반하게 된다. 따라서 개인의 정보요구 환경에 적합한 개인화검색에 대한 기법이 개발되었다. 개인화 검색은 이용자 개인의 성향과 상황, 의도를 파악하여 해당 이용자만을 위한 검색 결과를 제공하는 것으로 정의할 수 있다. 검색서비스 영역에서 개인화검색에 주목하기 시작한 배경은 검색 환경의 변화, 특히 검색 대상의 폭발적인 증가와 정보기술의 발전에 따른 이용자의 정보요구의 변화에 기인한다 할 수 있다. 정보 과잉의 상황에서 이용자는 검색을 통해 자신의 상황에 맞는 정보를 즉각적으로 제공받기를 원하고 있고 검색서비스는 이러한 이용자 요구를 만족시키기 위해 다양한 시도를 한 것이다.

현재 대부분의 검색서비스에서 제공하고 있는 개인화검색은 크게 세 가지 관점의 개인화 즉, 검색 대상의 개인화, 검색 과정의 개인화, 검색 결과의 개인화로 정리할 수 있다. 검색 대상의 개인화는 이용자가 사전에 등록된 정보원(데이터베이스)만을 대상으로 하여 검색을 수행, 그 결과를 보여주는 것이다. 검색 과정의 개인화는 이용자의 검색 의도와 상황에 맞는 검색어를 추천하여 검색의 방향을 제시하는 서비스이다. 이 서비스는 '키워드 추천 검색'이라고도 하며, 이용자가 직접 설정한 정보에 기반하는 것이 아니라 이용패턴 분석을 통한 정보 제공 방식으로 이용자 혹은 유사한 성향을 가진

이용자 집단의 검색 패턴을 분석하여 검색어를 추천하게 된다. 검색 결과의 개인화는 이용자 프로파일을 기반으로 이용자의 성향이나 취향에 알맞은 검색 결과를 우선적으로 제시하는 것이다. 이 또한 이용자가 사전에 등록한 개인 정보, 즉 이용자 출신, 소속기관, 관심사 등에 근거하거나 앞서와 마찬가지로 이용자의 검색 패턴 분석을 통해 추출한 정보에 근거하여 검색 결과를 순위화하는 데 반영하는 방식으로 서비스를 제공할 수 있다.

## 2.2 연계검색 모델

연계검색(linking search)은 콘텐츠의 외형(서지 사항)과 내형(내용)에서 관계 정보를 추출, 이를 두 콘텐츠 간 연계 정보로 활용하는 검색 모델이다. 연계검색은 매칭검색 모델의 기법 등으로 수행된 일차적인 검색 결과 콘텐츠에 적용되는 것과 독자적인 검색 모델로 수행되는 경우로 구분이 가능하다. 전자의 경우는 간단히 '콘텐츠 연계'라고 할 수 있으며, 대부분은 검색 결과로 나타난 콘텐츠를 토대로 연관이 있는 콘텐츠를 연계하는 경우를 의미한다. 검색 행위의 결과에 수반하여 이루어지는 콘텐츠 연계의 구체적인 사례는 다음과 같다. 첫째, 연계대상 개체의 URL을 이용하여 연계하는 경우이다. 둘째, OpenURL 연계이며, 연계되는 콘텐츠를 탐색하고자 하는 탐색자의 상황을 인식하여 연계결과를 제공하는 방식이다. 셋째, 꼬리물기 연계이며, 콘텐츠의 특정한 속성(저자, 주제 등)을 이용하여 다른 저작물을 연계할 경우에 사용한다. 특정 저자의 다른 저작 콘텐츠, 특정 주제의 다른 콘텐츠를 연계하

는 경우가 해당된다. 넷째, 의미기반 연계로 의미사전을 활용하여 특정한 용어의 관련어(동의어, 계층어, 유사어 등)를 연계하는 경우이다. 다섯째, 인용 연계이며, 특정 콘텐츠를 인용 또는 피인용의 관계에 있는 콘텐츠를 연계하는 경우이다. 이러한 콘텐츠 연계의 사례들은 연계의 체인을 따라가면서 특정한 콘텐츠들을 연계하거나 검색해 주는 방법이다.

한편, 독자적인 검색 모델로서의 연계검색은 크게 두 가지 유형으로 구분이 가능하다. 첫째 브라우저 연계검색으로, 검색 대상 DB의 개별 콘텐츠들을 카테고리로 구분하여 브라우징으로 검색을 수행하는 경우이다. 브라우징 연계검색은 검색자가 원하는 카테고리를 향해하면서 적합한 콘텐츠를 쉽게 탐색할 수 있는 이점이 있지만, 카테고리를 어떻게 분류할 것인가 하는 어려운 문제를 해결하여야 한다. 둘째 개체관계형 연계검색으로, 콘텐츠의 각종 관계적 속성을 활용하여 검색하는 경우이다. 콘텐츠의 관계는 서지적 관계, 사회적 관계, 의미적 관계 등과 같이 다양한 속성의 관계로 구분할 수 있다. 앞서 언급한 콘텐츠 연계에서의 OpenURL 연계나 인용문헌의 연계도 콘텐츠의 관계에 기반한다는 관점에서 보면 유사하다. 그러나 개체관계형 연계검색은 검색의 후속행위가 아니라 그 자체가 바로 검색행위이며, 콘텐츠 개체의 각종 관계적 속성을 이용하여 검색을 수행하는 것이다. 현재 가장 대표적인 사례는 FRBR 모델을 활용하는 개체관계 검색이다. 물론 콘텐츠 간 인용관계나 용어의 의미적 관계를 검색에 활용하는 사례도 개체관계형 연계검색의 사례라 할 수 있다.

### 2.3 세대별 검색 모델의 구분

앞서 구분한 검색 모델의 유형과 달리, 정보 검색의 발전 및 진화의 단계를 어떻게 인식하는가에 따라 검색 모델을 구분할 수 있다. 박영진 등은 웹 2.0에서의 정보검색의 미래에 대해 논한 바 있는데, 그는 정보검색 모델을 국내 웹 정보검색서비스의 발전 단계에 따라 아래와 같이 5단계로 나누고 있다(박영진 외 2006). 1단계는 타 정보검색서비스에 질의어를 전달하여 검색 결과를 모아서 보여주는 메타검색서비스이고, 2단계는 1997년에 등장한 야후가 사람에게 의해 카테고리별로 홈페이지를 등록하여 검색할 수 있도록 한 디렉토리 서비스이다. 3단계는 복합적인 검색도구를 제공하며, 자연어처리 기술을 기반으로 하는 자연어 검색서비스의 등장으로 보며, 4단계는 뉴스 연동, 통합검색과 더불어 네이버의 지식검색서비스가 등장한 2000년 이후의 정보검색으로 구분하였다. 그리고 5단계는 개인별 특성 및 사용행태에 최적화된 개인화 검색, 사용자의 의도와 정보를 추가한 대화형 검색, 사용자가 관심 있어 하는 정보를 자동 분석하여 제공하는 에이전트 기반 검색 등 의미검색(Semantic Search)으로 진화 발전할 것으로 예상하고 있다.

이경일은 검색서비스를 3세대로 나누어 접근하고 있다. 그가 제안하는 1세대 검색은 야후와 알타비스타와 같이 웹 페이지 콘텐츠를 단순 링크하거나 키워드 기반의 검색을 제공하는 인터넷 초창기의 검색서비스이다. 키워드 기반으로 접근하였기 때문에 검색 결과의 정확도와 신뢰도가 낮았다고 평가하고 있다. 2세대 검색은 링크 분석 방식의 페이지 랭크(Page Rank)

알고리즘을 적용한 구글이 대표적인 사례이다. 이 방식은 웹 페이지에 링크된 횟수가 높으면 신뢰도와 정확도 역시 높을 것이라는 가정에 근거한다고 한다. 하지만 재현율과 정확률에 초점을 둔 이러한 검색서비스 출현에도 이용자는 수많은 가비지 데이터(garbage data) 속에서 원하는 정보를 다시 한 번 찾아야 한다는 문제점은 여전히 해결하지 못한다. 한편, 3세대 검색은 개방, 공유, 참여에 기반한 웹 2.0이라는 새로운 트렌드를 반영하는 검색으로, 이용자 관점에서 UI를 강화하고 문서의 분류와 군집, 지능형, 개인화 검색, 전문검색 등의 다양한 서비스 제공하는 경우를 말한다. 이 세대의 검색서비스는 웹 2.0 환경과 더불어 텍스트 마이닝, 시맨틱 웹 기술을 융합한 형태의 서비스를 제공하고 SOA(Service Oriented Architecture) 환경 지원 및 SaaS(Software as a Service) 모델로 확장하고 있다. 이러한 논의를 기반으로 그는 1세대 검색과 2세대 검색을 검색 1.0, 3세대 검색을 검색 2.0이라고 분리하고 있다(이경일 2008).

### 3. 차세대 검색서비스의 내용분석

1990년대 말 닷컴 붐피를 이겨낸 인터넷 업체들의 특성 분석에 기초하여 최근에 웹 2.0이라는 용어가 등장하였다는 사실은 이제 상식이 되었다. 웹 2.0은 이전과는 다른 새로운 정보환경으로 크게 주목을 받았고, 정보검색 영역에 새로운 바람을 불러일으키는데 충분한 가능성을 열어 주었다. 이러한 가능성에 힘입어 이전의 검색 환경과는 다른 새로운 검색 환경을 지

칭하는 용어로 '검색 2.0'이라는 용어가 등장하여 사용되고 있다. 검색 2.0 또는 2세대 검색, 차세대 검색으로 불리는 새로운 검색 모델은 이의 등장 배경에 맞게 기본적으로 웹 2.0과 시맨틱 웹 기술 등을 정보검색에 활용하려는 노력으로 설명할 수 있다. 검색 2.0의 등장 배경은 정보환경의 변화로 인한 정보서비스 제공자의 입장 변화와 이용자의 역할 변화로 정리할 수 있을 것이다.

새로운 정보환경은 이상에서 언급한 정보서비스 제공자보다 이용자에게 더 큰 영향을 미쳤다고 볼 수 있다. 앞서 언급한 바와 같이, 과거의 이용자는 관리자 중심으로 개발된 자원과 서비스를 일방적으로 제공받는 입장이었으나, 웹 2.0시대의 이용자는 웹상에서 이루어지는 모든 서비스 주체로 군림하면서 콘텐츠를 생산하고 관리하고 공유하는 역할과 더불어 새로운 서비스를 창출하고 보급하는 역할도 수행하고 있다. 이용자의 이러한 역할 변화를 단순히 '이용자가 새로운 역할을 맡게 되었다'는 관점에서 해석해서는 안 될 일이다. 이것은 일종의 신분 상승이라고도 볼 수 있는 이용자의 위상 변화라 하겠다. '내게 필요한 정보가 어디에 어떠한 형태로 존재하고 있는지'하는 정도의 수준이 아니라 '내게 필요한 정보가 정확히 무엇인지'조차 알지 못했던 무지한 이용자가 콘텐츠 생산을 위해 적극적으로 관련 정보를 검색하고 본인이 생산한 콘텐츠를 공유하는 과정에서 적극적으로 관련 서비스를 활용하는 정보 프로슈머로 변모했다는 말이다. 말 그대로 전문적인 지식과 구체적인 요구를 가지고 생산에 적극적으로 참여하는 소비자가 바로 현 시점의 이용자인 것이다. 이다음의 논의는 분명해졌다. 정

보요구가 분명한 이용자는 더욱 효과적인 검색 서비스를 원할 것이고, 이는 차세대 검색서비스 등장의 필연성을 예고한 것이나 다름없다.

본 절에서는 검색 모델에 대한 이론적 접근에서 한 단계 더 나아가 차세대 검색서비스의 실제 구현 사례의 내용을 분석하여 차세대 정보검색환경에 개념적 속성을 도출하고자 한다. 여기에서는 차세대 검색을 기존의 검색 환경이 진화하고 발전된 다양한 현상이라고 보고, 그것을 대표하는 용어로 검색 2.0을 사용하기로 한다. 따라서 검색 2.0 환경에 대한 기존의 논의를 살펴보고 이 작업을 통해 검색 2.0의 주요 개념들을 추출하게 된다. 이 개념들을 바탕으로 문헌 리뷰, 뉴스를 포함한 웹 문서 검색을 수행하여 주요 사례를 선정하고, 이를 비교·분석하여 차세대 검색서비스의 속성을 도출하게 된다.

### 3.1 검색 2.0의 주요 개념

검색 2.0의 특성에 관한 논의는 국내외의 검색 2.0 서비스 업체의 등장과 함께 한다. 소쿨루(Sokullu)는 현재의 인터넷 검색이 여전히 유아수준에 머물러 있으며, 많은 사람들이 더 나은 검색 솔루션을 기대한다고 하면서 최근의 검색 경향을 3가지, 즉 이용자 인터페이스의 강화, 기술 강화, (Vertical Engine을 통한) 접근 강화로 보고 있다. 특히 기술 강화에서는 의미 검색, 검색 결과의 군집화, 집단지성, 개인화 검색, 이미지 검색, 소셜 검색 등을 지원하는 검색서비스의 예를 들고 있다(Sokullu 2006). 또한 에지(Ezzy)는 전통적인 검색이 불리언 질의, 질의 적합성과 링크 분석에 기반하고 있다고 지적하며, 검색 2.0은 3세대 검색 기술, 즉 협업, 집단 지성, 풍부한

이용자 경험 등을 활용하여 정보를 더욱 생산적으로 만드는 특별한 능력을 가진 새롭고 개선된 모델이라고 언급하였다(Ezzy 2006).

웹 기술 관련 뉴스와 리뷰, 분석을 제공하는 블로그인 ReadWriteWeb에서는 2007년에 '검색 2.0이란 무엇인가?'라는 주제로 설문을 실시하였다. 설문은 검색 2.0의 대표적인 특성이라 여겨지는 몇 가지 항목을 선정하여 실시되었는데, 여러 가지 항목 중에서 인공지능(Artificial Intelligence)에 가장 많은 표(23%)가 몰린 것으로 나타났다. 이어서 이용자 협력을 기반으로 한 인력 검색(People Powered Search), 전문 검색(Vertical Search), 개인화 검색(Personalized Search), 클러스터링(Clustering), 소셜 검색(Social Search), 시각화(Visualization), 미리 보기(Preview) 순으로 나타났다.

이상의 논의는 주로 상업 포털을 중심으로 한 인터넷 검색에 초점이 맞춰져 있다. 하지만 도서관에서도 1980년대 중반부터 온라인목록(OPAC)을 중심으로 도서관 이용자에게 검색서비스를 제공해왔다. 다시 말해, 도서관에서의 검색은 "도서관에 소장되어 있는 각종 자료에 대해 목록이라는 연결고리를 통해서(노지현 2007)" 이루어진 것이다. 그러나 탐색도구로서 기존의 도서관목록은 이용자의 요구와 기대에 제대로 부응하지 못했고, 이에 도서관 목록의 본질적 목적에 대한 오랜 논의 끝에 1998년 IFLA가 서지레코드 의 기능상의 요건인 FRBR 모델을 발표하기에 이른다.

FRBR 모델은 기존의 평면적인 서지레코드의 구조를 다양한 개체와 속성, 그리고 개체들 간의 관계를 통해 입체적으로 연결함으로써 이용자들이 원하는 자료를 보다 논리적으로 탐색

할 수 있는 구조를 제시하였는데, 이는 기존의 목록이 구현하지 못한 서지레코드들 간의 서지적 관계를 논리적으로 연결한 '의미론적 모델'이다. 검색의 관점에서 FRBR 모델은 개체 간 네트워크 구조를 통해 이용자들이 보다 손쉽게 원하는 정보를 찾아갈 수 있도록 지원하는 도구로 사용된다. 그리고 FRBR 모델은 '개체 간 관계 설정을 통한 검색 성능의 향상, 서지적 관계를 이용한 이용자 인터페이스 향상, 다양한 자료에 대한 윈스톱 검색' 등의 의의를 가지고 있다(노지현 2009). 한편, FRBR 모델은 너무 복잡하여 적용하기가 쉽지 않으며 전체적 개념을 이해하기도 어렵고 개체-관계모델이 지니는 구조적 한계가 있다는 지적이 있으나, 서지정보를 기술하는데 있어 FRBR은 미래 정보환경에 가장 적합한 모델이며 충분한 잠재력을 가지고 있다고 평가되고 있다. 따라서 이용자는 개체를 표현하는 하나 이상의 속성을 사용하여 탐색질의를 작성할 수 있고, 원하는 개체를 포괄적 또는 특정적으로 탐색할 수 있으며, 개체와 연계된 다른 부차적인 정보를 참조할 수 있다(서은경 2006). 다시 말해서 차세대 목록인 FRBR 모델을 기반으로 한 도서관검색 2.0은 관계 검색, 네트워크 검색으로 정의할 수 있을 것이다.

뿐만 아니라 디지털도서관 연구에서도 최근 새로운 검색환경과 관련하여 주목할 만한 움직임이 있다. 아일랜드의 DERI(Digital Enterprise Research Institute)에서 연구, 개발한 시맨틱 디지털도서관(Semantic Digital Library) 엔진인 JeromeDL이 바로 그것이다. 이 엔진은 디지털도서관의 혁신적인 검색서비스에 대한 연구의 결과물로 탄생한 것이다. JeromeDL 즉, 시맨틱 디지털도서관은 지식을 공유하는 데 있

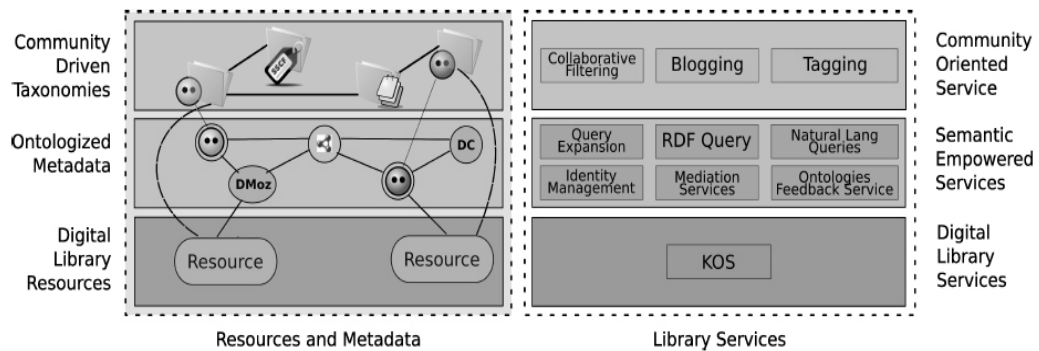


어 공동의 접근을 지원하는 디지털도서관이며, 이를 위해 시맨틱 서비스와 소셜 서비스를 제공해야 한다고 주장하였다.

JeromeDL의 아키텍처는 아래 <그림 1>과 같이 도식화된다. 간략히 설명하면, 아키텍처의 왼쪽은 자원 관리 영역으로 기존의 디지털도서관 자원을 의미적으로 검색할 수 있도록 확장 온톨로지를 구축하고, 그 상위에 이용자들의 참여를 기반으로 텍소노미와 폭소노미를 구축한다. 아키텍처 오른쪽은 도서관 서비스 영역으로 3가지 서비스 즉, 디지털도서관 서비스와 시맨틱 서비스, 커뮤니티 기반의 소셜 서비스를 제공한다.

이상의 논의를 분석하여 검색 2.0의 주요 개념을 추출하면 아래 <표 1>과 같다. 논의에서 언급된 키워드들을 언급된 순서대로 나열한 다음 유사 개념을 하나의 개념으로 보고 중복을 제거한 후 도서관 영역과 비도서관 영역으로 나누어 정리하였다.

도서관 영역은 앞서 FRBR 모델을 기반으로 한 검색서비스를 관계 검색, 네트워크 검색으로 정리하였기에 이 개념을 기술하였고, JeromeDL이 제공하고 있는 서비스 중 차세대 검색과 관련된 부분만을 정리한 것이다. 특히 소셜 서비스에서 SSCF(Social Semantic Collaborative



<그림 1> JeromeDL 아키텍처

<표 1> 기존의 논의에서 다루어진 검색 2.0의 주요 개념

	도서관 영역	비도서관 영역
검색 2.0 주요 개념	<ul style="list-style-type: none"> <li>• 네트워크 검색(Network Search) 또는 관계 검색</li> <li>• 자연어 검색(Linguistic Search)</li> <li>• 검색 결과 범주화</li> <li>• 커뮤니티 기반 검색(Community-driven Search)</li> </ul>	<ul style="list-style-type: none"> <li>• 개인화 검색(Personalized Search)</li> <li>• 검색 결과 군집화(Clustering)</li> <li>• 인공지능 혹은 지능형 검색</li> <li>• 전문 검색(Vertical Search)</li> <li>• 소셜 검색(Social Search)</li> <li>• 의미 검색(Semantic Search)</li> <li>• 이용자 협력 검색(People Powered Search)</li> <li>• 시각화(Visualization)</li> <li>• 대화형 검색(Real Conversation)</li> <li>• 미리보기(Preview)</li> </ul>

Filtering)는 이용자 네트워크를 통해 검색된 자원의 우선순위를 부여하기 위한 알고리즘으로 커뮤니티 기반 검색(Community-driven Search)으로 보았다.

비도서관 영역에서는 여러 논의에서 중복적으로 언급한 개념을 순서대로 나열하였다. 다시 말해 검색 2.0을 논하는데 있어 개인화 검색이 가장 많이 언급되었고, 다음으로 검색 결과 군집화, 지능형 검색 등의 순서로 언급되었다.

### 3.2 검색 2.0의 주요 사례

검색 2.0 사례의 선정은 다음의 두 단계로 진행되었다. 첫째, 검색을 통한 관련 사례의 수집 단계이다. 이를 위하여 먼저 본 연구의 주제어인 '검색 2.0(Search 2.0)'과 '차세대 검색(Next Generation Search)'을 질의어로 하여 인터넷

(Google, NAVER 등) 및 각종 웹DB 검색을 수행하였으며, 검색 결과 새로운 검색서비스 사례를 소개하고 있는 뉴스, 블로그, 학술논문 등의 콘텐츠에서 검색 2.0에 해당되는 구체적인 사례들을 도출하였다. 다음으로 앞의 <표 1>에 정리한 검색 2.0의 주요 개념들을 질의어로 하여 인터넷 및 웹DB 검색을 수행하였으며, 관련된 콘텐츠로부터 해당 사례들을 도출하였다. 두 번째 단계는 해당 사례의 웹 사이트에 직접 방문하여, 검색 2.0의 개념에 부합하는지를 확인하는 작업을 수행하였다. 그러한 작업의 결과 <표 2>와 같이 14가지 개념별 61가지 사례가 선정되었다.

### 3.3 검색 2.0의 주요 속성

<표 2>의 내용은 14가지 주요한 개념에 부합

<표 2> 검색 2.0 주요 사례

번호	주요 개념	주요 사례
1	검색 2.0	Swicki, Rollyo, Clusty, Wink, Lexxe, Gravee, iSearch, Krugle, LivePlasma, Qube, ZoomInfo, RedZee
2	관계 검색(혹은 네트워크 검색)	AustLit, FictionFinder, 한국연구재단 기초학문자료센터, OntoFrame
3	자연어 검색	Lexxe, Powerset, Hakia, Cognition
4	커뮤니티 기반 검색	Wikia, wiSkki, Twitter, Swicki, Rollyo, ChaCha
5	개인화 검색	Rollyo, Swicki, Collarity
6	검색 결과 군집화	JeromeDL, Clusty, [IN2], iBoogie, WebClust
7	지능형 검색	Hakia, Powerset, Cognition, Wolframalpha, Bing, [IN2], 스마트 앤서, 큐로보(Qrobo), 미르서치(Mir Search)
8	전문 검색	Google Universal Search, Bing, Simplyhired, Technorati, cRANKy, 해모수
9	소셜 검색	Rollyo, Swicki, Mahalo, Stumpedia, Wikia, Qitera, Scour, Wink, Baynote, Delver, OneRiot, SideStripe, ChaCha, Gist, twitter, Aradvark
10	의미 검색	OntoFrame, Hakia, SenseBot, Powerset, DeepDyve, Cognition
11	이용자 협력 검색	Wink, ChaCha, Mahalo, Boomja, AnooX, twitter
12	시각화	Quintura, Kartoo, Twine, RedZee, Like, Zillow, Cuil, 루키, 아울립
13	대화형 검색	Lexxe, 스마트 앤서, ChaCha
14	미리보기	Snap

한다고 판단한 61가지 사례들의 리스트이다. 여기에서는 각각의 사례들이 제시하는 구체적인 내용을 분석하여, 검색 2.0의 속성으로 명명할 수 있는 개념어들을 도출하고자 한다. 두 가지 관점, 즉 검색서비스의 이용자 관점과 관리자(혹은 개발자) 관점에서 내용분석을 시도하였다.

먼저 이용자 관점과 관련된 개념적 속성은 앞서 주요 사례의 확인을 위해 검색한 뉴스나 블로그의 웹 문서나 학술논문 등을 대상으로 추출하였다. 그리고 관리자 관점의 개념적 속성은 앞서 선정한 61가지 사례를 소개하는 각각의 웹 사이트를 대상으로 추출하였다. 다시 말해 외부에서 해당 사례를 소개하기 위해 사

용한 속성이 무엇인지, 그리고 내부에서 본인들의 검색서비스를 소개하기 위해 사용한 속성이 무엇인지를 보고 이를 추출하여 <표 3>과 같이 연구자가 직접 정리하였다. 속성은 전체 28개로 구분하였다.

이와 같은 분석 과정을 통해 기존의 논의에서 다루어진 검색 2.0의 특성들 이외에도 메타 검색, 탭 검색(Tap Search), 태깅(Tagging/Tag Cloud), 자연어 처리 기술, 인물 검색, 인력 검색(Human Powered Search), 맞춤 검색(Custom Search), 추천 엔진(Recommendation Engine) 등과 같은 새로운 속성들을 나타내는 키워드를 추출할 수 있었다.

<표 3> 검색 2.0의 속성과 그 내용

번호	속 성	내 용
1	소셜 검색	개인의 사회적 관계에 기반한 검색서비스, 이용자의 참여가 검색 결과에 영향을 미침
2	개인화 검색	검색 대상, 과정, 결과를 이용자 개인에 맞게 설정하여 맞춤형 검색 지원
3	커뮤니티기반 검색	이용자가 특정 분야에 대해 깊이 있는 검색을 수행할 수 있도록 지원, 검색 결과는 특정 커뮤니티의 검색 행태를 기반으로 하기 때문에 해당 주제 분야에 더욱 집중되어 있음
4	메타 검색	다수의 사이트(혹은 데이터베이스)를 대상으로 통합 검색한 결과를 재정렬하여 제공, 비표준 방식의 통합 검색
5	군집화	유사도에 따라 검색 결과 군집화하여 제공
6	탭 검색	검색 결과를 특정한 기준에 따라 분류하여 탭(tap) 형식으로 제공, 일반적으로 정보자원의 유형(이미지, 동영상, 웹 문서 등)이 분류지가 됨
7	태깅	이용자나 관리자가 검색 결과에 주제(topic)명을 부여, 태그 구름(tag cloud) 형태의 검색 결과 제공 및 검색어 추천
8	이용자 협력 검색	이용자가 직접 검색 결과를 제공하거나 검색 결과에 대한 이용자의 투표, 이용자 즐겨찾기(혹은 디렉터리)가 검색 결과에 직접적인 영향을 미침
9	자연어 처리 기술 (Natural Language Processing Technologies)	컴퓨터가 사람의 언어를 이해할 수 있도록 하는 기술로, 검색 쿼리의 의미를 분석하고 추출하여 쿼리에 직접적으로 응답할 수 있으며, 키워드 검색에 있어서도 더욱 적합한 결과를 생성함
10	자연어 검색	다양한 기술을 적용, 컴퓨터가 사람의 언어를 이해할 수 있도록 하여 질의어의 의미를 분석, 더욱 적합한 검색 결과를 제공
11	대화형 검색	질의응답 형식의 검색, 검색 결과는 시스템 혹은 이용자가 제공함
12	네트워크 검색	개체-관계 모델에 기반한 검색 혹은 사회 네트워크 분석에 따른 검색 결과 제공

번호	속 성	내 용
13	의미 검색	텍스트 구조화에 따라 이용자의 상황, 정황(context)에 맞는 검색 결과 제공
14	인공지능형 검색	단어와 구의 의미를 문맥에 따라 이해하여 사람과 같이 학습, 추론, 지각하고 자연어를 이해할 수 있도록 프로그래밍한 기술
15	인력 검색	사람의 노력과 지식에 근거한 검색 결과 제공
16	전문 검색	특정 분야(취업, 쇼핑, 여행 등)에 특화된 검색서비스 제공 혹은 검색 결과를 그 특성에 따라 특정 분야(지역, 쇼핑 등)로 구분하여 제공
17	시각화	텍스트 기반의 검색 결과가 아니라 비주얼 기반의 내용 검색
18	미리보기	검색 결과의 미리보기 제공, 텍스트 요약이나 사이트의 스냅 샷 제공
19	맞춤 검색	검색 환경 전반을 개인의 취향에 맞게 조절 가능, 주로 웹 형태 검색 엔진을 이용자가 주로 사용하는 사이트에 등록하여 사용할 수 있도록 지원함
20	사람 검색	인물 검색서비스, 특정 분야의 유명 인사나 취업정보와 연계한 구인/구직자 검색 등이 있음
21	추천 엔진/협력 필터링	협력 필터링 기반의 추천 검색 엔진, 유사 성향을 지닌 이용자가 선택한 자원을 추천하거나 이용 패턴 분석을 통해 검색어를 추천함
22	마인드 맵핑(Mind Mapping)	마인드 맵 형태의 검색 결과 제공
23	원클릭 검색 (One-Click, Browserless & Progressive Search)	한 번의 클릭으로 검색 결과로의 접근을 제공하는 데스크 탑 어플리케이션(desktop application)
24	소셜 네트워킹 툴 (Social Networking Tool)	검색 결과를 사회 네트워크 형태, 즉 노드와 링크 형태로 제공(사회 네트워크 분석을 적용하거나 온톨로지에 기반함)
25	FRBR 모델 적용	기존의 목적이 구현하지 못한 서지레코드들 간의 서지적 관계를 논리적으로 연결한 '의미론적 모델', 검색의 관점에서 FRBR 모델은 개체 간 네트워크 구조를 통해 이용자들이 보다 손쉽게 원하는 정보를 찾아갈 수 있도록 지원하는 도구임
26	수학공식/수치문법 적용	모든 단어를 수치로 변환하여 단어와 단어사이의 관계를 수학공식으로 분석, 단어사전과 형태소분석기 등과 같은 언어처리기술을 이용하지 않고 수치문법을 이용
27	텍스트마이닝(Text Mining)	텍스트에 포함된 명시적인 사실의 집합으로부터 메타적인 규칙을 추출
28	연구 엔진(Research Engine)	전문적인 주제 및 내용을 검색하길 원하는 이용자를 위한 검색엔진, 주로 연구자, 학생, 기술 전문가, 교수, 사업가 등을 대상으로 "Deep Web"을 검색하여 제공

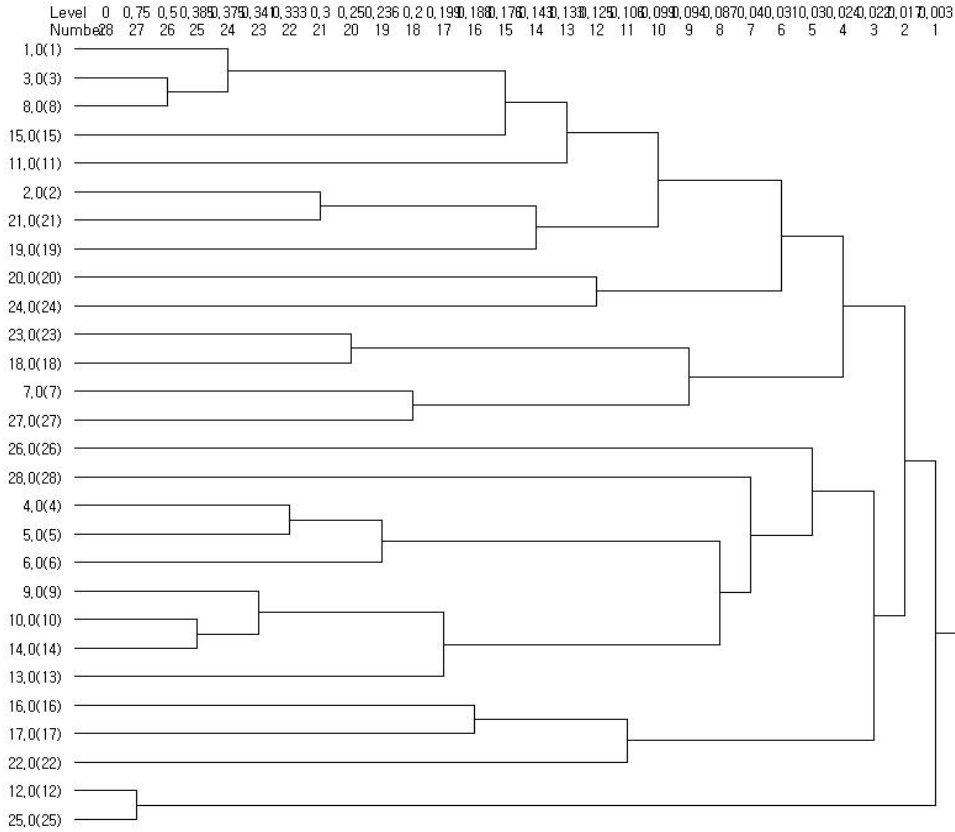
#### 4. 차세대 검색서비스의 군집분석

다음으로 차세대 검색서비스의 군집분석을 위해 검색 2.0의 28가지 속성에 따라 61가지 사례를 재평가하는 과정을 거쳤다. <부록 1>에 나타난 바와 같이 사례가 특정 속성을 나타내고 있을 때에는 '1'로 표시하고, 그렇지 않을 때에는 공란으로 남겨 두었다. 다시 말해 61가지 사례 각각과 28가지 속성에 대한 관계 테이블을

구성한 다음, 속성의 유사성을 기준으로 하여 사례의 클러스터와 속성의 클러스터를 도출하는 두 가지의 군집분석을 실시하였다.

##### 4.1 검색 2.0 속성의 군집분석

<부록 1>을 기반으로 먼저 검색 2.0의 28가지 속성에 대한 군집분석을 실시하였다. 속성 각각의 유사도에 기반하여 군집분석을 수행한 결과는 <그림 2>와 같다.



〈그림 2〉 검색 2.0 속성 간의 덴드로그램

여기에서는 계층적 클러스터링 기법 중 워드 기법을 사용하였고, 클러스터링 집단 개수 결정방법은 스투어즈스(Sturges) 공식, 즉  $L = [1 + \log_2 n]$ 에 따라 총 6개 클러스터로 구분하면 〈표 4〉와 같다.

군집분석의 결과로 나타난 6개 클러스터 중에서 1번, 4번, 6번 클러스터가 뚜렷한 특성을 보이고 있으며, 2번, 3번, 5번 클러스터는 그렇지 않음을 알 수 있다. 특히 1번 클러스터는 ‘소셜 검색’을 필두로 하여 ‘커뮤니티기반 검색’, ‘이용자 협력 검색’, ‘인력 검색’ 등이 포함되어 있다. 이들 모두 개인이 설정한 혹은 개인의 관

심사 등의 프로파일을 기반으로 한 사회적 관계, 특히 사람 관계를 기반으로 한 검색서비스라고 볼 수 있다. 이러한 서비스는 이용자들의 참여가 검색 과정 및 결과 전반에 있어 큰 영향을 미친다고 할 수 있다. 가장 많은 속성의 빈도를 가지는 것을 클러스터의 대표로 선정할 경우, ‘소셜 검색’이 해당된다. 1번 클러스터를 정리하면, 개인이 소속된 커뮤니티 내의 관계 혹은 소속되지는 않았지만 동일한 관심사나 검색 패턴으로 인해 가상으로 맺어진 관계 속에서 검색의 내용과 대상이 달라짐은 물론 연계나 추천을 통해 새로운 결과를 얻게 됨을 알 수 있

〈표 4〉 검색 2.0 속성의 클러스터링 결과

클러스터 번호	속 성
1	소셜 검색, 커뮤니티기반 검색, 이용자 협력 검색, 인력 검색, 대화형 검색, 개인화 검색, 추천 엔진/협력 필터링, 맞춤 검색, 사람 검색, 소셜 네트워킹 도구
2	원클릭 검색, 미리보기, 태깅, 텍스트마이닝
3	수학공식/수치문법 적용
4	연구 엔진, 메타 검색, 군집화, 탭 검색, 자연어 처리 기술, 자연어 검색, 인공지능형 검색, 의미 검색
5	전문 검색, 시각화, 마인드 맵핑
6	네트워크 검색, FRBR 모델 적용

다. 또한 이러한 속성을 가진 검색서비스들은 대부분 개인화 검색, 맞춤 검색을 지향하여 검색 환경 전반을 개인의 취향에 맞게 조절할 수 있도록 지원하고 있다.

4번 클러스터에는 ‘연구 엔진’, ‘메타 검색’, ‘군집화’, ‘탭 검색’, ‘자연어 처리 기술’, ‘자연어 검색’, ‘인공지능형 검색’, ‘의미 검색’이 묶여 있는데, 속성의 빈도를 체크하여 대표명을 부여한다면 ‘인공지능형 검색’ 또는 ‘의미 검색’이 될 수 있다. 1번 클러스터가 검색 대상, 과정, 결과에 〈사람〉을 내세웠다면 이 클러스터는 〈기술〉을 내세우고 있다 할 수 있다. 다시 말해 자연어 질의에 응답하고 동일한 질의어라 하더라도 그것을 사용한 사람의 정황에 맞는 검색 결과를 제시하는 등 검색시스템을 더욱 지능화시키고자 한 노력이다. 각각의 사례가 내세우는 이름은 다를 수 있지만, 그것들이 내포하는 내용은 이와 같이 정리할 수 있다. 군집화와 탭 검색은 수많은 검색 결과를 일정한 기준에 의해 분류하여 이용자가 접근하기 쉽도록 지원하는 서비스로 이 또한(현재까지는) 지능화된 시스템의 결과물이라 할 수 있다.

6번 클러스터는 클러스터 수준을 매우 낮게 보아도 뚜렷하게 구분되는 클러스터로 상기 클

러스터들과 확연히 구분되는 두 가지 특성, 즉 FRBR 모델을 기반으로 하는 검색과 사회 네트워크 분석에 따른 검색 결과를 제공하는 서비스들이다. 한편, 특성이 뚜렷하게 구분되지 않는 2번, 3번, 5번 클러스터가 포함하고 있는 속성 중에는 한 두 사례에서 나타나는 키워드이지만 해당 사례를 대표한다하여 여러 번 언급하고 있어 검색 2.0의 속성으로 추출할 수밖에 없었던 경우가 더러 있다. 따라서 속성 자체에 의미가 없다고 볼 수는 없으나 군집분석의 결과는 잘 나오지 않은 것이 사실이다. 이들의 클러스터 수준을 한 단계 낮추어 보면 각각 1번, 4번, 4번 클러스터에 속하게 되는데, 특히 3번 클러스터에는 ‘수학공식/수치문법 적용’이라는 하나의 속성만 포함하고 있는데 여기에 해당하는 사례는 울프럼알파와 큐로보 두 가지로 이들은 모두 지능형 검색엔진에 포함되는 것들이다. 5번 클러스터 또한 ‘인공지능형 검색’, ‘의미 검색’으로 대표되는 4번 클러스터에 포함되는데 특히 ‘시각화’나 ‘마인드 맵핑’ 기법은 시각화 기법을 적용하여 비주얼 기반의 검색 결과를 제공, 4번의 ‘군집화’와 ‘탭 검색’과 마찬가지로 이용자가 검색 결과에 접근하기 쉽도록 지원하는 서비스이다.

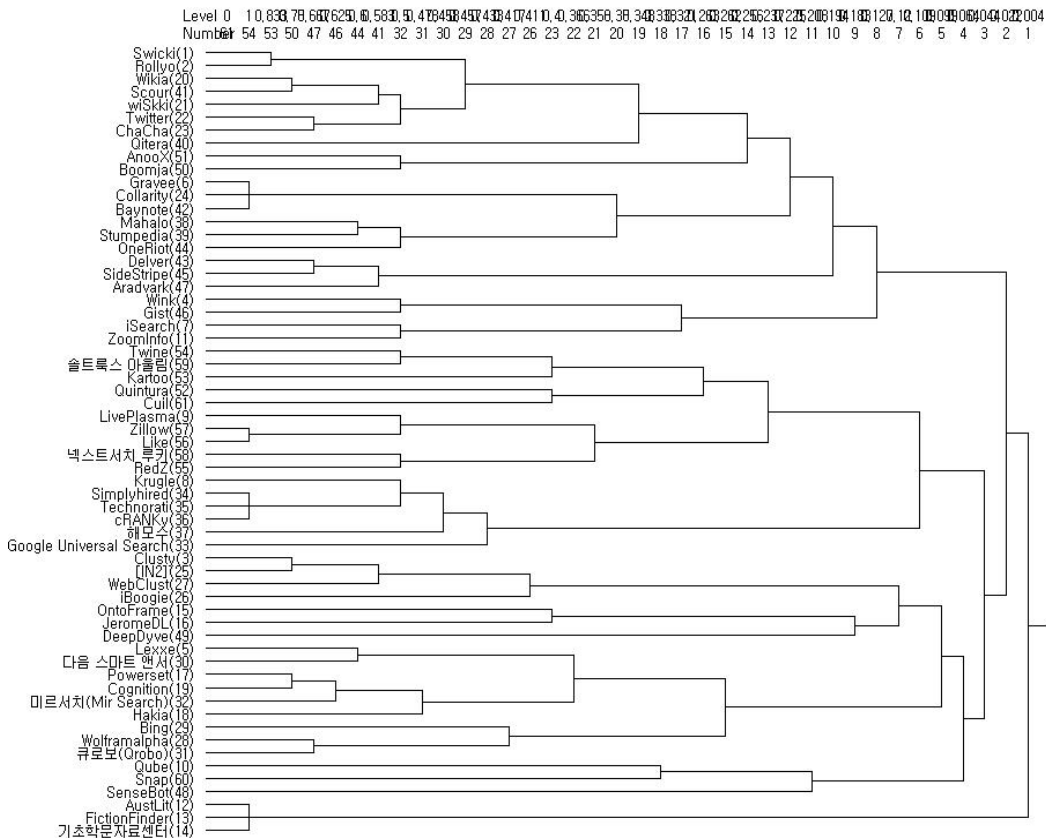
### 4.2 검색 2.0 사례의 군집분석

다음으로 <부록 1>을 기반으로 검색 2.0의 61가지 사례에 대한 군집분석을 실시하였다. 각 사례의 유사도에 기반하여 군집분석을 수행한 결과 아래 <그림 3>과 같이 나타났다. 이 분석에서도 계층적 클러스터링 기법 중 워드 기법을 사용하였으며 클러스터링 집단 개수 결정방법 또한 스투어즈 공식에 따라 총 7개 집단으로 구분하였다. 클러스터링 결과는 <표 5>와 같다.

이 결과는 5번을 제외한 나머지 클러스터가 비교적 뚜렷한 특성을 보이고 있다. 5번 클러스

터에 포함된 사례 또한 앞서와 마찬가지로 여러 가지 속성을 함께 가지고 있거나 한 두 개의 속성을 가지고 있으나 다른 사례와는 패턴이 크게 다른 경우라고 할 수 있다. 이를 제외한 나머지 6개의 클러스터를 하나씩 살펴보도록 하자.

1번 클러스터를 형성하고 있는 사례들을 살펴보면, 이들 사례가 공통적으로 ‘소셜 검색’, ‘개인화 검색’, ‘커뮤니티기반 검색’, ‘태깅’, ‘이용자 협력 검색’, ‘대화형 검색’, ‘인력 검색’, ‘맞춤 검색’, ‘사람 검색’, ‘추천 엔진/협력 필터링’, ‘소셜 네트워킹 도구’를 특성으로 하고 있다는 것을 알 수 있다. 특히 대부분의 사례가 ‘소셜



<그림 3> 검색 2.0 사례 간의 덴드로그램

〈표 5〉 검색 2.0 사례의 클러스터링 결과

클러스터 번호	속 성
1	Swicki, Rollyo, Wikia, Scour, wiSkki, Twitter, ChaCha, Qitera, AnooX, Boomja, Gravee, Collarity, Baynote, Mahalo, Stumpedia, OneRiot, Delver, SideStripe, Aradvark, Wink, Gist, iSearch, ZoomInfo
2	Twine, 아울림, Kartoo, Quintura, Cuil, LivePlasma, Zillow, Like, 루키, RedZ
3	Krugle, Simplyhired, Technorati, cRANKy, 해모수, Google Universal Search
4	Clusty, [IN2], WebClust, iBoogie
5	OntoFrame, JeromeDL, DeepDyve
6	Lexxe, 스마트 앤서, Powerset, Cognition, 미르서치, Hakia, Bing, Wolframalpha, 큐로보(Qrobo), Qube, Snap, SenseBot
7	AustLit, FictionFinder, 기초학문자료센터

검색'을 속성으로 가지고 있어, 앞서 속성에 대한 군집분석 결과가 틀리지 않았다는 것을 증명하고 있다고 할 수 있다. 다시 말해 앞의 1번 클러스터에 포함된 속성이 모두 사례 분석 결과에 해당하는 1번 클러스터에 포함된 사례를 설명하고 있다는 말이다. 정리하자면 〈표 5〉의 1번 클러스터는 소셜 검색을 대표하는 사례들로 하나의 클러스터를 이루고 있는 것이다.

2번 클러스터에 속한 사례들을 살펴보면 이 클러스터에 속한 모든 사례가 검색 결과의 '시각화'를 지원하고 있는 것을 알 수 있다. 그 외에는 몇몇 사례가 '군집화', '전문 검색', '의미 검색' 등의 속성을 가지고 있는데, 결국 이 클러스터는 시각화 기법을 적용한 비주얼 기반의 검색 결과를 제공하는 서비스들로 정리할 수 있다.

3번 클러스터는 '전문 검색'으로 대표되는 사례들의 집합이다. Krugle, Simplyhired, Technorati, cRANKy, 해모수, Google Universal Search, 이 모든 사례가 각각 프로그래밍 소스 검색, 일자리/인력 검색, 블로그 검색 등 검색 대상 혹은 결과를 수평적으로 제공하고자 하는 것이 아니라 특정 주제 영역으로 분할하여 수직적으로 깊이 있게 제공하고자 하는 서비스로

정의할 수 있다.

4번 클러스터는 모든 사례가 '메타 검색', '군집화' 속성을 가지고 있으며, Clusty와 [IN2]가 '탭 검색'을 속성으로 포함하고 있다. 다시 말해 이 클러스터는 다수의 검색 대상을 놓고 통합 검색한 결과를 군집화하여 제공하고 부가적으로 탭 형식으로도 제공하는 검색서비스 사례들로 이루어져 있다는 것이다.

6번 클러스터는 앞의 속성의 군집분석 결과에서 4번에 해당하는 속성들을 포함하고 있는 사례의 집합이다. 즉 여기에 포함된 사례들은 대부분 '인공지능형 검색'과 '의미 검색'을 앞세워 서비스하고 있는 사례들이다. 앞서 언급한 바와 같이 어떤 사례는 자연어 처리 기술을 기반으로 하고, 어떤 사례는 수치 문법에 기반하고 있고, 어떤 사례는 연구 엔진이라는 등 각각의 사례가 내세우는 이름은 다르지만, 결국 지능화된 검색서비스의 구현이라는 가치를 따르고 있는 사례임을 알 수 있다.

마지막으로 7번 클러스터에는 '네트워크 검색', 'FRBR 모델 적용'이라는 두 가지 뚜렷한 속성만을 지닌 사례 세 가지가 포함되어 있다. 속성 분석의 6번 클러스터에 해당하는 것으로



FRBR 모델을 기반으로 기존의 서지 구조를 개체(entity), 관계(relationship), 속성(attribute)이라는 입체적인 구조로 나타내어 이용자들이 원하는 자료를 논리적으로 탐색할 수 있도록 하는 데 목적을 둔 서비스 사례들이다.

## 5. 결론

이 연구를 통해 살펴본 검색 2.0 환경의 주요 개념은 소셜 검색, 인공지능 검색과 의미 검색, 그리고 FRBR 모델과 같은 관계기반 검색 등으로 대표된다 할 수 있다. 이들 개념은 검색의 방법과 결과 처리 등에 있어서 차이가 있다. 먼저, 소셜 검색은 수많은 이용자의 참여를 바탕으로 더욱 이용자 지향적인 검색 방법을 제공하며, 이용자의 상호작용이나 기여가 검색 결과의 적합성을 판단하는 특징이 있다. 인공지능 혹은 의미 검색은 지능 또는 의미에 기반하여 검색을 수행하며, 적합성 판단 기준 또한 알고리즘 혹은 기계 기반 방식을 사용한다. 따라서 이전보다 더욱 지능화된 시스템을 구축하여 이용자의 상황에 맞는 검색 결과 제공하고자 하는 것이다. 이것은 자연어 처리, 정보 분석, 클러스터링, 온톨로지 구현 등의 다양한 기술을 바탕으로 하고 있다.

FRBR 모델과 같은 관계기반 검색은 소셜 검색, 인공지능과 의미 검색을 포괄하는 것으로 눈여겨볼 만하다. 관계기반 검색은 검색 대상인 문헌에 내재된 다양한 속성, 즉 문헌의 형태나 출판 이력, 저자, 조직, 장소, 시대, 주제 등과 같은 다양한 개체들 간의 관계를 파악하고 표현하여 검색에 활용한다는 것이다. 이것

은 문헌으로부터 추출한 관계에 기반하여 매칭과 연계의 알고리즘을 제공하며, 검색 결과 또한 관계를 토대로 시각화하여 보여줄 수 있다. 이처럼 문헌들 간의 사회적, 의미적, 주제적 관계를 분석하여, 이용자가 미처 생각하지도 못한 새로운 문헌을 발견하고 연계하는 정보검색 환경을 제공하게 된다.

따라서 차세대 정보검색서비스는 다음과 같은 특성을 가진 검색서비스로 진화하고 발전할 것으로 판단된다. 첫째, 문헌정보의 생태계적 특성을 반영한 검색이다. 문헌정보 개체들의 생산과 소비의 구조를 반영하여, 각각의 관점에서 문헌들에 내재된 다양한 생태적 관계를 토대로 정보검색 환경이 구축된다. 이러한 정보검색 환경은 하나의 체계로 인식되며 자기조직적이고 역동적인 특성을 나타내 보일 것이다. 둘째, 정보가 내포하고 있는 구조적인 속성을 파악하여 그것을 검색에 반영하게 된다. 문헌정보를 구조와 행위의 관점에서 인식하고, 그것을 관계 또는 네트워크로 재구성하게 된다. 따라서 정보검색은 관계와 네트워크에 배태된 다양한 속성들을 활용하는 알고리즘으로 구현되는 것이다. 마지막으로 정보 분석 방법을 활용한 정보검색이다. 기존의 정보 분석은 인용분석, 공저자분석, 서지결합법, 의미망분석 등 계량정보학 등의 분야에서 제각기 접근하였으며, 이 중에서도 인용 분석 방법을 제외한 분석 방법은 정보검색 영역과는 별개로 다루어진 측면이 많다. 따라서 차세대 정보검색을 위한 모델에서는 정보환경의 생태계적 특성과 구조적 관계를 토대로 다양한 정보 분석 방법들을 통합적으로 수용하고, 이용자가 적절하게 활용할 수 있도록 하는 처리 알고리즘을 개발하여 적용하게 될 것이다.

## 참 고 문 헌

- 김용학. 2003. 『사회 연결망 분석』. 서울: 박영사.
- 노상규, 박진수. 2007. 『인터넷 진화의 열쇠 온틀로지 웹 2.0에서 3.0으로』. 서울: 가즈토이.
- 노지현. 2007. 한국의 도서관 환경에서 FRBR 모델의 의의. 『한국도서관·정보학회지』, 38(2): 224-231.
- 노지현. 2009. FRBR과 차세대 목록. 『2009 춘계 정보관리강좌 회의록』, 2009년 5월 22일.
- 박영진 외. 2006. 웹 2.0과 정보검색. 『정보통신연구진흥원 주간기술동향』, 1251. [인용 2009.10.5]. <<http://comin.itfind.or.kr/itcomp/getFile.htm?identifier=02-001-070601-013904&classCode=undefined>>.
- 서은경. 2006. FRBR 모델에 기반한 서지정보 인터페이스 개발에 관한 연구. 『정보관리학회지』, 23(4): 319-322.
- 송중철. 2001. 특허 및 기술정보의 연계 검색에 관한 연구. 『전자통신분석』, 16(15): 143-150.
- 윤영수, 채승병. 2005. 『복잡계 개론』. 서울: 삼성경제연구소.
- 이경일. 2008. 검색의 혁명 검색 2.0. [인용 2009. 10.26]. <<http://saltlux.egloos.com/849821>>.
- 이수상. 2006. 디지털도서관의 통합검색 방식에 관한 연구. 『한국도서관·정보학회지』, 37(2): 127-144.
- 이순영. 2007. 『지역기반 도서관 포털 구축의 요건에 관한 연구』. 석사학위논문, 부산대학교 대학원, 문헌정보학과.
- 이재윤. 2006. 계량서지적 네트워크 분석을 위한 중심성 척도에 관한 연구. 『한국문헌정보학회지』, 40(3): 191-214.
- 이충무, 이상기, 이병섭. 2009. 추천시스템과 소셜 네트워크를 융합한 지능형 연구자연결망 구축. 『정보관리연구』, 40(1): 199-215.
- 장정훈. 2007. 『네이버 스토리』. 서울: NEWRUN.
- 정영미. 2005. 『정보검색연구』. 서울: 구미무역.
- Bakkalbasi, Nisa and Thomas Krichel. "Patterns of research collaboration in a digital library for Economics." *Proceedings of the American Society for Information Science and Technology*, 43(1): 1-15.
- Ezzy, Ebrahim, 2006. "Search 2.0 vs Traditional Search." [cited 2009.10.5]. <[http://www.readwriteweb.com/archives/search\\_20\\_vs\\_tr.php](http://www.readwriteweb.com/archives/search_20_vs_tr.php)>.
- Kirchhoff, Lars, Katarina Stanoevska-Slabeva, Thomas Nicolai, and Matthes Fleck. 2008. "Using social network analysis to enhance information retrieval systems." *Applications of Social Network Analysis (ASNA)-Zurich*. [cited 2009.6.8]. <<http://www.alexandria.unisg.ch/Publikationen/46444>>.
- Merelo, Juan J. and Carlos Cotta. "Who is the best connected EC researcher? Centrality analysis of the complex network of authors in evolutionary computation." [cited 2009.8.29]. <[http://arxiv.org/PS\\_cache/arxiv/pdf/0708/0708.2021v1.pdf](http://arxiv.org/PS_cache/arxiv/pdf/0708/0708.2021v1.pdf)>.



