

# 디지털 아카이브즈의 문제점과 방향\*

- 문화원형 콘텐츠를 중심으로 -

## Digital Archives of Cultural Archetype Contents: Its Problems and Direction

함 한 희(Han-Hee Hahm)\*\*

박 순 철(Soon-Cheol Park)\*\*\*

### 초 록

본고는 문화원형콘텐츠를 유통시키고 있는 문화콘텐츠닷컴의 디지털아카이브 시스템에 주목해서 문제점을 분석하고 대안을 제시하는 것이 목적이다. 문화원형콘텐츠는 전통문화와 컴퓨터기술을 접목시켜 개척한 새로운 분야이다. 정부에서는 이 산업을 육성해서 한국문화의 세계화와 국가 경쟁력을 강화시킬 의도를 가지고 있다. 우리나라의 역사와 전통, 풍물, 생활·전승·예술·지리지 등 다양한 분야의 문화원형을 디지털 콘텐츠화하여 문화산업에 필요한 창작소재로 제공하는 것이 그 핵심내용이다. 아울러 디지털 콘텐츠 유통체계 정립과 저작권 관리를 통해서 공공부문 문화콘텐츠의 산업적 활용도를 제고하려는 의도도 포함된다. 본고에서 다루는 대상자료는 현재 문화콘텐츠닷컴에서 유통, 관리되고 있는 문화원형콘텐츠들이다. 이 성과물들은 2002년부터 2005년까지 개발되어서 문화콘텐츠닷컴 DB에 구축되어 있다. 이 자료들을 통해서 현재의 디지털아카이브 시스템의 문제점을 분석하였고, 현재의 시스템이 안고 있는 한계점을 요약하면 다음과 같다. 첫째는 각 자료에서 사용하는 주요 용어의 선택에 따라 유사한 자료들이 서로 다른 주제로 분류되면서 다른 항목에 속하게 되는 것이다. 둘째는, 따라서 서로 다른 항목 간에 교차검색이 이루어지지 않는 한계점이 있다. 현재의 제 문제를 해결할 수 있는 방법으로 본고에서는 온톨로지 기능을 포함한 데이터마이닝시스템을 이용해서 풍부한 지식정보표현과 활용이 가능한 디지털아카이브 시스템을 제안하고 있다. 데이터마이닝은 다섯 가지의 방법으로 가능하다. 의미검색, 문서요약, 문서클러스터링, 문서분류 그리고 주제추적이다. 최근에 빠르게 개발되고 있는 디지털 신기술도 인문학과 긴밀하게 연결되지 않으면 그 활용도가 제한적이라는 점을 본고를 통해서 지적하였다. 창작소재로서의 문화원형콘텐츠의 활용도를 크게 향상시킬 수 있는 길은 바로 신지식관리를 위한 통학적(uni-discipline) 접근이라는 점을 일깨우고자 한다.

### ABSTRACT

This is a study of the digital archives of Culturecontent.com where 'Cultural Archetype Contents' are currently in service. One of the major purposes of our study is to point out problems in the current system and eventually propose improvements to the digital archives. The government launched a four-year project for developing the cultural archetype content sources and establishing its related business with the hope of enhancing the nation's competitiveness. More specifically, the project focuses on the production of source materials of cultural archetype contents in the subjects of Korea's history, tradition, everyday life, arts and general geographical books. In addition, through this project, the government also intends to establish a proper distribution system of digitalized culture contents and to control copyright issues. This paper analyzes the digital archives system that stores the culture content data that have been produced from 2002 to 2005 and evaluates the current system's weaknesses and strengths. The summary of our findings is as follows. First, the digital archives system does not contain a semantic search engine and therefore its full function is lagged. Second, similar data is not classified into the same categories but into the different ones, thereby confusing and inconveniencing users. Users who want to find source materials could be disappointed by the current distributive system. Our paper suggests a better system of digital archives with text mining technology which consists of five significant intelligent process-keyword searches, summarization, clustering, classification and topic tracking. Our paper endeavors to develop the best technical environment for preserving and using culture contents data. With the new digitalized upgraded settings, users of culture contents data will discover a world of new knowledge. The technology we introduce in this paper will lead to the highest achievable digital intelligence through a new framework.

키워드: 문화원형콘텐츠, 문화콘텐츠닷컴, 디지털콘텐츠, 디지털아카이브즈, 데이터마이닝, 온톨로지, 의미검색, 통학적 연구, culture archetype contents, culturecontents.com, digital contents, digital archives, data mining, ontology, topic tracking, semantic search

\* 본고는 2006년 정부(교육인적자원부)의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구임(KRF-2006-321-A00012).

\*\* 전북대학교 고고문화인류학과 교수(hanheeh@chonbuk.ac.kr)

\*\*\* 전북대학교 전자정보공학부 교수(scspark@chonbuk.ac.kr)

논문접수일자 : 2006년 11월 26일 논문심사일자 : 2006년 11월 28일 게재확정일자 : 2006년 12월 19일

## 1. 문화원형콘텐츠사업의 특성

문화의 시대인 21세기를 열면서 정부에서는 문화산업화를 통해서 국가의 경쟁력을 키우고자 하고 있다. 그 가운데 문화원형콘텐츠사업의 창안은 국민적인 관심을 받아왔다. IT산업이 세계정상에 올라서면서 문화산업과 컴퓨터 기술을 접목시켜 만든 문화원형콘텐츠사업은 몇 가지 점에서 주목을 받아오고 있다. 하나는 정부가 주도하고 있는 국가적인 사업이라는 점이다. 문화관광부는 문화콘텐츠진흥원 이하 진흥원을 발족시키고, 곧 이어서 2002년부터 문화원형을 발굴하여 디지털콘텐츠화하는 사업을 시작하였다. 두 번째는 전통문화의 산업화를 내세운 점이다. 대중문화의 산업화는 익숙하고 당연해 보이는 반면에 전통문화의 산업화는 다소 생소한 분위기에서 출발하였다. 그러나 진흥원에서는 우리나라의 역사 전통·풍물·생활·전승·예술·지리지 등 다양한 분야의 문화적 내용을 디지털화하여 창작소재의 밑그림으로 사용케 함으로써 전통문화산업을 일으킬 수 있다고 보았다. 세째는 문화원형이라는 개념을 과감하게 사용함으로써 관련 학계와 문화계에 충격을 주었고, 한편 많은 영향을 준 것도 사실이다.<sup>1)</sup> 넷째는 문화콘텐츠의 산업적 활용도를 높이기 위해서 유통체계 정립 및 저

작권 관리체계를 세우고자 한다는 점이다.<sup>2)</sup> 이 사업에는 2006년까지 총 550 억원이 투자될 계획이었고, 2005년까지는 모두 141개의 문화원형사업이 공개경쟁을 통해서 선정되었다.

정부가 막대한 예산을 문화원형사업에 투입하게 된 배경은 취약한 창의적 기획환경 때문이었다. 디지털기술과 정보통신망은 세계적인 수준을 자랑하고, 숙련된 제작인력도 갖추고는 있지만, 풍부한 소재발굴이 턱없이 부족하였다. 그래서 정부는 많은 예산을 들여서라도 디지털 기술을 전통문화에 접목시켜서 창작기반이 되는 소재를 발굴하고, 그것을 문화산업으로 연결시키고자 하였다. 이렇게 시작한 전통문화원형콘텐츠사업은 그동안 성장을 거듭해서 전통문화에 숨어있던 원형적인 요소가 드러나기 시작했고, 그것들이 디지털콘텐츠로 만들어져서 산업현장으로 나갈 태세를 갖추게 되었다.

지금까지 진흥원이 공모한 문화원형콘텐츠는 다음의 4가지 소재를 중심으로 개발되었다.

- (1) 이야기형 소재: 신화, 전설, 민담, 역사, 문학 등의 문화원형 창작소재
- (2) 예술형 소재: 회화, 서예, 복식, 문양, 음악, 춤 등
- (3) 경영 및 전략형 소재: 전투, 놀이, 외교, 교역 등
- (4) 기술형 소재: 건축, 지도, 농사, 어로, 음

1) 본고의 목적이 문화원형이나 문화콘텐츠의 개념규정과 관련된 것이 아니기 때문에 더 이상의 논의를 전개하지 않기로 한다. 대신에 한국문화콘텐츠진흥원에서 규정한 바대로 "문화원형은 민족정체성을 구성하고 있는 집단적 무의식의 내용물이 구체화된 보편적 표상이나 결과물로서의 민족문화로 정의내릴 수 있다"를 잠정적으로 받아들이고자 한다(한국문화콘텐츠진흥원 문화원형사업 설명참조, www.kocca.or.kr). 한편 이와 관련해서는 다음의 글들을 참고할 수 있다: 송현동, 2006, 문화콘텐츠와 한국학, 『종교연구』; 김기덕, 2005, 문화원형의 층위와 새로운 원형개념, 『인문학콘텐츠』, 제 6호.

2) 한국문화콘텐츠진흥원 2005: 22-26, 문화관광부 2005: 100-112, 2003 년도에 수립한 『우리문화원형의 디지털콘텐츠화 사업 종합계획 2003-2007』 과 2005년도에 개발한 2005년 『한국문화콘텐츠진흥원 지원사업 평가지표 개발연구』 참고.

식, 의학 등

이 네 가지 소재를 기초로 해서 개발자들은 디지털콘텐츠를 제작하여 왔고, 그 결과물들은 문화콘텐츠닷컴(www.culturecontent.com)에서 유통되고 있다. 이 곳에는 개발자들이 만든 문화원형 관련 각종 자료들과 개발자들의 홈페이지가 연결되어 있다. 문화원형콘텐츠의 포털 사이트를 표방하면서 그 동안 개발된 자료들을 데이터베이스(DB)로 구축해서 이용자들에게 제공되고 있다. 본고에서는 문화콘텐츠닷컴이 운영하고 있는 디지털자료실 즉 디지털아카이브시스템의 현황에 주목해 보고자 한다. 먼저, 디지털자료실의 기본이 되는 DB 구성이나 체계의 문제점을 지적한 후 그것들을 해결할 수 있는 방안을 고려해 보고자 한다.

## 2. 문화원형콘텐츠 디지털화와 DB

한국문화콘텐츠진흥원에서는 그동안 개발된 자료의 유실을 막고 원활한 자료의 유통관리를 위해서 여러 가지 방법으로 노력해 왔다. 특히 문화원형콘텐츠는 디지털화된 자료의 특성상 온라인에서 관리·유통될 수 있는 환경을 만들 고자 하였으며 현재까지 소정의 성과를 거두었다. 디지털 자료의 접근성과 활용성을 강화했으며 투명한 유통질서를 조성하기 위해서 편리하고 안전한 시스템 인프라를 갖추게 되었다. 그

결과는 '문화콘텐츠닷컴(culturecontent.com)을 통해서 확인할 수 있다.<sup>3)</sup> 이 포털사이트에서는 무엇보다도 먼저 디지털 콘텐츠의 표준화를 이루기 위해서 콘텐츠에 대한 표준메타데이터를 정하고 COI(Content Object Identifier)기반의 저작물 식별체계를 만들어서 확산 보급시키고 있다. COI는 보다 체계적인 온라인과 오프라인 문화콘텐츠 환경을 조성하는데 필수 전제조건이며 나아가서는 국제적 경쟁력을 갖춘 문화콘텐츠 산업으로 성장기 위해서 마련되었다.<sup>4)</sup>

문화콘텐츠닷컴의 특징과 장점이 해당 홈페이지에 다음과 같이 정리되어 있다.<sup>5)</sup> (1) 공공 문화콘텐츠에 특화된 인터넷 포털시스템, (2) 저작권 기반의 문화콘텐츠 전자상거래 시스템 (3) 표준화된 Content Object Identifier 시스템 등이다. 그리고 그 구성은 (1) 공모사업 선정과제를 모은 문화원형 전시관 (2) 여러 과제의 내용을 유형별로 재구성한 창작소재 뱅크 (3) 방대한 메타데이터, 온톨로지 기법을 도입한 첨단 검색기능 (4) 문화콘텐츠 창작자를 위한 커뮤니티이다.

문화콘텐츠닷컴이 현재 제공하고 있는 문화원형자료는 2002년부터 2005년까지 공모를 통해서 선정된 139개 과제의 결과물들이다. 이 자료들을 문화산업분야에서 널리 활용할 수 있도록 문화콘텐츠진흥원에서는 공개의 원칙을 세워 놓고 있다. 그런데, 개발된 자료는 일괄적으로 유료이용의 원칙을 지키고 있어서 공공문화콘텐츠로서 이용률이 떨어지고 있다. 희귀성,

3) 한국문화콘텐츠진흥원에서는 문화콘텐츠닷컴을 구축하기 위해서 약 27억원의 경비가 들었다. (www.culturecontent.com 내용 참고)

4) 한국문화콘텐츠진흥원, COI 핸드북:문화콘텐츠 식별체계(COI)사업, 2006년 4월 12일 Ver 1.0,

5) www.culturecontent.com

독창성의 우선 순위를 매겨서 차별적으로 유료 이용을 유도하는 것이 바람직하다고 생각한다. 다른 사이트에 가면 이용자들이 원하는 유사한 자료들이 무료로 제공되는 경우도 흔하다. 한편, 이용자들이 구매를 하도록 이끌 수 있는 좋은 장치가 필요하다. 이용자들이 하여금 사전에 충분히 검토할 수 있도록 하는 장치가 고안되어야 하고, 또 좋은 자료들을 일목요연하게 찾아볼 수 있는 검색시스템도 마련되어야 할 것 같다. 현재로서는 이용자들이 원하는 자료를 제대로 찾기도 힘들고, 자세하게 살펴볼 수 있는 장치도 부족하다. 문화원형 자료의 가치를 높이기 위해서는 무엇보다도 자료의 정리, 분류 및 이용을 위한 연구가 뒷받침되어야 한다. 본고에서는 개발된 문화원형콘텐츠의 활용도를 높이고 문화산업으로 성장시키기 위해서 바로 이 시점에서 풀어야 할 문제 가운데 하나인 디지털자료의 정리시스템에 대해서 논의를 전개해보고자 한다.

### 3. 디지털아카이브즈의 특성

21세기는 학문영역에 있어서도 커다란 변화가 감지되고 있다. 지식기반에 대한 총체적인 변화가 일어나고 있으며 여기에서의 성패가 국가경쟁력과도 밀접한 관련이 있다고 본다. 그렇기 때문에 지식정보의 창출과 활용문제가 국가적인 과제로 떠오르고 있다. 최근 몇 년 동안 외

국에서는 대학과 시민사회의 문화선점(cultural initiatives) 방향의 중요한 과제로 디지털아카이브즈나 전자문화지도가 속속 개발되고 있다. 이러한 움직임은 컴퓨터기술의 발달과 맥을 같이 한다. 모든 자료가 디지털로의 변환이 자유로워지면서 방대한 자료를 보유하고 있는 아카이브즈계의 변혁은 필연적이다. 각 나라마다 아카이브즈의 디지털화를 위해서 많은 예산과 노력을 투입하고 있다.<sup>6)</sup>

기존의 아카이브즈가 문서 중심의 기록물을 수집, 보관하는 일에 중심을 두었다면, 디지털 아카이브즈는 문서의 영구보존은 물론이고 이용자들의 접근성을 향상시킴으로써 새로운 국면으로 나아가고자 한다. 새로운 기술과의 접목을 통해서 거듭 나게 되는 디지털아카이브즈는 단순히 자료를 모아놓는 보관 창고의 이미지를 넘어서고 있다. 접근성을 향상시킨다는 말은 결국 자료를 효율적으로 이용하게 된다는 뜻이고, 그렇게 함으로써 전과는 다른 차원의 아카이브즈로 거듭나게 된다는 것이다. 사실 유럽이나 미국처럼 아카이브즈의 역사가 오래된 곳에서조차 디지털아카이브즈를 구축하는 일은 현재로서는 초기단계이고 전자문서의 관리, 시스템 개발, 메타데이터, 전자보존과 진본성의 문제에 대해서 활발하게 토론이 진행되고 있다.<sup>7)</sup> (박은정 2004).

한국의 아카이브즈 역사는 짧지만 디지털아카이브즈는 새로운 시도이기 때문에 다른 나라들과 어깨를 겨루며 나아갈 수 있는 좋은 기회

6) 과거에는 연구와 학습의 기본대상이 대체로 문서 중심의 자료였다. 그러나 이제는 문서만이 아니라 음성, 이미지 등을 포함하는 다양한 형태의 자료가 만들어지며 또 손쉽게 디지털 자료로 환원되고 있다. 이렇게 만들어지는 멀티미디어자료의 효율적인 관리가 시급해지고 있는 것이다.

7) 2004년 10월 서울에서 열린 ICOM 2004 Seoul 대회에서 미국, 일본, 및 유럽의 아카이비스트나 박물관 큐레이터들이 디지털아카이브즈의 필요성에 대해서는 한목소리를 냈으나 나라마다 예산이나 인력 확보에서는 차이를 보였다.

이다. 단 이 과정에서 핵심은 기술적인 선점에 있기 보다는 원본성, 진본성을 유지하고 완성도가 높은 메타데이터의 구축 등과 같이 내용을 충실하게 만드는 일이 더 중요하다. 보관 자료를 디지털화시킨 것만으로 디지털아카이브즈가 만들어진 것은 결코 아니다. 새로운 차원의 자료수집, 분석, 분류체계의 연구 등과 같이 심도있는 체계적인 연구도 필요하다.

위에서 언급한 바와 같이, 디지털화된 자료의 접근성이 향상되었다고 하는 것은 지식이용의 차원이 달라진 것을 의미한다. 지식과 정보의 이용자층이 두터워졌을 뿐 아니라 이용방법도 크게 변화하고 있다. 인터넷이 발달하면서 디지털 자료의 대부분이 웹상에서 구현되고 또 상호작용모드로 바뀌면서 각 가정에서도 국가나 기관의 아카이브즈에 접속이 가능해졌다. 과거에는 아카이브즈의 방문객만이 직접 자료를 볼 수 있었지만, 디지털화된 자료는 그러한 수고를 덜 수 있다. 언제 어디에서고 누구를 막론하고 열람할 수 있는 아카이브즈는 지식의 공유화를 재촉하고 있다. 새로운 차원으로 전개되고 있는 디지털자료의 이용에 대한 정확하고도 체계적인 기본적 연구가 절실히 요구되고 있다.

일찍부터 문화콘텐츠진흥원에서는 디지털자료의 DB화 및 자료관 운영의 필요성을 가지고 일을 추진해 온 것은 다행한 일이다. 그러나 디지털 아카이브즈가 지식의 공유화를 촉진하기 위해서 이용자의 입장을 중시한다기 보다는 자료 생산자의 편의성만 따져서 구축되고 있다는 점이 안타깝다. 자료관의 외형은 화려하고 높

은 기술수준을 가지고는 있지만, 정작 이용자들이 필요로 하는 정보를 찾기에는 부족한 점이 많다고 느껴지게 된다. 이 점을 보완하기 위해서는 이용자들의 입장에 서서 접근성을 높이는 방법의 연구가 이루어져야 한다.

다시 말해서 디지털시대의 아카이브즈에서는 이용자들이 자료를 충분히 활용할 수 있도록 하는 장치에 대한 연구가 필요하다는 뜻이다. 심도있고 체계적인 연구를 통해서 자료에 대한 접근성을 높이면 이용자들의 창의성을 자극하는 일은 쉬어진다. 이용자들이 주도적이고 적극적일 때 새로운 지식창출의 가능성이 높아지게 되는 것 또한 당연하다(함한희 2006, 22-25). 디지털화된 자료들은 그 특성상 자료와 자료사이를 넘나들기 쉬우며, 하나의 자료가 나누어지면서 다른 자료와 합쳐지면 새로운 자료가 된다. 이러한 과정 속에서 신지식의 터전이 다져지게 된다. 디지털의 속성이 충분히 자료관에서 발휘될 때, 이용자의 저변도 확대된다. 물론 전자문서의 관리 진본성 여부 등의 윤리적이고 법적인 문제들도 나타나겠지만, 이러한 단점에도 불구하고 디지털아카이브즈의 확산은 시대적인 사명이다.

#### 4. 문화원형콘텐츠 디지털아카이브즈의 문제와 방향

문화원형콘텐츠닷컴의 DB에는 우리나라의 역사와 문화에 유익한 자료들이 디지털로 가공되어서 상당 수 축적되어 있다. 이용자들이 이

최근(2006.10) 본 연구자는 프랑스와 벨기에의 국립아카이브즈에서 진행시키고 있는 자료의 디지털화작업의 현황을 파악하고 돌아왔다. 두 나라 모두 국가가 장기적인 사업으로 추진 중에 있다.

자료를 쉽고도 친근하게 이용할 수 있도록 하는 방법을 모색하는 것이 본고의 목적이다. 이용자 중심의 디지털아카이브즈가 되려면 우선적으로 이용자에 입장에 서서 자료를 제공하는 정책방향이 필요하다. 이용자들이 원하는 것이 어떤 것인가에 대한 고려가 필요하다는 뜻이다. 문화원형콘텐츠의 생산자의 입장에서는 자신들이 공들여서 만든 마지막 결과물만을 제공하는 것으로 그 소임을 다했다고 생각할 수 있다. 그러나 아카이브즈에는 1차 자료에 해당하는 자료들이 수집되어 있어야만 본래의 설립목적에 부합한다. 각 과제마다 독특한 주제와 내용을 가지고 한국의 문화원형을 발굴하였기에 연구과정에서 나온 원천자료도 중요한 자료로 취급되어서 아카이브즈에 보관되어야 한다. 문화원형에 관심이 있는 문화산업에 종사하는 사람들, 여러 분야의 전문가, 연구자, 학생 그리고 교육자들은 마지막 결과물뿐만 아니라 문화콘텐츠의 생산과정에 대한 관심 또한 지대하다. 그러므로 문화원형콘텐츠 개발의 전 과정 자체를 자료화하여 DB에 보관하는 일은 중요하다. 개발과정에서 나온 회의록, 공문, 메모, 이메일을 포함하는 서신 등도 따로 보관해 둘 필요가 있다. 최종결과물만이 중요한 것은 아니라는 점은 분명하다. 다양한 범주의 원천자료로 불릴 수 있는 것들 문화원형 원자료와 개발과정을 보여주는 기초자료들을 포함해야만 비로소 문화원형콘텐츠 자료실은 이용자들의 창의성이 개발될 수 있는 곳으로 올라서게 된다.

둘째로 문화원형콘텐츠가 이용자들을 적극적인 지식생산의 주체로 전환시킬 수 있어야 한다. 즉 이용자들이 기초자료에 해당하는 문화원형콘텐츠로부터 자극을 받고 스스로 지식산업

의 생산자가 될 수 있어야 한다. 이것이 진흥원이 추진하는 최종목적이기도 하다. 그러기 위해서는 두 가지가 필요하다. 하나는 이용자들이 편리하게 이용할 수 있는 시스템의 창안이고, 다른 하나는 이들이 현재 유통되고 있는 자료를 통합적으로 접근해서 새로운 차원의 아이디어를 만들어내는 적극적인 생산자로 전환시킬 수 있는 시스템의 구현이다. 과거에는 책이나 논문을 쓰는 학자, 예술작품을 만드는 예술가 등이 지식생산의 주체가 되었다. 그러나 디지털세계에서는 새로운 계층의 지식생산자들의 탄생을 예고하고 있다. 문화원형콘텐츠의 이용자들도 이제는 더 이상 소극적인 수용자가 아니라는 점을 인식하는 것이 중요하다. 그리고 이들이 현재 만들어진 문화원형콘텐츠를 단순가공하는 차원에 머물기 보다는 새로운 것으로 창조하는 차원으로 유도하는 것이 바람직하다.

지식생산의 저변이 크게 확대되었고, 그 경계도 사라졌다. 특히 인터넷환경에서는 지식생산과 소비를 구별하는 의미가 크지 않다(피에르 레비 2000, 218-230). 그렇다면 이렇게 저변이 확대된 지식생산자들을 겨냥해서 새로운 지식을 창출할 수 있도록 유도하는 방법이 구체적으로 어떤 것일까를 생각해 보기로 한다.

## 5. 문화원형콘텐츠 분류와 검색의 문제점

### 5.1 자료정보 제공의 한계점

본 연구자는 한국의 도깨비를 소재로 시나리오를 쓰고 싶은 이용자를 가정해서 문화콘텐츠

닷컴을 방문해 보았다. 제일 처음 한 일은 문화콘텐츠닷컴의 전체구조를 파악하는 일이었다, 그 결과 여느 포털사이트와 크게 다른 점을 발견할 수는 없었다. 그 다음으로 한 일은 통합검색을 통해서 '도깨비'를 검색하였다. 도깨비검색 결과는 총 6,012건이다.<sup>8)</sup> 그 가운데 콘텐츠 검색은 2,975건이고, 사이트는 1곳, 이미지는 2,863건이다. 이 콘텐츠는〈한국도깨비 캐릭터 이미지 콘텐츠 개발과 시나리오 제재유형개발〉과제의 결과물이다.<sup>9)</sup> 이 과제는 한국도깨비에 대해서 자세한 서술과 애니메이션을 약 5,000건 이상 개발하였다(본고의 문화원형과제분류표 1의 3-10, 표1 참고). 개발된 도깨비를 보다 상세히 보려면 '자세히'라고 하는 곳으로 가도록 되어 있다. 그러나 막상 '자세히'라는 곳에는 자세한 설명은 없고, 샘플보기와 메타데이터만 보인다. 더 자세하게 보려면, 해당 도깨비에 대한 자료를 구입해야만 한다. 가격은 시나리오 제재(식별번호: cp0313000269)는 81,500원이다. 낮도깨비-도깨비 보물얼가 도깨비 감투류-1(식별번호: cp0313000474)는 35,000원이다. 도깨비 보물얼기(식별번호: cp0313000276)를 모두 구입하려면 4,731,500원이 든다.

이용자로서 본 연구자가 이 자료들을 과연 구입할 수 있을지, 막상 어떤 자료를 구입해야

하는지 등 망설여지는 이유들은 다음과 같다. 첫째는 도깨비관련 콘텐츠의 가격이 지나치게 높다는 점이다. 둘째는 이용자들이 구입을 결정할 수 있는 정보가 크게 부족하다는 점이다. 도깨비에 대한 자세한 내용을 알아야만 구입 여부를 결정짓게 되는 것은 당연하다. 아마도 간단한 샘플보기만을 보고 구입할 수 있는 이용자들은 극히 드물 것이다.<sup>10)</sup> 이용자가 도깨비를 시나리오에 사용할 수 있을 만큼의 기본적인 정보는 주어져야 한다. 더 바람직한 것은 최종적으로 구입을 결정했을 때까지 무제한 사용을 허락하는 것도 공공부문 문화콘텐츠의 진가를 높이는 것이라고 생각된다. 한편, 이용자들에게 주어진 정보 가운데 메타데이터도 소중한 자료가 된다. 그럼에도 불구하고 문화콘텐츠닷컴에서 제공하고 있는 디지털자료의 메타데이터에는 몇 가지의 기본적인 정보가 들어있지 않다. 자료가 제작된 시기, 자료의 길이(도큐먼트), 크기(이미지), 상영시간(동영상) 등의 정보가 있어야만 이용자들은 그 자료에 대한 개략을 짐작하게 된다.

## 5.2 의미검색의 필요성

본 연구자는 도깨비에 대한 자료를 찾다가

- 
- 8) 본고를 작성하면서 문화콘텐츠닷컴의 자료를 2006년 10월 말부터 11월 20일까지 살펴보았다. 자료의 검색결과와 통계수치 등은 이 시기 이후와는 차이가 있을 수 있음을 밝힌다. 사이트의 운영자들이 자료를 추가하거나 변경할 수 있기 때문이다.
- 9) 제작기관, (주)네오그래프, <http://dokkaebi.culturecontent.com>. 도깨비 설화에 들어가 있는 한국도깨비의 역사와 문화체험하기, 등록일 2004년 11월 11일. 본 발표를 위해서 재미있는 주제의 하나로 선택하였을 뿐 고의적으로 이 사이트나 개발자를 폄하하려는 의도는 없음을 밝힌다.
- 10) 샘플보기의 내용을 그대로 옮기면, "시나리오 제재 유형 자료 제어번호cp0313t30192제 제명낮도깨비 - 도깨비 보물 얼기 - 도깨비 감투류 - 1 캐릭터 도깨비 종류 낮도깨비 보조인물 게잡는 사람 성격 장난 배경시간 낮공간 개울가서사 규칙 도깨비가 조화를 부린 것은 시간이 지나고 나면 정체가 드러난다. 도깨비감투의 덕을 사람이 영원히 누리지는 못한다."

유사한 과제가 있음을 발견했다. 표 1의 9-5, '인귀세상'이 바로 그것이다. 그리고는 검색창을 통해서 찾아보려고 하였다. 통합검색에서 '인귀세상'을 넣고 검색을 하면 관련 자료가 나타난다. 그러나 도깨비를 가지고는 '인귀세상'을 찾을 수 없다. 그 역도 마찬가지이다. 의미검색시스템이 작동되지 않기 때문이다. 인귀세상은 <한국 인귀설화의 원형 콘텐츠 개발> 과제로 2005년 11월 11일에 등록되었다. 귀신에 대한 자료를 토대로 콘텐츠를 개발한 것이다.<sup>11)</sup>

이용자들의 입장에서 문화원형자료를 제대로 활용하기 위해서는 '한국의 도깨비'와 '인귀세상'이 동시에 검색된다면 매우 편리하고 유용할 것이다. 도깨비와 인귀가 죽음이나 민속 신앙 그리고 구비설화로 전해지는 문학이라는 공통점이 있다. 의미검색(semantic search)이 가능하다면 도깨비와 인귀를 서로 비교해가며 둘의 유사점, 차이점, 다양성, 특수성, 일반성 등을 비교분석하여 한국의 도깨비를 소재로 창의적이고 재미있는 시나리오를 쓸 수 있을 것이다. 뿐만 아니라, 죽음, 상례, 한옥(변소, 장독대), 민화, 부적, 탈을 다룬 과제들도 도깨비나 귀신과 관련된 내용을 포함하고 있어서 의미검색의 필요성은 더욱 절실해진다. 문화콘텐츠닷컴이 이러한 시스템을 갖추게 되면 문화원형에 대한 총체적이고 종합적인 정보를 제공하는 곳으로 각광을 받게 될 것이다.

본 연구자는 문화콘텐츠닷컴에 올라와 있는 문화원형과제들을 현재 진흥원이 분류한 9가지

주제 항목에 따라서 정리해 보았다. 각 분류항목의 문화콘텐츠과제의 수는 아래와 같다(개발된 번호는 편의상 필자가 붙여놓은 것임). 문화콘텐츠분류를 자세히 들여다보면 기준이 모호하다는 것을 쉽게 발견할 수 있다. 어떤 과제는 4-5개로 중복 분류되기도 하고, 어떤 과제는 전혀 그렇지 못하다. <디지털 악학궤범>과제의 경우, 의식주, 의례/신앙, 역사/민족, 예술, 문학/문헌, 5개 주제에 동시에 포함되어 있지만, 한국의 도깨비는 의례/신앙에만, 인귀세상은 문학/문헌에만 속해 있다.<sup>12)</sup> 상식적으로도 한국의 도깨비는 문학/문헌에도 속해야 하고, 인귀세상은 의례/신앙에도 속해야 한다. 기준이 없는 분류체계는 이용자들의 창의적 사고를 가로 막으며 문화원형콘텐츠의 활용도를 떨어뜨리고 있음을 알 수 있다.

문화원형콘텐츠를 자료의 형태를 기준으로 해서 분류해 놓은 것에서도 역시 문제가 나타난다. 문화콘텐츠닷컴에서 자료의 소스는 디자인, 멀티미디어, 오디오, 그리고 도큐먼트로 분류하였다. 각각의 소스는 다시 특성에 따라서 세부분류항목으로 나누어진다(표 2 참조). 각 세부분류 항목은 인물/캐릭터, 동물/식물, 전설/신화/신앙, 의복/장신구/물건, 민속/놀이, 역사/군사/사건, 사찰/건축/도면, 천문/지리의 주제로 나뉜다.<sup>13)</sup> 이러한 분류를 뛰어넘어서 이용자들이 자유롭게 소스간, 세부항목간, 주제별 답을 넘나들 수 있을 때 문화원형자료는 정보로서의 가치가 크게 올라갈 수 있을 것이다.

11) <http://koreaghost.culturecontent.com>에 자세한 내용이 수록되어 있다. 문화콘텐츠닷컴의 통합검색에서는 콘텐츠검색 0, 이미지 검색 0, 사이트검색 1건으로만 나온다.

12) 많은 사례를 일일이 다 소개하는 것은 그다지 의미가 없다고 생각해서 생략한다.

13) 오디오소스는 예외이다. 그리고 각 소스의 세부항목에서도 약간의 예외는 있다(표 2 참조).

(표 1) 문화원형콘텐츠과제의 주제별 분류

주제별과제	과제수	과제내용	중복분류 및 분류상 문제
1 의식주	16	(1) 전통고기잡이	
		(2) 사냥	
		(3) 전통머리모양과 머리치레거리	
		(4) ****	
		(5) 죽음의 전통의례와 상징세계	3-7/9-9
		(6) 조선시대 동물화첩	분류상 문제 3-8/7-17
		(7) 길상이미지	6-20
		(8) 디지털 약학계법	3-11/6-21/7-20/9-10
		(9) 조선시대 식문화원형	3-12
		(10) 조선후기 시장	6-22
		(11) 상인과 상업활동	4-2/5-5
		(12) 전통지수문양	
		(13) 한국의 고유복식	7-34
		(14) 전통건축과 장소	2-10
		(15) 고려복식	
		(16) 사진으로 보는 한국전통건축	2-11
2 건축	11	(1) 근대서울의 공간	
		(2) 궁궐의례와 공간	3-4/6-11
		(3) 조선시대 수영의 디지털 복원	
		(4) 목조건축 부재별 조합 조립체험	
		(5) 사찰 건축 디지털세트 절집	
		(6) 한국의 고인돌	6-23/7-21
		(7) 디지털한양	4-6/8-6
		(8) 사이버 전통 한옥마을	6-33
		(9) 단청문양	3-18/7-31
		(10) 전통건축과 장소	1-14
		(11) 사진으로 보는 한국전통건축	1-16
3 의례/신앙	22	(1) 한국의 굿	
		(2) 불교설화	
		(3) 건국설화 이야기	
		(4) 궁궐의례와 공간	2-2/6-11
		(5) 강릉단오제	
		(6) 우리문화 상징 아이콘 부적	6-15/7-16
		(7) 죽음의 전통의례와 상징세계	1-5/9-9
		(8) 조선시대의 동물화첩	1-6/7-17
		(9) 조선왕실의 관혼상제	6-18
		(10) 한국의 도깨비	
		(11) 디지털 약학계법	1-8/6-21/7-20/9-10
		(12) 조선시대의 식문화 원형	1-9
		(13) 감로탕	6-25/7-23
		(14) 바리공주 서사창작	9-14
		(15) 종묘제례악	
		(16) 오색채운	7-29/9-18
		(17) 오방대제	6-32/9-19
		(18) 단청문양	2-9/7-31
		(19) 고려 팔관회	6-35/9-21

		(20) 토정비결과 한국인	9-23
		(21) 한국의 불화	7-32
		(22) 디지털 민화	6-36
		(23) 신화의 섬, 제주	6-43/9-27
4	교통/통신	6	(1) 문화와 산업의 대동맥 옛길
			(2) 상인과 상업활동
			(3) 대동여지도
			(4) 한국의 배
			(5) 고선지 실크로드 개척사
			(6) 디지털한양
			1-11/5-5
			8-5/9-16
			5-6/6-30
			5-8
			2-7/8-6
5	군사/외교	10	(1) 한국의 산성
			(2) 태견의 이야기와 동작
			(3) 조선시대 수영의 디지털복원
			(4) 한국무예 원형 및 무과시험 복원
			(5) 상인과 상업활동
			(6) 한국의 배
			(7) 진법
			(8) 고선지실크로드 개척사
			(9) 조선의 국왕경호
			(10) 한민족 전투원형
			1-11/4-2
			4-4/6-30
			9-17
			4-5
			6-31
			6-40
			(1) 천하명산 금강산
			(2) 조선시대 아동교육
			(3) 전통놀이와 춤
			(4) 해동성국 발해
			(5) 백두산 문화상징 디지털콘텐츠화
			(6) 1910년, 과거로 가는 시간여행
			(7) 기생사랑
			(8) 고려사담
			(9) 한국인 얼굴유형
			(10) 신여성문화
			(11) 궁궐의례와 공간
			(12) 독립신문화 만민공동화
			(13) 조선시대 형구와 형벌이야기
			(14) 궁중문양
			(15) 우리문화 상징 아이콘, 부적
			(16) 화랑세기 속의 신라 화랑
			(17) 죽음의 전통의례와 상징세계
			(18) 조선왕실의 관혼상제
			(19) 조선시대 기녀문화
			(20) 길상이미지
			(21) 디지털 약학궤법
			(22) 조선후기 시장
			(23) 한국의 고인돌
			(24) 국악
			(25) 감로탕
			(26) 한국무예 원형 및 무과시험복원
			(27) 한국천문 우리하늘 우리별자리
			(28) 한국의 탈, 탈춤
			(29) 풍속화콘텐츠
6	역사/민족	43	
			7-7
			2-2/3-4
			9-6
			3-6/7-16
			3-9
			1-7
			1-8/3-11/7-20/9-10
			1-10
			2-6/7-21
			3-13/7-23

		(30) 한국의 배	4-4/5-6
		(31) 조선의 국왕경호	5-9
		(32) 오방대제	3-17/9-19
		(33) 사이버 전통 한옥마을	2-8
		(34) 압행어사	
		(35) 고려 팔관회	3-19/9-21
		(36) 디지털 민화	3-22
		(37) 소리은행	
		(38) 화성의계	
		(39) 검안	
		(40) 한민족 전투원형	5-10
		(41) 전통놀이	
		(42) 한국신화	
		(43) 신화의 섬, 제주	3-23/9-27
		(1) 저작권위탁관리(한예중)	
		(2) 저작권위탁관리(국립국악원)	
		(3) 저작권위탁관리(민속박물관)	
		(4) 한국의 전통장신구	
		(5) 한국근대의 음악원형 디지털화	
		(6) e-조선궁중여성	
		(7) 전통놀이와 춤	6-3
		(8) 거례의 노래 아리랑	
		(9) 한국의 전통 춤	
		(10) 음성원형콘텐츠 웨어	
		(11) 한국의 암각화	
		(12) 디지털 한국석탑	
		(13) 유랑예인집단 남사당	
		(14) 현대 한국대표 서예가의 디지털폰트	
		(15) 중국 집대성문학의 환타지 '태평광기'	
		(16) 우리문화 상징 아이콘 부직	3-6/6-15
		(17) 조선시대 동물화첩	1-6/3-8
		(18) 조선시대 기녀문화	
		(19) 길상 이미지	
		(20) 디지털 악학계법	1-8/3-11/6-21/9-10
		(21) 한국의 고인돌	2-6/6-23
		(22) 국악	
		(23) 감로탕	3-13/6-25
		(24) 한국의 탈, 탈춤	
		(25) 연오랑과 세오녀	9-13
		(26) 종묘제례악	
		(27) 풍속화콘텐츠	
		(28) 디지털폰트	
		(29) 오색채운	3-16/9-18
		(30) 전통국악기 샘플데이터	
		(31) 단청문양	2-9/3-18
		(32) 한국의 불화	3-21
		(33) 디지털민화	
		(34) 한국의 고유복식	1-13/3-18
		(35) 소리은행	
7 예술	35		

8 과학기술	8	(1) 전통 한의학 및 한약재	
		(2) 한국의 풍수지리	
		(3) 조선 궁중 과학기술관-천문	
		(4) 한국천문 우리하늘 우리별자리	9-12
		(5) 대동여지도	4-3/9-16
		(6) 디지털한양	2-7/4-6
		(7) 화성의궤	9-24
		(8) 겸안	9-25
9 문학/문헌	27	(1) 한국고서의 표지문양 및 장정	
		(2) 조선후기 여향문화	
		(3) 조선 한의학과 한약재	
		(4) 새롭게 펼쳐지는 신화의 나라	
		(5) 인귀세상	
		(6) 조선시대 형구와 형벌이야기	6-13
		(7) 중국 집대성문학의 환타지 '태평광기'	
		(8) 화랑세기 속의 신라 화랑	
		(9) 죽음의 전통의례와 상징세계	1-5/3-7
		(10) 디지털 약학개범	1-8/3-11/6-21/7-20
		(11) 한국무예 원형 및 무과시험 복원	
		(12) 한국천문 우리하늘 우리별자리	8-4
		(13) 연오랑과 세오녀	7-25
		(14) 바리공주 서사창작	3-14
		(15) 디지털 폰트	
		(16) 대동여지도	4-3/8-5
		(17) 진법	5-7
		(18) 오색채운	3-16/7-29/9-18
		(19) 오방대제	3-17/6-32
		(20) 압행어사	
		(21) 고려 팔관회	3-19/6-35
		(22) 조선시대 대하소설	
		(23) 토정비결과 한국인	3-20
		(24) 화성의궤	8-7
		(25) 겸안	8-8
		(26) 한국신화	
		(27) 신화의 섬, 제주	3-22/6-43

(표 2) 문화원형자료 소스 통계표

○ 오디오소스

음 악		음 향		음 원	
악보/가사	1,912	악기소리	2,678	디지털음원	3,084
연주곡/반주	218	자연소리	1		
목소리/무반주	200	사람소리	0	효과음	13
노래/가요	0				
합 계	2,330	합 계	2,679	합 계	3,097

○ 디자인소스

	일러스트	포 토	문 양	
			옷문양	
인물/캐릭터	1,161	5,692	옷문양	0
동물/식물	1,063	4,009	건축문양	978
전설/신화/신앙	3,267	3,633		
의복/장신구/물건	289	11,008	자연	3,091
민속/놀이	2,479	1,463	동물/식물	1,740
역사/군사/사건	262	2,234		
사찰/건축/도면	556	14,915	현대식문양	1,063
천문/지리	274	1,787		
합 계	9,351	44,741		

○ 멀티미디어 소스

VR(가상체험)		VOD 동영상		Flash animation		Cell animation	
인물/캐릭터	9	인물/캐릭터	350	사찰/건축/도면	544	0 (카테고리 없음)	
의복/장신구/물건	12	단편영화	0				
		민속/놀이	245				
역사/군사/사건	1	역사/군사/무술	122	전통상장례	0		
		음식	68				
사찰/건축/도면	51	역사/지리	104	전설/신화/신앙	290		
		자연	0				
천문/역사/지리	28	동물/식물	0	동화	0		
		전설/신화/신앙	154				
전설/신화/신앙	2	생활	153	민속/놀이	96		
		천문/역사	93				
민속/놀이	0	합 계	1,289	합 계	930		
합 계	103	합 계	1,289	합 계	930		

○ 도큐먼트 소스

	시놉시스	번역서	해설/분석서	실화/기록서	이야기	시나리오	창작/소설
인물/동물/식물	0	0	371	1,881	278	344	0
사찰 / 건축	3	0	290	2	12	0	0
천문 / 지리	0	0	370	343	3	0	0
전통문화 / 생활	0	1	6,883	119	257	22	5
역사 / 사건	17	0	193	260	7	2	72
무 술	0	0	13	395	0	0	0
합 계	20	1	8,120	3,000	557	368	77

6. 디지털아카이브즈의 새로운 모델:  
통합적 분석시스템의 창안

문화원형콘텐츠는 현재 두 가지 방법으로 분류·정리되어있다. 하나는 자료의 주제별 분류

이고(표 1), 다른 하나는 자료의 유형별 분류(표 2)이다. 현재 이용자들이 느끼는 불편함은 의미검색 시스템개발의 한계리는 점을 앞의 절에서 지적하였다. 더욱이 문화원형콘텐츠를 애써서 개발한 개발자들의 노력을 무력하게 만드

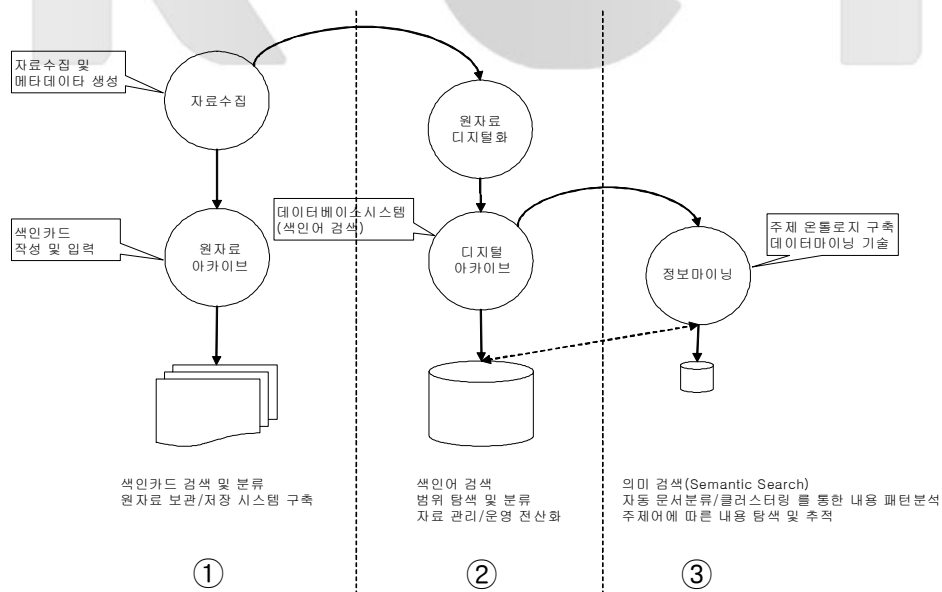
는 측면도 있다. 문화원형콘텐츠를 통해서 창의적인 사고를 불러일으키려면 현재와 같이 일정하게 틀에 박힌 구분법은 바람직하지 않다. 그래서 이용자들의 상상력을 저해하는 경우도 적지 않다. 예를 들면, 조선시대식문화원형(표 1의 1-19과제)은 조선시대 사람들과 가옥, 궁궐(표 1의 2-1, 2, 8과제), 그리고 시장(표 1의 1-10과제) 등과 만나야만 비로소 종합적인 시나리오, 연극, 영화 등의 창작물이 만들어질 수 있다. 그런데 현재의 아카이브시스템에서는 이러한 통합적인 검색이 불가능하다.

본 연구자는 창의성을 유도하는 통합적 방법에 대해서 간략하게 설명해보고자 한다. 그림 1은 현재 다른 과제에서 실험적으로 진행하고 있는 디지털아카이브화 과정을 간략하게 표현한 것이다.<sup>14)</sup> 문화콘텐츠닷컴의 디지털아카이브

는 ①과 ②의 일부만 구축된 상태이기 때문에 ③의 과정을 추가적으로 개발할 필요가 있다. 세 번째 단계의 디지털아카이브과정은 온톨로지 구축과 데이터마이닝이라고 하는 새로운 기술의 도입을 의미한다.

### 6.1 데이터 마이닝의 필요성

문화원형자료가 디지털아카이브를 통해서 그 가치를 되찾기 위해서는 무엇보다도 데이터 마이닝의 과정을 거쳐야 한다. 사람들이 일일이 손으로 항목을 지정해서 분류한 후 데이터 베이스에 저장해 두었다가 검색하는 방법도 있지만, 컴퓨터가 자료사이에 숨어있는 새로운 자료를 찾아내는 작업을 병행시키면 매우 유용한 결과를 얻게 된다. 이용자들은 모든 문서를



(그림 1) 디지털아카이브의 구축 방법과 과정

14) 현재 진행 중인 과제는 '온톨로지 구축을 통한 인지인류학적 연구방법의 개발'이다.

일일이 읽지 않아도 문서 내에 숨겨진 정보들을 쉽게 찾을 수 있다.<sup>15)</sup> 그 과정은 컴퓨터가 자료들을 새로운 평가에 의해서 중요도를 매기고, 또 분류하기 때문이다. 그 결과 지금까지와는 다른 새로운 분류체계를 만들어낼 수도 있다. 또한 이용자들이나 연구자들은 지금까지 쉽게 알아차릴 수 없었던 새로운 내용의 패턴을 발견하게 될 가능성이 높아진다. 이 기술의 장점은 수작업으로 빠뜨리고, 개인적으로 편협된 시각으로 제외시켰던 자료들이 컴퓨터에 의해서 복원될 수 있다는 점과 자료와 자료의 장르를 넘어서서 통합적인 검색이 가능해 진다는 데 있다.

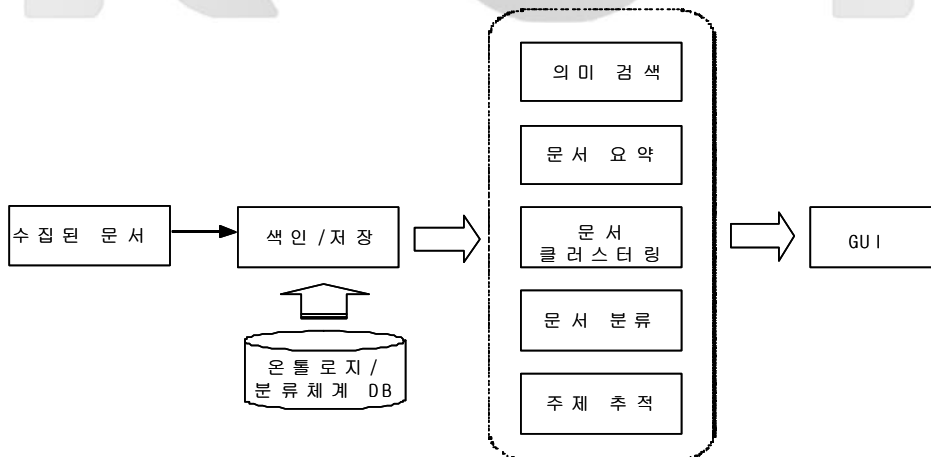
그림 2는 위에서 제시한 데이터마이닝시스템의 구성도이다. 각 데이터마이닝의 기능을 살펴보면 다음과 같다.

① 의미검색(semantic search): 특정 단어가

있는 문서를 찾을 때, 그 단어가 의미하는 내용이 있는 문서를 검색한다.

② 문서 요약(summarization): 문서요약은 사용자가 원하는 자료의 내용을 요약하여 제공한다. 통계적 요약방법에 MMR(Maximal Marginal Relevance) 기법을 적용하여 요약내용이 훨씬 더 함축적이고 분명하도록 만들 수 있다.<sup>16)</sup> 여기에서 문서요약은 컴퓨터가 자동으로 하는 것을 말한다(박순철 2004; Carbonell and Goldstein 1998).

③ 문서 클러스터링(clustering): 비슷한 문서들을 클러스터링하여 가시적으로 보여준다. 이 기능을 이용함으로써 단순한 검색보다 훨씬 더 빠르고 정확하게 원하는 정보를 찾을 수 있을 뿐더러 내용이 포함하고 있는 숨어있는 패턴도 유추해낼 수 있는 장점이 있다.



(그림 2) 데이터마이닝시스템 구성도

15) 여기에서 문서(document)라고 함은 컴퓨터가 인식하는 하나의 개체를 의미한다. 따라서 사진, 그림들도 하나의 문서라고 말한다.

16) MMR 알고리즘은 문서를 요약할 때 중복을 피하기 위해서 고안된 방법이다.

- ④ 문서 분류(classification): 전문가에 의해 작성된 분류체계에 따라 문서를 분류한다. 문서분류의 기능을 향상시키기 위해서 온톨로지 구축이 필수적이다. 문화원형콘텐츠에서는 전문가에 의한 심도있는 분류체계가 연구되면 분류의 질이 저절로 높아지게 될 것이다.
- ⑤ 주제추적(topic tracking): 특정 주제에 따라 문서의 내용을 순차적으로 찾는다.

수집된 자료와 테이터마이닝시스템을 이용함으로써 문화원형의 다양한 자료는 새로운 체계를 갖추고 이용자들이 원하는 정보를 가장 효율적으로 신속하게 만들어주게 된다. 예를 들면, 도깨비와 관련된 모든 문서, 이미지, 소리들을 통합적으로 클러스터링하고, 요약해서 검색결과를 제공할 수 있다. 이 방법에 따르면 인귀세상(표 1의 9-5), 한국의 굿(3-1), 죽음의 전통의례와 상징세계(3-7), 부적(6-15), 한국의 탈(6-28), 디지털민화(7-33) 등등 모든 문화원형콘텐츠 속에 들어있는 도깨비관련 정보를 다양한 방법으로 검색 및 분석할 수 있다.

## 6.2 온톨로지 기능 추가

최근 컴퓨터분야에서 온톨로지(ontology)에 대한 관심이 높아지고 있다. 원래 온톨로지란 철학의 한 분야로서 존재론이라고 하며, 실제(reality)에 대한 정확한 이해를 추구하는 학문이다. 실재하는 것을 규명하기 위해서는 이 세상에 존재하는 실체에 대한 명확한 이해와 정의가 필요하다. 컴퓨터분야에서 사용하는 온톨로지도 이러한 개념규정에서 크게 벗어나지 않는

다. 정보시스템의 대상이 되는 분야에서 다루는 내용들의 구성요소 또는 개념을 온톨로지라는 철학적 용어를 차용해서 사용하고 있다. 컴퓨터분야에서 구축한 온톨로지는 외국의 경우 주로 의료, 기계제조, 부동산, 금융 등의 분야가 활발하게 개발되었다(Aronson 2001, Srinivasan 2002). 대표적인 예로는 SENSUS, KACTUS, OTK, CommonKADS, Tove, Methoontology, Kikrokosmos, ONIONS, PHYSSYS 등이 있다.

한국의 전통문화원형과 관련된 용어와 개념에 대한 온톨로지가 구축되면 위에서 언급한 테이터마이닝 시스템의 기능을 강화시킬 수 있다. 이처럼 특정한 분야 또는 시스템 내에 구축된 온톨로지는 풍부하고도 신뢰도가 높은 검색결과를 제공해 줄 수 있기 때문이다. 궁극적으로 온톨로지는 컴퓨터로 하여금 인간과 유사한 개념구조를 갖도록 만드는 것이다. 한 예를 들어보면, 한국의 구비문학 속에 들어있는 귀물(鬼物)에 대한 온톨로지를 구축하고 나면 도깨비나 귀신에 대한 관련 자료들을 통합검색해서 재미있는 결과를 보여준다. 그림 3은 본 연구자들이 개발하고 있는 온톨로지를 응용해서 간단한 귀물 온톨로지를 구축해 보았다. 귀물이라고 하는 범주 안에 도깨비와 인귀가 서로 밀접한 관계가 있다는 것을 컴퓨터가 인식하게 되면서 다양한 자료들을 검색해 주고 흥미로운 결과를 보여준다.

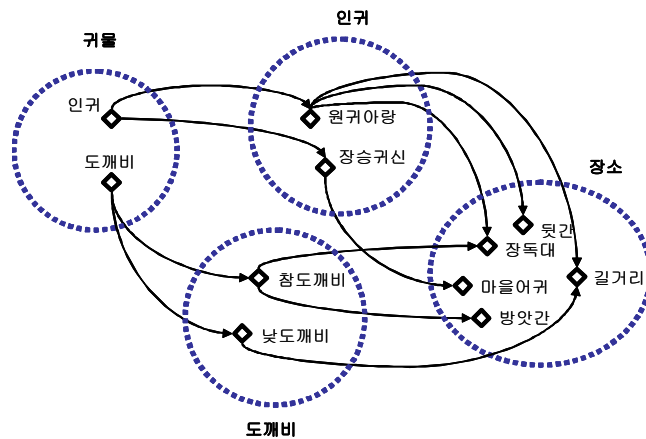
그림 3이 보여주는 온톨로지는 한국의 전통 문화와 문학 속에 나타나는 인귀와 도깨비의 개념과 관계지도의 극히 일부분이다. 이 둘의 정체는 귀물이라고 하는 보다 높은 항목으로 묶여지며, 이들이 존재하는 공간 또한 겹쳐진다. 또한 이들의 속성도 비슷한 경우도 많다. 인

귀와 도깨비가 각각 원한, 보은, 사랑, 복수, 수호 등 독특한 특징을 가지는 다양한 귀신과 도깨비로 나타나서 때로는 서로의 경계를 넘나들고 있다. 인귀에도 수많은 종류가 있다. 원한을 가진 원귀아랑으로부터 마을을 지키는 수호신의 역할을 하는 장승귀신에 이르기까지 다양하다. 도깨비 또한 마찬가지로 참도깨비, 낮도깨비, 총각도깨비 등등 그 종류와 속성이 가지가지다. 이렇게 세부적으로는 다른 것을 서로 견주어서 통합적으로 살펴다보면 보다 재미있고 창의적인 시나리오가 나올 수 있다. 기존의 틀에 박힌 원귀와 도깨비로 나뉘는 이야기가 아니라 이 둘의 범주를 뛰어넘어서 새로운 귀물 세상을 창조해 낼 수 있을 지도 모른다. 예를 들어서, 인귀나 도깨비가 출몰하는 장소를 중심으로 보면 이들은 서로 다른 공간에 존재하는 것들이 아니다. 길거리, 서낭당, 마을 어귀, 나무 숲속에 함께 등장하며 사람들에게는 그것이 인귀인지 도깨비인지 조차도 분명치 않게

나타나기 때문이다. 이처럼 온톨로지 기능을 충분히 이용하면 문화원형콘텐츠에서 개발한 결과물 안에 들어있는 귀물에 관한 모든 것들을 동시에 검색해 볼 수 있다. 한국의 굿, 죽음과 상례 그리고 부적 등도 귀물과 관련된 내용이 들어있다. 그러나 현재 시스템에서는 굿, 상례, 부적 속의 귀물들을 도깨비나 인귀와 동시에 찾아낼 수 없다. 이러한 한계를 극복하기 위해서는 우선적으로 통합적인 검색시스템의 구축이 시급하다.

## 7. 맺는 말

디지털화된 문화원형 자료의 가치가 제대로 평가받기 위해서는 앞으로 더 많은 연구가 필요하다. 특히 인문학자와 정보학 및 컴퓨터공학자들과의 통학적(uni-disciplinary)인 연구가 시급하다.<sup>17)</sup> 학제간의 벽을 허물면서 서로의 전문



(그림 3) 간략한 귀물 온톨로지의 예

17) 여기에서 통학적(uni-disciplinary)이란 학제간(interdisciplinary)의 연구를 뛰어 넘고자 하는 개념이다. 디지털 시대에 출현 가능한 새로운 학문분야를 염두에 둔 것이다.

성을 받아들이고, 그러한 자세에서 새로운 문화원형사업이 발전해 나갈 수 있을 것으로 판단된다. 특히 우리나라의 역사와 전통·풍물·생활·전승·예술·지리지 등 다양한 분야의 문화원형을 디지털 콘텐츠화하여 문화산업에 필요한 창작소재로 제공하려면 새로운 틀로 짠 지식관리시스템이 필수적이다. 이러한 목적을 가지고 본고에서는 현재의 문화콘텐츠닷컴의 디지털아카이브 시스템에 주목해서 문제점을 분석하고 대안을 제시해 보았다. 현재의 시스템으로는 의미검색이 이루어지지 않기 때문에 문화원형콘텐츠의 이용도가 떨어지고 있다는 점을 지적하였다. 유사한 자료들이 서로 다른 주제로 분류되면서 다른 항목에 속해서 이용자들은 통합적으로 정보를 검색할 수 없다. 즉 서로 다른 항목 간에 교차검색이 이루어지지 않고 있다는 점이다. 현재의 제 문제를 해결할 수 있는 방법으로 본고에서는 온톨로지를 이용한 의미검색 시스템을 개발하는 방법과 데이터마이닝을 이용해서 풍부한 지식정보표현과 활용이 가능한

디지털아카이브시스템을 제안하고자 한다.

문화원형자료가 디지털화되면서 지식이용의 차원이 혁명적으로 변화하고 있음을 언급하였다. 지식과 정보의 이용자층이 두터워졌고, 이들이 단순히 수동적이고 소극적인 이용자가 아니라는 점도 강조하였다. 현재의 생산자 중심의 정보유통체계를 이용자중심으로 전환시키는 정책방향의 설정도 시급함을 지적하였다. 문화원형콘텐츠사업이 관련 분야에 창작소재제공을 궁극적인 목적으로 삼고 있기 때문에 더욱 이용자위주의 시스템 창안으로 나아가야 한다.

문화원형콘텐츠가 통합적이면서도 획기적인 디지털아카이브 시스템을 갖추게 되면 우리 문화원형사업의 미래도 밝다. 특히 공공문화콘텐츠로서 거듭나고 있는 이 자료들이 창의성을 갖춘 이용자들에 의해서 효율적으로 활용될 수 있을 것이고, 그 결과 문화원형콘텐츠는 날개를 달고 세계의 문화유산으로 다시 태어날 수 있을 것이다.

## 참 고 문 헌

20세기민중생활사연구단. 『민중생활사연구 학술연구보고서』. 2003-2005.  
 김기덕. 2005. 문화원형 디지털콘텐츠사업의 사회적 효용. 『인문콘텐츠』, 5: 7-28.  
 김기덕. 2005. 문화원형의 층위와 새로운 원형 개념. 『인문콘텐츠』, 6: 55-74.  
 김선미. 1997. 『Dublin Core 메타데이터 형식 확장 설계에 관한 연구』. 이화여대 석사

논문.  
 김주관. 2005. 생활사아카이브구축의 의미와 방법. 『지방사와 지방문화』, 8:1.  
 박순철. 2004. 단일 문서의 인위적 요약과 MMR 통계요약의 비교 및 분석. 『대한전자공학회』, 41(2): 43-50.  
 박순철, 변동률, 최경웅. 2005. Text Mining for Ubiquitous Computing on PDA. U.S.-

- Korea Conference on Science, Technology, and Entrepreneurship, UKC 2005*. University of California, Irvine, USA.
- 박은정. 2004. 『북미지역 디지털아카이브즈 구축의 현황과 전망』. 한국기록관리협회 워크샵 제7회 기록관리 워크샵. 50-65.
- 박은정. 2005. 전자문서관리와 디지털아카이브즈: 성장과 전망. 『한국기록관리학회 학술논문집』. 23-32.
- 서은경. 2004. 디지털아카이브의 연구적 보존을 위한 개념적 모형 설계에 관한 연구. 『한국문헌정보학회지』, 38(1).
- 송현동. 2006. 문화콘텐츠와 한국학 문화원형 사업과 한브랜드를 중심으로. 『종교연구』.
- 윤택림 · 함한희. 2006. 『새로운 역사쓰기를 위한 구술사연구방법론』. 서울: 아르케.
- 이경순. 2001. 『정보검색에서 벡터공간모델과 클러스터 분석을 통한 문서 순위 결정 모델』. 한국과학기술원 박사학위논문.
- 이계학. 2002. 『향토문화전자대전 분류체계의 기본 아이디어』. 향토대전분류체계 세미나 발표원고.
- 이상훈. 2001. 디지털아카이브에 대한 이해의 지평: 정보화시대의 영상, 영상자료 그리고 새로운 테크놀로지. 『한국방송학보』, 15(2): 269-302.
- 이재윤, 최보영, 정영미. 2000. 문헌 자동분류에서 용어 가중치 기법에 대한 연구. 『정보관리학회지』, 41-44.
- 정영미, 이재윤. 2001. 지식분류의 자동화를 위한 클러스터링 모형 연구. 『정보관리학회지』, 18(2): 203-230.
- 한국문화콘텐츠진흥원. 2003. 『우리문화원형의 디지털콘텐츠화 사업 종합계획 2003-2007』.
- 한국문화콘텐츠진흥원. 2005. 『한국문화콘텐츠진흥원 지원사업 평가지표 개발연구』.
- 함한희. 2006. 전북지역 민중생활사 자료와 아카이브구축의 의미. 『자료조사연구로 본 전라문화』. 전주: 도서출판 선명.
- 피에르 레비. 2000. 『사이버 문화 뉴테크놀로지와 문화협력 그리고 커뮤니케이션 -』. 서울: 문예출판사.
- Aronson AR, 2001. *Effective mapping of biomedical text to the UMLS meta-thesaurus: the MetaMap program*. Proc AMIA Symp. Washington DC, USA
- Baeza-Yates, Ribeiro-Neto. 『최신 정보 검색론』, 홍릉과학출판사, 2001.
- Berry, M. W., and Murray Browne. *Understanding Search Engines*. University of Tennessee.
- Carbonell, James and Jade Goldstein. 1998. The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *Proceedings of the 21st ACM-SIGIR International Conference on Research and Development in Information Retrieval*, Melbourne, Australia.
- Croft, W. B. 2000. *Advances In Information Retrieval*. Kluwer Academic Publishers.
- Everett, Stephen E. 1992. *Oral History Techniques and Procedures*. Washington:

- U.S. Army Center of Military History.
- Goldstein, Jade et al. 1999. *Summarizing Text Documents: Sentence Selection and Evaluation Metrics*. In Proceedings of ACM-SIGIR'99. Berkeley, CA.
- Grele, Ronald J. 1987. On Using Oral History Collections: An Introduction. *The Journal of American History*, 74(September).
- Ketelaar, Eric. 2003. Being Digital in People's Archives. *Archives and Manuscripts*, 31(2): 8-22.
- Korfhage, Robert, R. 1997. *Information Storage And Retrieval*. Hohn Wiley Sons, Inc.
- Mani, Inderjeet, and Mark Maybury. 1999. *Advances in Automatic Text Summarization*. MA: MIT Press.
- Moore, Frank W. ed. 1961. *Readings in Cross-cultural Methodology*. New Haven: HRAF.
- HRAF et al. 1987. *Outline of Cultural Materials*. New Haven: HRAF. 1983 data from <http://www.yale.edu/hraf>
- G. Salton and M. J. McGill. 1983. Introduction to Modern Information Retrieval. McGraw-Hil.
- Srinivasan, S. et al. 2002. Finding UMLS metathesaurus concepts in MEDLINE. *Proc. AMIA Symp.* San Antonia: TX, USA.
- Stielow, Frederick. 1986. *The Management of Oral History Sound Archives*. Westport, CN: Greenwood Press.
- 한국문화콘텐츠진흥원.  
<[www.culturecontent.com](http://www.culturecontent.com)>
- 한국역사정보통합시스템.  
<[www.koreanhistory.or.kr](http://www.koreanhistory.or.kr)>
- Dublin Core Metadata Initiative.  
<[www.dublincore.org](http://www.dublincore.org)>
- Electronic Cultural Atlas Initiative.  
<[www.ecai.org](http://www.ecai.org)>
- HRAF.  
<[www.yale.edu/hraf](http://www.yale.edu/hraf)>
- Memory of the World.  
<[www.unesco.org/webworld](http://www.unesco.org/webworld)>
- search cast.  
<[www.searchcast.net](http://www.searchcast.net)>
- search mate.  
<[www.searchmate.co.kr](http://www.searchmate.co.kr)>
- vivisimo.  
<[www.vivisimo.co.kr](http://www.vivisimo.co.kr)>