

---

# 디지털콘텐츠의 특성을 고려한 백업 시스템의 설계 및 구현

## Design and Implementation of a Backup System for Digital Contents

---

이석재, 윤종현, 황석철, 유재수  
충북대학교 정보통신공학과

Seok-Jae Lee(sjlee@netdb.chungbuk.ac.kr), Jong-Hyun Yun(blurette@netdb.chungbuk.ac.kr),  
Sok-Cheol Hwang(bbacdo@netdb.chungbuk.ac.kr), Jae-Soo Yoo(yjs@chungbuk.ac.kr)

---

### 요약

첨단 IT 기술의 발전으로 유무선 통신망을 통하여 다양한 환경에서 사용되는 대용량 디지털콘텐츠들이 급증하고 있다. 이와 함께 불시의 사고에 의한 디지털콘텐츠의 손실을 방지하기위해서 지속적인 데이터 백업의 필요성이 증가하고 있다. 본 논문에서는 디지털콘텐츠의 특성을 고려하여 동일한 데이터가 중복되어 저장되는 것을 방지하기 위해 콘텐츠를 입출력 I/O 크기 단위로 객체화하고, 유일한 ID를 부여하여 백업장치에 저장하는 백업 시스템을 설계하고 구현한다. 제안하는 백업 시스템은 중복되어 발생하는 데이터에 대해 하나의 사본만 백업함으로써 백업되는 데이터의 양을 효과적으로 감소시켜 디지털 콘텐츠를 한정된 저장 공간에 효율적인 백업이 가능하도록 하였다.

■ 중심어 : | 디지털콘텐츠 | 백업 시스템 | CAS |

### Abstract

With the development of IT technology, the amount of digital contents used in various environments of wired/wireless networks have been increased hugely and rapidly. To protect the loss of the digital contents from the sudden accident, continuous data backup is required. In this paper, we design and implement the backup system that stores digital contents in backup storage by objectifying the contents with a unit of I/O size and giving them the unique ID using the properties of digital contents to avoid duplicated store of the same data. The backup system reduces the amount of backup data efficiently by backing up the only one copy of the duplicated data. as a result, the backup system can back up the digital contents more efficiently in a constrained storage space.

■ keyword : | Digital Contents | Backup System | CAS |

---

## 1. 서론

네트워크 기술과 인터넷의 발전에 따라 수많은 사용자

들은 유무선 전기 통신망을 통하여 텍스트, 이미지, 동영상, 애니메이션, 음악 등의 다양한 형태로 이루어진 방대한 양의 데이터를 사용하고 있다. 이렇게 유무선 전기 통

---

\* 본 연구는 한국과학재단 목적기초연구(특정기초연구 과제번호 F01-2003-000-10627-0) 지원 및 산업자원부의 지역혁신 인력양성사업의 연구결과로 수행되었습니다.

접수번호 : #050926-001

접수일자 : 2005년 09월 26일

심사완료일 : 2005년 12월 15일

교신저자 : 유재수, e-mail : yjs@chungbuk.ac.kr

신망에서 사용하기 위해 첨단 IT 기술을 사용하여 부호, 문자, 음성, 음향, 영상 등을 디지털 포맷으로 가공 처리한 정보를 디지털콘텐츠라고 정의한다[1-3].

최근의 네트워크 전송과 그래픽, 무선인터넷 등 나날이 발전하는 IT 기술은 다양한 대용량 디지털콘텐츠의 개발과 전송을 가능하게 한다. 또한 다양한 종류의 디지털콘텐츠들이 기하급수적으로 증가되는 것을 가속화시키고 있으며, 많은 사용자들로 하여금 다양한 형태의 디지털콘텐츠를 접하고 사용할 수 있도록 한다. 인터넷 게임의 예와 같이 최근의 디지털콘텐츠들은 대용량으로 작성되어 사용자들에게 현실감 높은 화면과, 웅장한 입체사운드, 영화와 같은 애니메이션의 전개, 그리고 여러 가지의 다양한 서비스 등 기존의 콘텐츠들보다 발전된 형태로 제공된다. 이렇게 발전된 디지털콘텐츠들은 더욱 많은 사용자들을 불러 모으고 있다[2][3].

그러나 이러한 디지털콘텐츠들은 불시의 기계 고장, 네트워크를 통한 침입자, 각종 바이러스 등에 매우 취약하며, 최악의 경우 디지털 콘텐츠가 훼손될 수 있다. 이러한 예기치 못한 콘텐츠의 훼손은 업무에 커다란 영향을 주어 심각한 문제를 발생시키고, 서비스 제공자와 사용자에게 엄청난 손실을 가져온다. 따라서 디지털콘텐츠가 손실 또는 손상되기 전에 미리 안전한 곳에 백업하고, 사고 이후 신속한 복구를 통해 업무 흐름을 원활하게 만드는 백업과 복구가 필요하다[4-6]. 이와 더불어 백업되는 디지털콘텐츠의 급증에 따른 백업 데이터의 대용량의 저장 공간에 대한 요구를 항상 충족시킬 수는 없기 때문에, 한정된 저장 공간을 효율적으로 이용하여 백업하는 효과적인 저장 기술이 필요하다[7][8].

이에 본 논문에서는 기하급수적으로 늘고 있는 디지털콘텐츠의 특성을 고려한 효율적인 백업을 위해 디지털콘텐츠를 객체 단위로 저장하고 관리하는 백업 시스템을 제안한다. 디지털콘텐츠의 경우 그 특성상 데이터의 중복 비율이 매우 높다. 무수히 많은 디지털콘텐츠들 중 비슷한 종류의 디지털콘텐츠를 일정한 크기로 나누어 분석해 보면, 전체의 데이터들 중에서 일부분의 데이터들이 서로 중복되어 나타나는 것을 발견할 수 있다. 이렇게 부분적으로 중복되어 나타나는 디지털콘텐츠 데이터를 하나의 객체 단위로 저장 장치에 저장하고, 저장된 데이터

와 같은 데이터를 부분적으로 갖는 디지털콘텐츠들 간에 저장된 데이터를 공유시킨다. 따라서 중복된 디지털콘텐츠 데이터를 모두 저장하는 기존의 백업 시스템에 비해, 중복되어 저장되는 디지털콘텐츠를 효과적으로 제거할 수 있어, 적은 저장 공간으로 방대한 양의 디지털콘텐츠 백업 데이터를 관리할 수 있다.

본 논문의 구성은 다음과 같다. 먼저 2장에서는 디지털콘텐츠 저장기술과 백업 기술에 대한 관련연구를 알아보고, 3장에서는 제안하는 디지털콘텐츠 백업 시스템에 대해 시스템 구조와 각 모듈별 특징, 그리고 백업 및 복구 처리 방법으로 나누어 설명한다. 4장에서는 제안하는 디지털콘텐츠 백업 시스템의 인터페이스를 설명하고, 5장에서는 성능 평가와 분석을 통해 논문에서 제안하는 백업 시스템의 우수성을 증명한다. 마지막으로 6장에서는 결론을 맺고 향후 연구방향을 제시한다.

표 1. 디지털콘텐츠 저장기술 비교

	SAN	NAS	CAS	OSD
메타데이터 생성	응용 프로그램	응용 프로그램	메타데이터 서버	메타데이터 서버
저장 데이터	블록	파일	객체	객체
데이터 유형	가변 콘텐츠	가변 콘텐츠	고정 콘텐츠	고정 콘텐츠
데이터 주소	Location	Location	콘텐츠	콘텐츠
데이터 중복	발생	발생	발생하지 않음	발생

## II. 관련 연구

### 1. 디지털콘텐츠 저장기술

최근에 연구되고 있는 디지털콘텐츠 저장 기술로 객체 기반 스토리지 장치(Object-based Storage Device : OSD)와 콘텐츠 기반 스토리지 장치(Content Addressed Storage : CAS)가 있다[7][8].

[표 1]은 기존의 디지털콘텐츠 저장기술과 CAS와 OSD를 비교한 것이다[7-9]. 기존의 디지털콘텐츠 저장 기술인 NAS(Network Attached Storage)와 SAN(Storage Area Network)을 보완하기 위하여 새롭게 등장한 CAS와 OSD는 데이터를 파일 단위로 저장하는 기

존의 파일 기반 저장장치와 달리 데이터를 일정한 입출력 크기로 나눈 후, 나눈 데이터의 개개별로 고유 ID를 부여하여 객체 단위로 저장 장치에 저장하고, 접근한다[7]. 데이터를 객체 단위로 저장하면 스토리지로부터 객체에 직접 접근함으로써 시스템 성능을 향상시킬 수 있으며, 시스템 성능의 저하 없는 확장을 제공할 수 있다. 또한 객체에 대한 메타데이터에 독립적인 객체 접근 인터페이스를 사용함으로써, 이기종 플랫폼에서 객체를 안전하게 공유할 수 있는 특성을 제공할 수 있다[7][8].

2. 일반적인 백업 기술

백업은 천재지변이나 사용자의 실수, 전산 장비의 고장 등에 따른 불시의 사고에 대비하여 파일 또는 데이터 베이스를 복사해 두는 행위를 말한다. 데이터의 백업은 대형 컴퓨터를 운영하는 대규모 사업체에서는 물론 개인 컴퓨터에서도 필수적이고 일상적인 업무이다. 백업은 백업의 형태, 대상, 기능, 백업 시스템의 구조 등에 따라서 백업 방식을 [표 2]와 같이 분류할 수 있다[6][10-12].

표 2. 다양한 백업 방식 분류

구분	백업 방식	
형태	전체백업	백업을 수행할 시스템에 존재하는 모든 파일 백업
	부분백업	이전의 백업 데이터와 비교하여 변경된 파일만 백업
대상	파일 기반 백업	파일 구조를 파악하여 파일과 디렉토리 구조를 백업
	장치 기반 백업	파일 구조를 무시하고 디스크 블록을 백업
기능	온라인 백업	작업 수행 중에도 백업 지원
	오프라인 백업	백업을 위해 일반적인 작업을 중지
시스템 구조	S/W 구조	중앙집중식백업 지역관리백업
	H/W 구조	LAN 백업 LAN-Free 백업 Serverless 백업

적으로 백업된 디지털콘텐츠 데이터가 저장될 저장 공간의 크기를 고려해야 한다. 디지털콘텐츠는 사용되어지는 분야에 따라 매우 높은 데이터의 중복도를 보인다. 즉 디지털콘텐츠를 일정한 입출력 크기로 분할하여 분석하면, 다른 유사한 디지털콘텐츠와 부분적으로 유사한 데이터를 가지고 있거나, 혹은 두 디지털콘텐츠가 일치할 가능성이 매우 높다. 그러나 기존의 백업 시스템은 이러한 디지털콘텐츠의 특성을 고려하지 않고, 중복된 디지털콘텐츠 데이터 전부에 대해 백업을 수행하므로, 백업되어지는 디지털콘텐츠의 양이 늘어남에 따라 사용되어지는 저장 공간 또한 기하급수적으로 커져야 하는 문제점이 있다[8].

이러한 문제점을 해결하기 위해 논문에서는 CAS 기술을 응용하여 디지털콘텐츠의 특성을 파악하고, 중복된 디지털콘텐츠 데이터에 대해 하나의 사본만 백업한다. 백업된 사본은 백업된 사본과 동일한 데이터를 갖는 디지털콘텐츠가 백업될 경우, 제안하는 백업 시스템에 의해 공유된다. 따라서 중복된 디지털콘텐츠에 대해 중복되어 백업되는 데이터의 양을 효과적으로 감소시켜 한정된 저장 공간에서도 효과적인 백업이 가능하다.

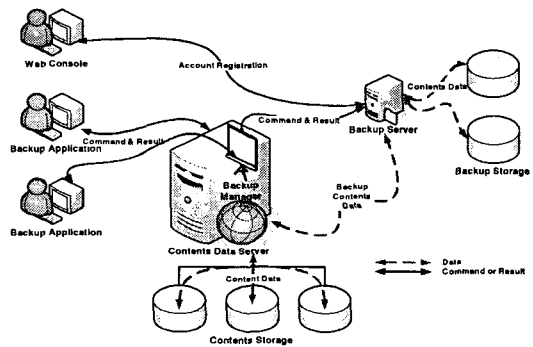


그림 1. 제안하는 백업 시스템의 전체 구조

1. 시스템 구조

[그림 1]은 제안하는 백업 시스템의 전체적인 구조를 나타낸다. 제안하는 디지털콘텐츠 백업 시스템의 구조는 크게 백업 웹 단말기(Backup Web Console), 백업 관리자(Backup Manager) 및 백업 어플리케이션(Backup Application), 백업 서버(Backup Server)로 구

III. 디지털콘텐츠 백업 시스템 구조 설계

최근의 디지털콘텐츠들은 다양한 종류의 대용량 데이터로 구성되며 그 수 역시 기하급수적으로 늘고 있다. 따라서 디지털콘텐츠 백업 시스템의 설계를 위해서는 우선

분할 수 있다.

백업 웹 단말기는 백업 시스템을 사용하는 사용자들의 계정을 관리한다. 관리자는 백업 웹 단말기를 통해 현재 등록된 사용자의 계정정보를 확인 할 수 있다. 또한 새로운 계정을 추가하거나 기존에 존재하는 계정을 관리하는 것이 가능하다. 백업 웹 단말기를 통해 관리되는 데이터는 백업 서버의 데이터베이스에서 유지한다. 이렇게 사용자 계정 정보를 백업 서버의 데이터베이스에서 관리하는 이유는 백업 서버에서 계정 정보를 이용해서 백업 관리자의 접근을 관리하기 때문이다. 또한 백업 웹 단말기는 각 사용자에게 할당된 백업 서버의 백업 장치 할당량을 관리할 수 있다.

백업 관리자는 디지털콘텐츠를 관리하는 디지털콘텐츠 관리 서버에 위치한다. 백업 관리자는 백업 어플리케이션을 사용하여 사용자로부터 백업 및 복구 작업 요청을 입력받고, 이를 백업 서버에 전달하여 작업을 처리한다. 스케줄러에 의해 수동 및 자동 백업을 수행할 수 있다. 또한 실시간 모니터링 기능을 사용하여 변경된 디지털콘텐츠를 실시간 백업할 수 있다. 백업된 디지털콘텐츠를 확인하고 필요에 따라 선택한 백업 데이터를 복구한다.

백업 어플리케이션은 백업 관리자에게 사용자의 명령을 전달하기 위한 단말기로 동작한다. 사용자는 백업 어플리케이션을 통해 다양한 명령을 백업 관리자에게 전달하고, 전달한 명령에 따라 추가적인 데이터를 요청받거나 수신할 수 있다.

백업 서버는 별도의 백업 장치와 데이터베이스를 갖는 서버로 구성된다. 백업 서버는 백업 관리자를 통한 사용자의 작업 요청을 처리한다. 작업을 요청한 사용자의 계정 정보에 대한 인증 과정을 거친 후, 사용자가 요청한 백업 및 복구 작업을 처리한다. 백업된 디지털콘텐츠를 지정된 저장소에서 관리한다.

## 2. 구성모듈

제안하는 디지털콘텐츠 백업 시스템의 백업 관리자와 백업 서버는 [그림 2]와 같은 모듈로 구성되어 있다. 제안하는 디지털콘텐츠 백업 시스템에서 수행되는 작업은 크게 백업 관리자에서 수행되는 작업과 백업 서버에서

수행되는 작업으로 분류된다. 백업 관리자에서는 백업 서버로 보낼 백업 및 복구 작업에 대한 요청 작성 및 디지털콘텐츠 데이터의 송수신을 처리한다. 백업 서버는 백업 관리자가 보낸 작업 요청을 분석하고, 백업 관리자로 부터 백업할 디지털콘텐츠를 수신한다. 또한 백업된 디지털콘텐츠를 백업 관리자에게 송신한다.

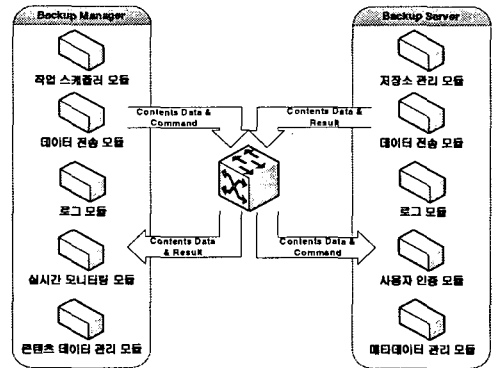


그림 2. 백업 관리자와 백업 서버의 주요 구성 모듈

### 2.1 백업 관리자의 구성 모듈

백업 관리자는 크게 작업 스케줄러 모듈, 데이터 전송 모듈, 로그 모듈, 실시간 모니터링 모듈, 콘텐츠데이터 관리 모듈로 구성된다.

작업 스케줄러는 사용자의 작업 요청을 관리한다. 사용자가 작성한 작업 스케줄에 의해 백업 및 복구 작업을 수행하도록 명령한다. 작업 스케줄러 모듈은 수동 및 자동 백업과 복구 작업을 처리한다. 자동 백업을 처리하기 위해 작업 스케줄러는 시스템의 시간을 주기적으로 검사하는 타이머(Timer)를 갖는다. [그림 3]은 작업 스케줄러의 동작 과정을 나타낸다. 사용자는 먼저 작업 스케줄러를 사용하여 스케줄을 작성한다. 사용자는 스케줄에 의해 수행할 작업과 스케줄이 동작할 시간을 지정한다. 백업의 경우 스케줄에 의해 백업된 백업 정보를 저장할 백업 서버를 선택한다. 스케줄이 작성되면 스케줄에 의해 백업되거나 복구될 디지털콘텐츠 정보를 선택한다. 작업 스케줄러 모듈은 모듈 내의 타이머를 사용하여 시스템의 시간을 주기적으로 검사하고, 등록된 스케줄의 수행 여부를 결정한다.

데이터 전송 모듈은 백업 및 복구 시 서버로 전송하거나, 전송받을 콘텐츠 데이터 및 다수의 명령어를 처리한다. 백업의 경우 데이터 전송 모듈은 사용자가 지정한 백업 작업 목록에 따라, 백업 서버에게 백업 준비 명령과 백업할 디지털콘텐츠 데이터를 전송한다. 복구의 경우 백업 서버로부터 백업된 디지털콘텐츠 목록과 디지털콘텐츠 데이터를 수신한다. 또한 백업이나 복구 시 사용자 인증 및 사용자 정보 획득과 관련된 다수의 명령어를 백업 서버에게 전송하고, 그 결과를 수신한다.

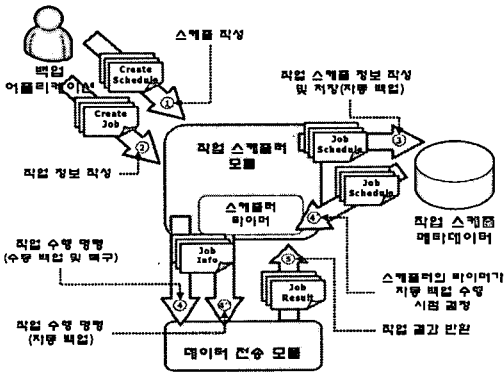


그림 3. 작업 스케줄러의 동작 과정

표 3. 제안하는 백업 시스템의 로그(Log) 종류

로그 종류	정의 및 기능
디지털콘텐츠 히스토리 로그	백업된 디지털콘텐츠의 정보를 기록하는 로그로 백업된 콘텐츠의 메타 정보 및 백업 정보와 백업 버전 정보를 기록
작업 로그	백업 및 복구 작업의 성공 실패 여부를 기록하는 로그로 작업의 성공 또는 실패 여부와 저장 장치 사용과 관련된 통계 정보 기록
시스템 에러 로그	백업 시스템에서 발생한 에러를 기록하는 로그로 백업 및 복구 작업 시 발생 가능한 다양한 에러 정보 (저장 장치 에러, 통신 에러, 인증 에러 등) 기록

로그 모듈은 백업 및 복구 작업을 처리하면서 발생한 다양한 로그 정보를 기록한다. 제안하는 디지털콘텐츠 백업 시스템의 로그는 [표 3]과 같이 크게 디지털콘텐츠 히스토리 로그, 백업 및 복구 작업 로그, 시스템 에러 로그로 구성된다. 제안하는 백업 시스템은 로그 정보를 분리하여 시스템 에러에 따른 백업 및 복구 작업의 중단점

을 정확하게 파악하고, 경우에 따라 발생된 에러로부터 시스템 복원이 가능하다.

실시간 모니터링 모듈은 사용자가 지정한 특정 위치의 디지털콘텐츠에 대해 콘텐츠의 변경을 수시로 검사하여, 변경된 콘텐츠만을 선택적으로 백업하는 모듈이다. 실시간 모니터링 모듈은 사용자가 지정한 디지털콘텐츠에 대해 이전 백업 정보가 있는지 검사하고, 디지털콘텐츠의 메타데이터를 분석하여 콘텐츠의 변경 여부 및 백업 버전 정보를 확인한다. 만약 이전 백업 정보가 없으면 실시간 모니터링을 설정한 파일은 자동으로 백업되며, 이후 디지털콘텐츠의 변경 여부에 따라 실시간 모니터링에 의해 자동으로 백업된다. 이전 백업 정보가 있는 디지털콘텐츠의 경우 디지털콘텐츠의 변경 날짜와 이전 백업 정보의 백업 날짜를 비교하여 백업 여부를 결정한다.

콘텐츠 데이터 관리 모듈은 백업될 디지털콘텐츠를 백업 서버에 기록하기 위해 메타 정보를 생성한다. 제안하는 디지털콘텐츠 백업 시스템은 CAS와 유사한 방법을 사용하여 디지털콘텐츠의 백업 데이터를 관리한다. 콘텐츠데이터 관리 모듈은 디지털콘텐츠의 데이터와 메타 정보를 조합하여 백업될 디지털콘텐츠에 대한 고유한 ID와 콘텐츠 메타데이터를 생성한다.

콘텐츠 데이터 관리 모듈은 MD5 알고리즘을 사용하여 디지털콘텐츠에 고유한 ID를 할당한다. MD5 알고리즘은 R. L. Rivest가 1990년에 개발한 MD4 알고리즘을 개선한 것으로, 입력 데이터(길이에 상관없는 하나의 메시지)로부터 128 비트 메시지 축약을 만듦으로써 데이터 무결성을 검증하는데 사용되는 해시 기반 알고리즘이다 [13]. 일반적인 해시 알고리즘에서 서로 다른 메시지에 대한 해시 알고리즘의 결과 값이 같을 확률은 매우 높으며, 근사적으로 볼 때 n개의 결과 값을 가질 수 있는 알고리즘에서 두 결과 값이 같을 확률이 0.5 이상이 되게 할 최소의 결과 값의 개수는 n의 제곱근이다. 여기서 만약 결과 값이 n bit라면 2^n개의 서로 다른 결과 값들이 나올 수 있는 경우의 수가 있으며, 이 결과 값들 중 2^(n/2)개의 서로 다른 메시지에 대하여 검사했을 때 그 값이 서로 같은 값이 나오게 될 확률이 0.5 이상이 된다고 할 수 있다. 그러므로 알고리즘의 결과 값에 대해서로 같은 값이 나오는 충돌(Collision) 현상이 발생하는 것

을 막기 위해서는 이  $n$ 이 상당히 큰 수이어야 하며, 이론상  $n$ 이 128 정도이면 적당하다고 알려져 있다. 따라서 MD5에 의해 산출된 결과 값은 128bit로 구성되며, 이에 의해 MD5 알고리즘으로 입력된 어떤 두 개의 메시지가 동일한 결과 값을 내는 것은 계산적으로 불가능하다[13].

콘텐츠 데이터 관리 모듈은 디지털콘텐츠를 백업 서버의 저장소 성능을 고려한 데이터 크기로 분할한다. 분할된 데이터의 순서에 따라 차례대로 MD5 알고리즘을 적용시켜, 분할된 데이터 각각에 ID를 할당한다. 이 분할된 콘텐츠 데이터에 대한 ID는 백업 서버의 백업 저장소에서 데이터가 저장될 위치를 결정하며, 복구 시에는 백업된 데이터로의 접근을 가능하게 하는 주소가 된다. 또한 할당된 각각의 ID는 각 디지털콘텐츠의 이름 또는 위치 경로와 같은 메타데이터와는 독립적으로 디지털콘텐츠의 데이터만을 사용하여 할당되므로, 중복된 데이터를 갖는 두 디지털콘텐츠 간에 파일 이름이나 경로에 상관없이, ID만으로 백업된 콘텐츠 데이터의 공유가 가능하다. 따라서 각 콘텐츠 데이터에 할당된 ID는 콘텐츠 데이터의 공유를 보다 편리하게 한다.

백업된 콘텐츠에 대한 사용자별 메타데이터는 콘텐츠에 할당된 콘텐츠 ID들과 콘텐츠에 대한 메타데이터 정보 및 사용자 정보로 구성된다. 콘텐츠에 대한 사용자별 메타데이터 정보는 복구 작업을 위해 콘텐츠데이터 관리 모듈이 관리하는 데이터베이스에 저장되며, 복구 작업 요청에 따라 필요한 메타데이터를 제공한다.

## 2.2 백업 서버의 구성 모듈

백업 서버는 크게 저장소 관리 모듈, 데이터 전송 모듈, 로그 모듈, 사용자 인증 모듈, 메타데이터 관리 모듈로 구성된다.

저장소 관리 모듈은 백업 서버의 데이터 전송 모듈을 통해 입력된 사용자의 요청에 따라 전달된 콘텐츠 데이터를 지정된 백업 저장소에 저장하거나, 또는 지정된 저장소로부터 읽는 기능을 수행한다. 또한 사용자별로 할당된 저장 장치 할당량을 관리하는 기능을 수행한다. 저장소 관리 모듈에 의해 저장되는 콘텐츠 데이터는 백업 관리자가 할당한 콘텐츠 ID를 파일 이름으로 갖는다. 백업 관리자가 할당한 콘텐츠 ID는 MD5 알고리즘에 의해

작성되어 동일한 데이터를 갖는 콘텐츠가 백업되는 경우 콘텐츠 데이터를 저장하지 않고, 사용자에게 할당된 백업 장치 사용량만 갱신한다. 이 정책에 의해 제안하는 백업 시스템은 저장 공간의 사용을 크게 줄일 수 있다.

데이터 전송 모듈은 백업 관리자가 전송한 디지털콘텐츠를 저장소 관리 모듈의 정책에 따라 백업 저장소에 저장한다. 백업 관리자의 복구 작업 요청 시 백업 저장소로부터 백업 관리자가 요청한 백업 콘텐츠 정보를 읽고 백업 관리자로 전송한다. 사용자가 요청한 백업 서버의 각종 통계 정보를 전송한다.

로그 모듈은 백업 및 복구 작업을 처리하면서 발생한 다양한 로그 정보를 기록한다. 백업 관리자의 로그 모듈과 동일하게 백업 서버의 로그 모듈은 디지털콘텐츠 히스토리 로그, 백업 및 복구 작업 로그, 시스템 에러 로그를 기록한다. 백업 서버의 로그 모듈은 백업 관리자가 요청한 작업 단위로 로그를 기록한다. 또한 기록되는 로그 정보를 효과적으로 관리하기 위해 로그 모듈의 로그 기록 주기를 설정할 수 있다.

사용자 인증 모듈은 백업 관리자를 통해 백업 서버에 접근하는 사용자들의 인증 정보를 검사하고, 사용자별로 할당된 백업 저장 공간 정보를 저장소 관리 모듈에게 통보한다. 사용자의 인증 정보는 관리자가 백업 웹 단말기를 사용하여 입력한다. 사용자의 인증 정보는 크게 사용자 계정 정보와 사용자 백업 장치 할당량으로 구성되며, 입력된 인증 정보는 백업 서버의 메타데이터 관리 모듈에 의해 데이터베이스로 저장, 관리된다.

메타데이터 관리 모듈은 백업 서버에서 백업 웹 단말기를 통해 입력된 사용자의 계정 정보와 백업 서버에서 발생한 로그 정보 등과 같은 각종 로그 정보를 백업 서버 내의 데이터베이스에 저장, 관리하는 기능을 수행한다. 메타데이터 관리 모듈은 백업 서버 내의 다른 모듈이 요청하는 메타데이터를 데이터베이스로부터 읽고, 쓰는 작업을 수행하며, 백업 웹 단말기의 요청에 따라 새로운 사용자의 인증 정보를 데이터베이스에 저장, 관리한다.

## 3. 백업 처리

[그림 4]는 제안하는 백업 시스템의 백업 처리 과정을 나타낸다. 제안하는 백업 시스템의 백업 처리 순서는 다

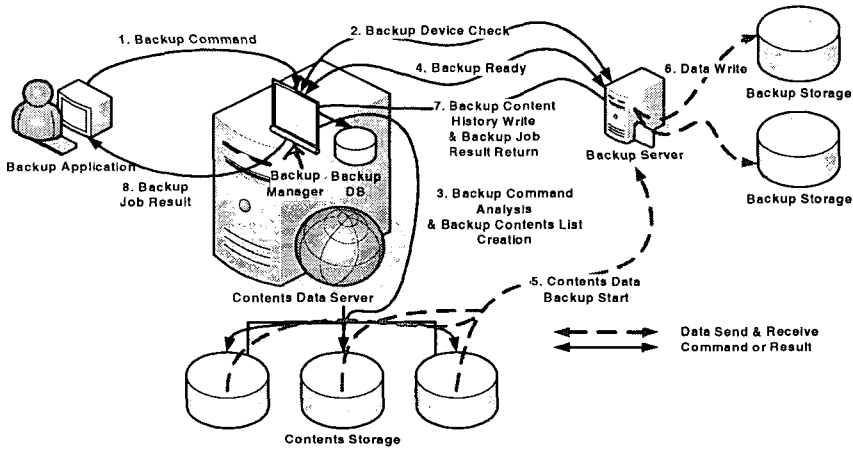


그림 4. 제안하는 백업 시스템의 백업 처리 과정

음과 같다.

- ① 사용자는 백업 어플리케이션을 사용하여 백업 명령을 백업 관리자에게 전달한다.
- ② 백업 관리자는 백업 서버의 점검과 사용자의 계정 정보를 점검한다.
- ③ 백업 관리자는 작업 스케줄 정보를 분석하고 백업할 콘텐츠들에 대한 백업 목록을 작성한다.
- ④ 백업 관리자는 백업 목록에 기록된 디지털콘텐츠데이터에 대해 콘텐츠 데이터 관리 모듈을 사용하여 콘텐츠 데이터를 분석한다.
- ⑤ 백업 관리자는 콘텐츠 데이터를 백업 서버로 전송한다.
- ⑥ 콘텐츠 데이터는 백업 서버의 저장 장치에 콘텐츠 ID를 이름으로 하는 파일로 저장된다.
- ⑦ 백업 서버는 작업 상태를 백업 관리자에게 반환한다. 백업 관리자는 백업된 콘텐츠 데이터의 메타 정보와 콘텐츠 ID를 백업 콘텐츠 히스토리 정보로 작성하여 기록한다.
- ⑧ 백업 작업 결과를 사용자에게 전달한다.

백업 처리 시, 백업 어플리케이션을 사용하여 사용자가 작성하는 백업 작업 정보는 크게 자동 백업 정보와 수

동 백업 정보로 구분된다. 자동 백업은 사용자가 선택한 디지털콘텐츠에 대해 주기적인 백업을 수행한다. 자동 백업의 경우 사용자는 디지털콘텐츠가 백업될 주기를 월, 일 단위로 선택할 수 있다. 또한 백업 작업이 일어나는 시간을 설정할 수 있다. 수동 백업의 경우 백업 주기를 선택하지 않으며, 작업 스케줄이 작성되는 즉시 백업 작업을 수행한다. 작업 스케줄에서는 백업된 콘텐츠 정보가 저장될 백업 서버를 선택할 수 있으며, 부가적으로 특정 확장자를 갖는 파일이나, 특정 크기를 갖는 파일에 대해 백업을 제한할 수 있다. 이는 시스템 관련 파일이 임의의 사용자에 의해 백업되는 것을 막기 위함이다. 스케줄이 작성되면 백업할 콘텐츠를 선택한다. 제안하는 백업 시스템은 백업할 콘텐츠를 개별적으로 선택할 수도 있고, 디렉터리 단위로 선택할 수도 있다.

[그림 5]는 백업 작업 시 백업 관리자와 백업 서버의 동작을 나타낸다. 콘텐츠 데이터 관리 모듈은 백업할 콘텐츠에 대한 메타 정보를 작성한다. 또한 콘텐츠를 일정한 크기의 데이터 단위로 분할하고, 분할된 콘텐츠 데이터들에게 MD5 알고리즘을 적용하여 각각의 콘텐츠 ID를 할당한다. 이때, 각기 다른 콘텐츠의 한 부분인 Data1과 Data1'이 동일한 내용을 갖는다면 콘텐츠 데이터 관리 모듈의 MD5 알고리즘에 의해 동일한 콘텐츠 ID를 할당받는다.

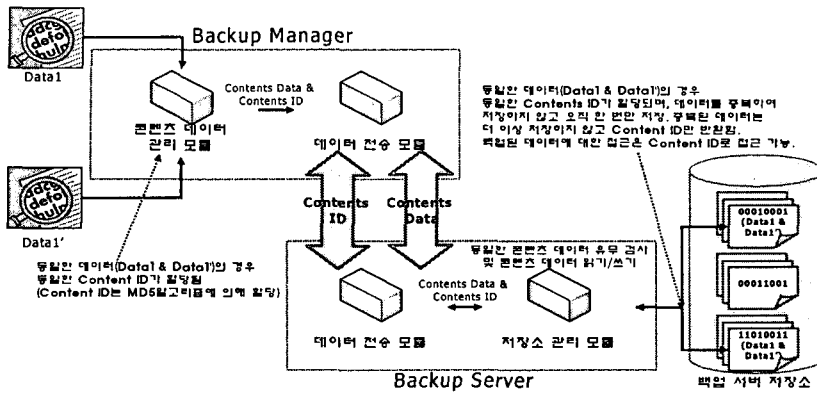


그림 5. 백업 작업 시 백업 관리자와 백업 서버가 데이터 처리 흐름

백업 관리자는 콘텐츠 데이터를 백업 서버로 전송할 때 분할된 콘텐츠 데이터들에게 할당된 콘텐츠 ID를 백업 서버로 먼저 전송한다. 백업 서버는 메타데이터 관리 모듈을 사용하여 백업 관리자가 보낸 콘텐츠 ID가 이미 백업 서버에 등록되어 있는지를 검사한다. 같은 ID의 콘텐츠 데이터가 저장되어 있으면, 이는 이미 동일한 데이터가 백업 서버에 등록되어 있는 것이므로 백업 관리자는 해당 콘텐츠 데이터를 백업 서버로 전송하지 않고, 다음 콘텐츠 데이터에 대해 백업 요청을 수행한다. 만약 백업 관리자가 보낸 콘텐츠 ID가 존재하지 않으면, 이는 동일한 데이터가 백업 서버에 등록되어 있지 않은 것이므로, 백업 관리자는 콘텐츠 ID에 해당하는 콘텐츠 데이터

를 백업 서버로 전송한다. 따라서 Data1이 먼저 백업된 상태에서 Data1'의 백업이 수행되면 Data1'은 실제 백업 서버의 저장소에 기록되지 않고, 백업되었다는 기록과 함께 사용자에게 할당된 백업 저장소의 크기만 갱신하게 된다.

백업이 완료되면 콘텐츠 데이터 관리 모듈이 디지털 콘텐츠에 할당된 콘텐츠 데이터의 ID와 메타데이터가 데이터베이스에 저장된다. 또한 백업 관리자 및 백업 서버에서 작업을 처리하면서 발생한 로그 정보 역시 저장된다. 작업이 완료되면, 해당 작업 결과는 메시지 형태로 백업 어플리케이션에 통보된다.

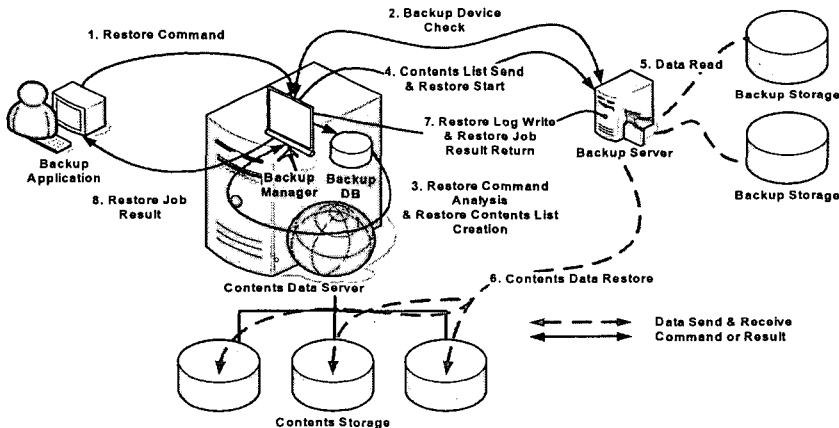


그림 6. 제안하는 백업 시스템의 복구 처리 과정



#### 4. 복구 처리

[그림 6]은 제안하는 백업 시스템의 복구 처리 과정을 나타낸다. 사용자는 복구 요청 시 복구 되어진 데이터의 위치를 선택할 수 있다. 사용자가 임의의 복구 위치를 지정하지 않으면 복구 명령에 의해 복구되는 디지털콘텐츠의 저장 위치는 기본 설정에 의해 원래 디지털콘텐츠가 있던 위치로 복구된다.

제안하는 백업 시스템의 복구 처리 과정은 일반적인 백업 시스템의 복구 처리 과정과 유사하다. 제안하는 백업 시스템의 복구 처리 순서는 다음과 같다.

- ① 사용자는 백업 어플리케이션을 사용하여 복구 명령을 백업 관리자에게 전달한다.
- ② 백업 관리자는 백업 서버의 점검과 사용자의 계정 정보를 점검한다.
- ③ 백업 관리자는 복구할 콘텐츠에 대해 히스토리 정보를 검색하여, 백업된 콘텐츠의 메타데이터와 콘텐츠 ID 정보를 가지고 복구 목록을 작성한다.
- ④ 복구 목록을 백업 서버로 전송한다.
- ⑤ 백업 서버는 콘텐츠 ID를 이름으로 갖는 파일을 읽는다.
- ⑥ 읽은 콘텐츠 데이터를 순서에 따라 디지털콘텐츠 서버에 전달한다.
- ⑦ 백업 서버는 복구 작업 결과를 백업 관리자에게 전달하고, 백업 관리자는 복구와 관련된 로그 정보와 작업 로그 정보를 기록한다.
- ⑧ 작업 결과를 사용자에게 전달한다.

### IV. 제안하는 디지털콘텐츠 백업 시스템의 구현

본 논문에서 제안하는 백업 시스템의 구현 환경은 펜티엄IV 3GHz 프로세서에서 512Mbytes의 메모리를 가지며, 운영체제로 Windows Server 2003을 사용하였다. 개발 언어로는 MFC, C++를 사용하였다.

#### 1. 디지털콘텐츠 백업 서버 인터페이스

[그림 7]은 디지털콘텐츠 백업 서버의 전체적인 레이

아웃이다. 로그 뷰어는 서버의 동작 상태를 알려주는 메시지를 출력하는데 처음 서버를 구동시켰을 때 서버에게 할당된 로컬 IP와 서비스에 사용될 포트번호를 출력한다. 또한 백업 서버가 수신한 백업 관리자의 여러 명령 및 데이터 관련 메시지를 확인할 수 있다. 제어판에서는 서버의 시작과 중지를 선택할 수 있고 서버에 접속이 가능한 계정 정보를 보여준다. 또한 서버의 동작을 설정할 수 있다.

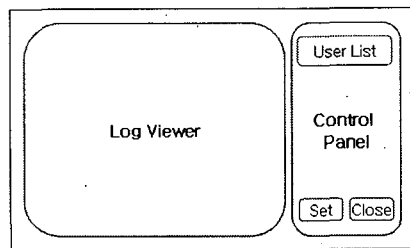


그림 7. 제안하는 백업 서버의 레이아웃

#### 2. 디지털콘텐츠 백업 관리자 인터페이스

디지털콘텐츠 백업 관리자의 기능은 탭을 사용하여 백업 스케줄 관리, 정기 백업 및 복구 기록, 환경 설정으로 구분되어 있다. [그림 8]은 백업 스케줄 설정 화면이다.

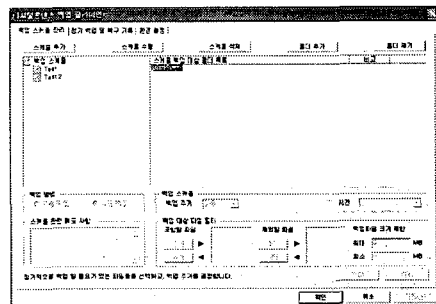


그림 8. 백업 스케줄 관리 화면

백업 스케줄 관리 탭에서는 백업 작업의 단위라고 볼 수 있는 백업 스케줄을 등록하고 등록된 스케줄을 수정하거나 삭제가 가능하다. 스케줄의 이름을 입력하고 서비스를 제공하고 있는 백업서버를 선택할 수 있다. 그리

고 사용자가 수동으로 백업 작업을 수행시킬 것인지 아니면 특정 시간이 되면 자동적으로 백업 작업이 수행되게 할 것인지를 선택하며 자동 백업의 경우 언제 수행되게 할 것인지를 지정 한다. 다음으로 각각 파일마다 실제 백업이 이루어질 때 파일의 확장자를 이용해 어떤 파일을 백업 작업에 포함을 시킬 것인지 어떤 파일을 백업 작업에서 제외시킬 것인지 지정하고 파일의 크기를 이용해 필터링이 가능하다. -1의 값은 사용하지 않는다는 의미이며 최대값과 최소값을 지정하여 그 이내의 크기를 가지는 파일만을 백업 하는 것이 가능하다.

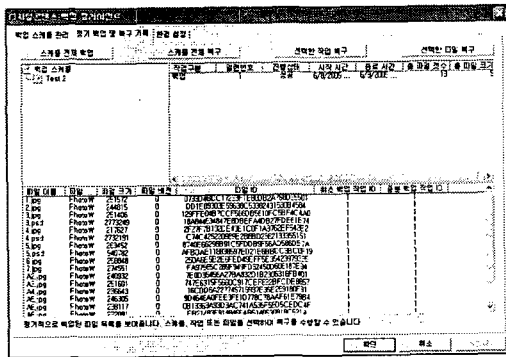


그림 9. 정기 백업 및 복구 기록 화면

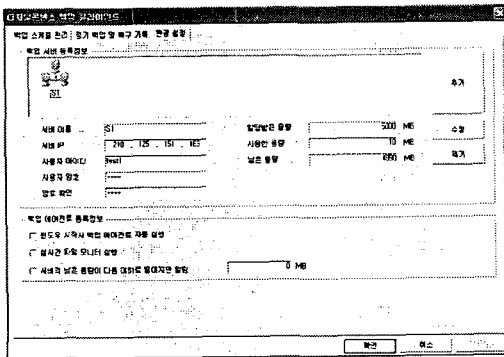


그림 10. 환경 설정 화면

[그림 9]는 실제 수행된 정기 백업 및 복구 기록 화면이다. 정기 백업 및 복구 기록 탭은 앞의 백업 스케줄 관리 탭에서 만들어진 백업 스케줄의 동작 및 복구 작업을 요청하고 결과를 확인할 수 있다. 먼저 왼쪽의 트리 뷰 창에서 앞서 만들었던 백업 스케줄들이 나타나게 되고

백업 스케줄에 대해 실제 백업 작업을 수행하는 것이 가능하다. 백업 스케줄에 의해 백업이 수행되면 화면 오른쪽의 작업 리스트에 진행 상태가 표시되고, 수행된 백업 작업이 등록된다. 백업 작업이 성공적으로 수행되었으면 백업된 데이터로부터의 복구 작업이 가능하다. 이때 백업된 데이터는 백업 작업 전체를 복구할 수도 있고, 특정 파일을 선택해서 복구할 수도 있다.

[그림 10]은 백업 서버의 환경 설정 화면이다. 환경 설정 탭에서는 현재 등록되어있는 백업서버의 이름, 서버 IP, 접속할 때 사용하는 계정정보, 해당 계정에 대해 할당된 용량, 사용한 용량 그리고 남은용량 등 상세한 정보를 볼 수 있고 만약 백업 서버가 하나이상 존재할 경우 추가 버튼을 눌러 여러 개의 백업 서버를 등록하는 것도 가능하다.

### V. 성능평가 및 분석

제한하는 백업 시스템은 방대한 양의 디지털콘텐츠를 분석을 통해서 일정한 크기로 나누고 중복되는 데이터를 없애도록 설계하고 구현되었다. 구현한 백업 시스템의 성능을 평가하기 위하여 다양한 종류의 백업 데이터를 사용하여 성능을 비교, 분석하였다. 성능평가에 사용된 시스템은 펜티엄IV 3GHz 프로세서에서 512Mbytes의 메모리를 가지며, 운영체제로 Windows Server 2003을 사용하였다.

실험에 사용된 디지털콘텐츠 데이터는 인터넷 사이버 강의에서 사용되는 강의 자료 중에서 총 15장으로 이루어진 자료구조 인터넷 강의 자료를 이용하였다. 강의 자료는 텍스트, 동영상, 애니메이션, 이미지 등의 다양한 디지털콘텐츠들로 구성된다. 강의 자료를 구성하는 각 데이터의 중복도는 [표 4]와 같다.

기존 백업 시스템들은 대부분 널리 알려진 압축 기술을 백업에 활용하고 있다. 따라서 본 논문에서는 성능 평가를 위해 기존 백업 시스템들과 직접 비교를 하지 않고, 일반적으로 널리 사용되는 압축 소프트웨어를 사용하여 데이터를 압축한 경우와 제한하는 백업 시스템을 이용한 경우를 비교한다.

실험은 실험 데이터를 10명의 사용자에게 배포 후, 일반 압축과 제안하는 백업 방식을 수행하여 백업된 데이터의 크기를 비교한다. [그림 11]은 백업하기 위한 원본 데이터 크기, 백업 데이터를 zip 압축한 데이터 크기, 논문에서 제시한 백업 시스템을 이용하여 백업한 데이터의 크기를 각각 비교한 것이다. 실험 결과 일반 압축 프로그램을 이용하여 백업한 데이터와 비교하여, 논문에서 제안하는 백업 시스템에 의해 백업된 데이터의 사이즈가 크게 줄어든 것을 확인할 수 있다. 이는 논문에서 제안하는 백업 시스템은 중복되는 데이터를 하나의 사본만 저장하여 공유토록 함으로써, 많은 중복 데이터를 제거하여 실제 저장되는 데이터의 양을 줄일 수 있기 때문이다.

표 4. 실험에 사용된 데이터의 중복도

데이터 종류	중복도
텍스트	91.5%
동영상	38.7%
애니메이션	83.6%
이미지	73.2%

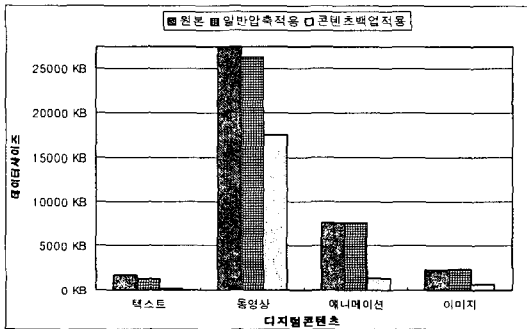


그림 11. 저장기술별 백업된 디지털콘텐츠 크기

실험결과 데이터의 중복이 가장 많은 텍스트 파일의 경우 원본 파일에 비하여 91.2%, zip 압축 데이터보다 30.1% 높게 저장 데이터의 사이즈가 줄었다. 또한 실험한 여러 콘텐츠들도 높은 비율로 백업 데이터의 사이즈가 감소하여 논문에서 제안하는 백업 시스템이 저장 공간 활용 면에서 높은 성능을 보이고 있다는 것을 알 수 있다.

## VI. 결론

본 논문에서는 첨단 IT 기술의 발전과 비례하여 방대해지는 디지털콘텐츠의 효율적인 백업을 위해 디지털콘텐츠를 객체 단위로 저장하고 관리하는 백업 시스템을 설계하고 구현하였다. 구현한 백업 시스템은 콘텐츠 기반 스토리지(CAS) 기술을 적용하여 방대한 양의 디지털콘텐츠를 분석을 통해 일정한 입출력 단위로 나누고, 유일한 ID를 부여하여 백업 및 복구를 수행한다. CAS 기술을 활용하여 중복되는 데이터는 하나의 백업 사본만 유지하도록 함으로써, 기존의 백업 시스템에 비하여 디지털콘텐츠의 백업에 사용되는 저장 공간 사용량을 획기적으로 줄였다. 또한 백업된 콘텐츠 데이터를 데이터에 할당된 ID를 사용하여 OSD나 CAS와 유사하게 객체 단위로의 접근이 가능하므로, 중복된 디지털콘텐츠 간에 하나의 백업 사본을 통한 백업 데이터 공유가 가능하도록 하였다.

향후 연구방향으로 저장 공간을 보다 효율적으로 사용하기 위하여 백업 데이터를 압축하여 저장하는 방법과 기존 백업 시스템과 비교하여 제공되지 않는 기능에 대한 연구를 진행할 예정이다.

## 참고 문헌

- [1] IITA, IT 차세대 성장 동력 기획보고서(소프트웨어 솔루션 및 디지털콘텐츠), 2003.
- [2] KISDI, 콘텐츠의 산업화에 따른 시장변화 및 발전 전략 연구, 2002.
- [3] KIPA, 2003 국내 디지털콘텐츠산업 시장조사 보고서, 2004.
- [4] Storage Networking Industry Association, Backup/Recovery Tutorial, 2001
- [5] [http://www.veritas.co.kr/Products/www?c=collateral&refId=2&dn=ko\\_KR](http://www.veritas.co.kr/Products/www?c=collateral&refId=2&dn=ko_KR)
- [6] <http://www.microsoft.com/korea/technet/security/guidance/secmod201.asp>
- [7] M. Mesnier, G. R. Ganger, and E. Riedel,

"Object-Based Storage," IEEE Communications Magazine, pp.84-90, 2003.

[8] <http://korea.emc.com/>

[9] W. C. Preston, *Using SANs and NAS*, O'REILLY, 2002.

[10] A. L. Chervenak, V. Vellanki, Z. Kurmas, and V. Gupta. "Protecting File System, A Survey of Backup Techniques," Proceedings of Joint NASA and IEEE Mass Storage Conference, 1998.

[11] W. C. Preston, *UNIX Backup and Recovery*, O'REILLY, 1999.

[12] K. Keeton and E. Anderson, "A Backup Appliance Composed of High-Capacity Disk Drives," HotOS, p.171, 2001.

[13] R. L. Rivest, "The MD5 Message Digest Algorithm," Request for Comments(RFC) 1321, Internet Activities Board, 1992.

저 자 소 개

이 석 재(Seok-Jae Lee)

정회원

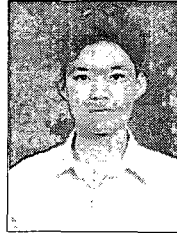


- 2000년 2월 : 충북대학교 정보통신공학과(공학사)
- 2002년 2월 : 충북대학교 정보통신공학과(공학석사)
- 2002년 3월~현재 : 충북대학교 정보통신공학과 박사과정

<관심분야> : 데이터베이스 시스템, 저장 시스템, 주 기억장치 데이터베이스 시스템, 센서 네트워크 등

윤 종 현(Jong-Hyun Yun)

준회원

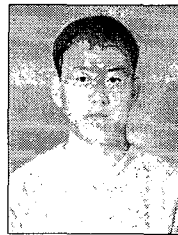


- 2003년 2월 : 충북대학교 전기전자및컴퓨터공학부 정보통신공학전공(공학사)
- 2005년 2월 : 충북대학교 정보통신공학과(공학석사)
- 2005년 3월~현재 : 충북대학교 정보통신공학과 박사과정

<관심분야> : 데이터베이스 시스템, 저장 시스템, 시공간 색인 구조, 이동 객체 데이터베이스, 센서 네트워크 등

황 석 철(Sok-Cheol Hwang)

준회원



- 2004년 2월 : 한국방송통신대학교 컴퓨터과학과(이학사)
  - 2004년 3월~현재 : 충북대학교 정보통신공학과 석사과정
- <관심분야> : 데이터베이스 시스템, 센서 네트워크 등

유 재 수(Jae-Soo Yoo)

중신회원



- 1989년 2월 : 전북대학교 컴퓨터공학과(공학사)
- 1991년 2월 : 한국과학기술원 전산학과(공학석사)
- 1995년 2월 : 한국과학기술원 전산학과(공학박사)

- 1995년~1996년 8월 : 목포대학교 전산통계학과 전임강사
- 1996년 8월~현재 : 충북대학교 전기전자컴퓨터공학부 부교수

<관심분야> : 데이터베이스 시스템, 정보검색, 멀티미디어 데이터베이스, 분산 객체 컴퓨팅 등