

DDDS(Dynamic Delegation Discovery System) 알고리즘을 활용한 콘텐츠 식별 시스템 설계 에 대한 연구

김균영*, 김창현*

*고려대학교 컴퓨터과학기술대학원

e-mail : rbsdud@hitel.net

A Design of Digital Content Identification System based on DDDS Algorithm

Kyun-Young Kim*, Chang-Hun Kim*

* Graduate School of Computer Science and Technology, Korea University

요 약

오늘날 인터넷 상의 디지털 콘텐츠는 기하급수적으로 증가하여, 콘텐츠를 관리하기 위한 다양한 시스템과 관리 체계가 연구 개발되고 있다. 특히 2000 년대부터 식별체계에 대한 표준화 연구가 진행되면서 논문등의 문헌 정보와 음악, 영상 등의 콘텐츠에 식별 체계를 적용하고 관리하려는 연구가 활발히 진행중이다. 특히 국내에서는 정보통신부와 한국소프트웨어진흥원이 디지털콘텐츠 유통기반을 구축하기 위해 CONPIA.COM 사이트를 개발하였고, 2000 년 정보화지원사업으로 한국데이터베이스진흥센터에서 디지털콘텐츠 식별체계 사업을 진행하였다. 이후 2001 년 정보통신부가 한국전산원을 디지털콘텐츠 운영기관으로 지정하면서 확장개발 및 운영 시험 등 향후 서비스 운영에 대한 기술개발에 집중하고 있다. 이러한 시스템에 적용된 식별 체계는 체계적인 구문 구조를 가지고 있으며 구문구조에 맞는 식별 코드를 해석하여 디지털콘텐츠를 찾는 구조를 가진다. 그렇지만 기존의 시스템은 식별체계의 확장과 변화에 따라 시스템 인터페이스를 재설계하고 내부 처리 알고리즘을 수정해야 하는 등 많은 수정이 요구된다. 이에 본 논문에서는 DDDS 알고리즘을 적용하여 식별 코드를 해석하기 위한 시스템을 구성하고 사용자 인터페이스를 설계하였다.

1. 서론

오늘날 디지털콘텐츠는 양과 질적인 측면에서 폭발적인 성장을 하였다. 이러한 디지털콘텐츠를 서비스하는 기관, 단체, 업체에서는 자체적인 식별 체계와 서비스 체계를 구축하여 서비스를 실시하고 있다. 이러한 서비스는 인터넷에서 URL(Uniform Resource Locator) 기반으로 제공하고 있다.

URL 은 인터넷에서 접근 가능한 디지털콘텐츠의 주소를 표현할 수 있는 형식을 말한다. URL 은 디지털콘텐츠에 접근하는데 필요한 프로토콜의 이름과 컴퓨터를 구별하기 위한 도메인 이름, 그리고 디지털콘텐츠의 위치를 계층적으로 표현한 경로명 등으로 구성된다. 이러한 구성은 컴퓨터의 식별과

디지털콘텐츠의 위치를 표현하는 면에 있어 효율적이다. 그러나 이러한 구성이 가지고 있는 단점은 디지털콘텐츠의 위치변경이나 시스템의 주소변경이 발생하면 디지털콘텐츠에 접근할 수 없으며 또한 디지털콘텐츠의 복제나 변경이 발생하였을 경우에 추적이 어렵다는 점이다. 이러한 URL 의 단점을 보완하기 위해 URN(Uniform Resource Name)의 개념이 나타났으며, 현재 문헌 정보, 논문등의 학술정보 식별과 디지털콘텐츠 유통을 유통하고 서비스하는 여러 분야에서 빠르게 확산되고 있다.

URL 은 고유한 의미를 가지는 이름으로 인터넷 자원을 찾는 방식이다. 이는 콘텐츠의 인터넷 위치가

아닌 콘텐츠 그 자체를 식별하기 위한 체계로서 URL 기반으로 처리되는 기존의 인터넷 주소체계를 보완할 수 있다. 따라서 고유성과 영구성을 지닌 URN 을 차세대 인터넷 식별체계로 채택하는 것이 세계적인 추세이며, 현재 CID(Content Identifier), DOI(Digital Object Identifier), ISAN(International Standard Audiovisual Number), ISBN(International Standard Book Number) 등 다수의 식별체계가 국제적으로 통용되고 있다.[1]

URN 을 사용함으로써 콘텐츠의 불법 유통 방지와 통합 과금, 콘텐츠의 통합 관리가 가능해 지며 콘텐츠의 이동, 변경에 따른 콘텐츠 접근의 어려움이 줄어든다. URN 체계를 적용하여 URL 식별 코드와 콘텐츠의 위치를 맵핑하기 한 다양한 시도가 진행되고 있다. 본 논문에서는 URN 식별 코드를 콘텐츠 위치로 맵핑하기 위해 DDDS 알고리즘을 적용한 핸들 시스템을 설계하고, 이를 평가하는 것이 목적이다.

본 논문의 구성은 다음과 같다. 2 장 관련연구에서는 국내외 기술 개발 현황과 표준화 추진 현황 및 적용 기술에 대해 서술하고, 3 장에서는 식별체계를 해석하는 DDDS 알고리즘과 사용자 간의 인터페이스 구조를 나타내었다. 4 장에서는 DDDS 알고리즘을 적용한 식별코드 해석기와 사용자 인터페이스에 대한 설계 및 평가하였고, 5 장에서는 결론 및 향후 계획을 서술하였다.

2. 관련연구

2.1 기술개발 현황

2.1.1 국내기술 개발현황

국내에서는 2000 년 정보화지원사업으로 한국 데이터베이스 진흥센터 "디지털콘텐츠식별 체계 사업"으로 시스템을 구축하였다. 2000 년 8 월부터 2001 년 5 월까지 약 11 억 6 천만원이 소요되었으며, 등록관리, 검색, 유통관리, 참조링크, 변환시스템으로 구성되었다.

해당 시스템은 URN 체계의 수정, 생성, 삭제 등 변환모델 응용에도 우수하도록 적합성이 뛰어나게 구축되었다. 현행 시스템은 미국의 메타데이터 체계인 DOI(Digital Object Identifier)를 기준으로 구축되었다.

현재는 2001 년 10 월 정보통신부가 한국전산원을 디지털콘텐츠식별시스템 운영 기관으로 지정하여, 2002 년 1 월부터 동 시스템에 대한 확장개발 및 운영 시험등 향후 서비스 운영에 대한 기술개발에 박차를 가하고 있다.[2]

2.1.2 국외기술 개발현황

메타데이터 체계의 기술개발은 미국, 일본에서 상용화 수준의 개발중에 있다. 미국의 경우 1994 년 CNRI 에서 URN 체계의 핵심 시스템으로 개발 운영 중인 Handle System 을 기반으로, 1997 년 미국출판협회(AAP)의 주도로 과학, 기술, 의학분야 전자저널(E-Journal)의 식별 및 유통을 위해

DOI(Digital Object Identifier)를 개발, 운영 중에 있다.

일본은 cIDf(Content ID Forum)에서 콘텐츠 및 유통속성을 기술하기 위해 개발한 식별기호 및 메타데이터 관리시스템으로서 CID 는 등록번호, 콘텐츠 메타데이터, 유통메타데이터, 시스템 제어 데이터로 구성된다.

2.2 표준화 추진현황

가. 국내 표준화추진현황

지난 2002 년 TTA 응용데이터 연구반에서는 미국의 ANSI/NISO 의 디지털콘텐츠 식별체계 표준인 DOI 를 기반으로 국내 고유의 식별자 구문구조 명세를 정의한 "디지털콘텐츠 식별자 구문구조"에 관한 표준을 개발, 확정하였다.

즉, DOI 문자집합 및 접두부, 접미부를 정의함으로써 국내 디지털콘텐츠의 유통을 위한 기본체계를 마련하였다. 그 외, 메타데이터 체계를 위한 구성 요소에 대한 정의는 한국인터넷정보센터에서 수행한 "URN 활용 체계를 위한 메타데이터 체계개발 연구"를 통해 총 6 가지의 필수 메타데이터를 제안했다.

요 소	정 의	예
URNs	디지털콘텐츠를 식별할 수 있는 유일한 식별자	MCI, DOI, cID, DCI
Identifier	지존에 부여받은 콘텐츠의 식별번호	URL, ISBN, ISSN
Title	디지털콘텐츠의 제목	URN 체계 활용을 위한 메타데이터 개발
Type	디지털콘텐츠의 유형	Text, Music, Movie
Format	디지털콘텐츠의 형식	Hwp, Doc, Avi, Mp3
Agent	디지털콘텐츠와 관련된 주요 행위자의 이름이나 식별자	한국인터넷정보 센터, IETF
Agent Role	주요 행위자의 역할	저작권자, 유통업자

[표 1] 제안된 6 가지의 필수 메타데이터

이상과 같은 필수 메타데이터는 향후 TTA 단체 표준 및 국제표준을 추진할 예정이며, URN 체계 활용 및 메타데이터 검색 활용에 이용될 전망이다.[3]

나. 국외 표준화 추진현황

URN 체계의 식별자에 대한 표준은 미국의 DOI 및 일본의 CID 및 미국 OCLC/NCSA 의 더블링크어가 있다.

DOI는 디지털 환경에서 지식정보자원의 식별을 위해 이용되는 DOI의 정의, 형식 및 특징등을 정의한

‘Syntax for the Digital Object Identifier’ 로 2000년 미국의 표준화 기구인 ANSI/NISO에서 국가표준으로 채택되었다.[4, 5]

DOI를 구성하는 각 요소는 다음과 같다.

- DOI AP : 공통되는 특성을 가진 객체의 범주를 나타낸다.
- 식별자 : DOI 이외의 식별자로 문자/숫자로 구성된다.
- 제목 : 객체가 나타내는 이름을 뜻한다.
- 타입 : 객체의 주요한 구조 형태를 뜻한다.
- 모드 : 콘텐츠의 형태로서 Visual, Audio 등으로 표현될 수 있다.
- 주요 행위자 : 주요 행위자(Primary Agent)의 이름이나 식별자를 뜻한다.
- 행위자의 역할 : 주요 행위자에 의해 행해진 역할을 뜻한다.

cIDf는 2000년 3월 spec 1.0을 발표하였으며, 현재 국제표준화기구(ISO/IEC)를 통해 표준화를 추진 중이다. cIDf는 Identifier, Public Description, Full Description으로 구성되어 있으며, 각각의 내용은 다음과 같다.

- Identifier : 식별자를 뜻하며, html/xml 문서에는 텍스트 형식으로 표현되며, 인쇄물의 경우에는 인쇄,바코드,워터마킹의 형식으로 삽입하여 사용한다.
- Public Description : 검색의 목적으로 공개 저장소에 저장, 공개되는 정보로 발급 후 변하지 않는 정보이다. 제목/저자/창작일 등을 의미한다.
- Full Description : 비공개 저장소에 저장되어 일반 대중의 접근이 통제되는 정보로 발급 후 계속 변할 수 있는 정보이다. 로열티 정보/저작권자 등을 의미한다.

더블린 코어는 네트워크 기반 정보자원의 검색에 적합하면서도 작성이 간단한 핵심 데이터 요소를 제안하였다. 현재 15개의 데이터 요소로 제안되어 있으며, 각 범주에 따른 구성요소는 다음과 같다.[1]

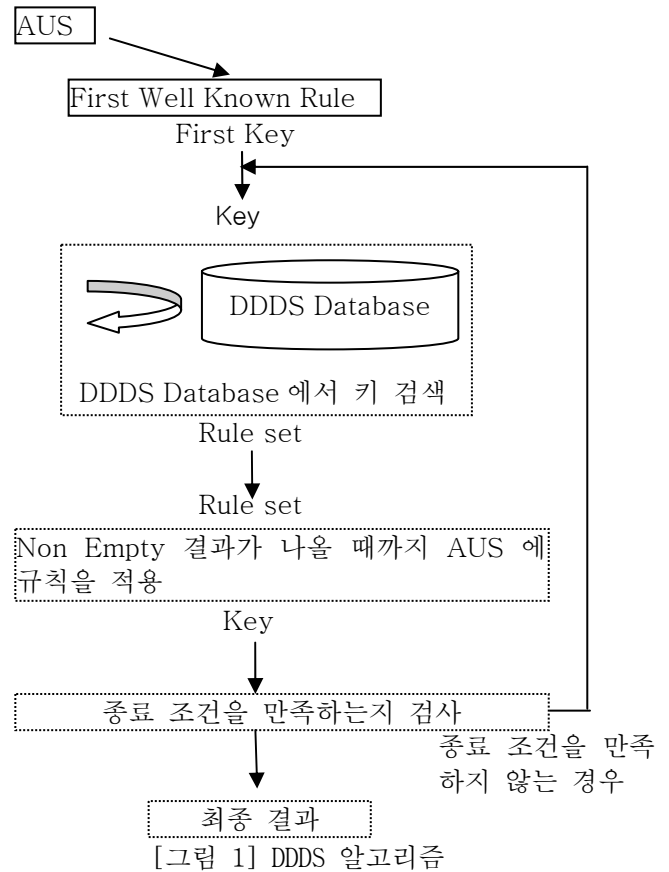
- 내용 : 제목, 주제, 표현, 출처, 관계, 범위
- 지적 소유권 : 창작자, 출판자, 기여자, 권리
- 형식 : 날짜, 타입, 포맷, 식별자

3. DDDS 알고리즘과 사용자간의 인터페이스 구조

3.1 DDDS 알고리즘

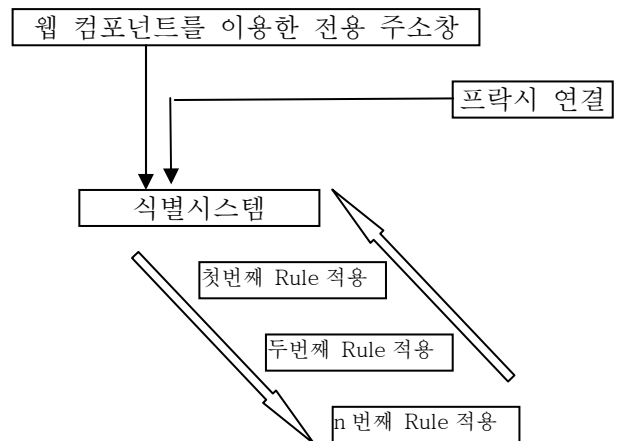
The DDDS 알고리즘은 Rewrite Rules 의 개념에 기반하며 이 규칙[판결]은 DDDS Rule Database 안에 모인다. 그리고 고유한 키 정보에 의해 접근한다.

Application Unique String 에 적용되었을 때 주어진 규칙은 Rule Database 에서 새로운 Rule 를 검색하는 데 이용될 수 있는 새로운 Key 로 그 String 을 변형시킨다. 그리고 나서 이 새로운 규칙[판결]은 독창적인 Application Unique String 까지 다시 신청되고 그 주기는 끝나는 조건이 도착될 때까지 되풀이한다.



3.2 사용자와의 인터페이스 구조

사용자와의 인터페이스는 웹 주소창을 이용하여 프락시로 식별 코드를 입력하면 시스템에서 이를 받아 식별 코드에 해당하는 콘텐츠 주소를 반환하게 되는 구조를 가지며 다음과 같다.



사용자 인터페이스는 프락시 서버를 이용하여 AUS 를 입력 받는 경우와 웹 주소창 외의 별도의 입력창을 만들어 입력 받을 수 있는 구조를 갖는다.

4. 식별코드 분석과 사용자 인터페이스에 대한 설계 및 평가

AUS 해석기는 DDDS 알고리즘을 기반으로 하여

주어진 콘텐츠 식별코드를 정해진 규칙에 의해 해석하여 콘텐츠의 위치(URL)를 찾는 구조로 되어 있다.

구축에 사용된 환경은 다음과 같다.

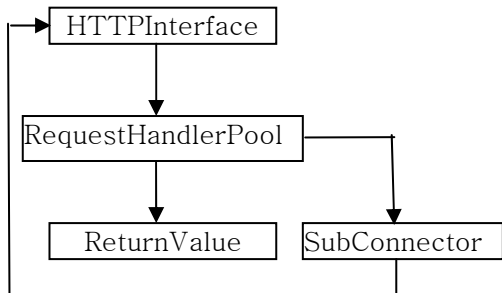
	식별코드 해석기	식별코드 입력기
운영체제	Solaris	Windows XP
개발언어	Java	C++
데이터베이스	oracle	-

[표 2] 시스템 구축 환경

시스템 구성은 크게 식별코드 입력단과 규칙 정의에 따른 식별코드 분석단과 규칙에 따라 나온 키 값을 타 시스템에 전달하는 전송단으로 나뉘어진다. 프로그램 구성을 보면 다음과 같다.

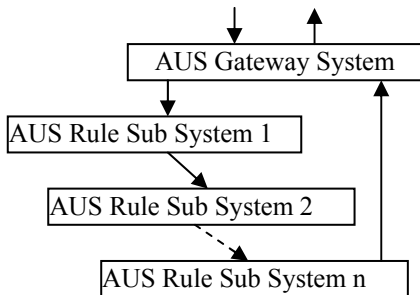
식별코드 입력단 : Class HTTPInterface (Class NetworkInterface 에서 상속) 에서 프락시 포트 8000 번을 통해 데이터 받음

식별코드 분석단: class RequestHandlerPool 에서 AUS 를 파싱하고 다음 키값을 추출하여 정의된 규칙에 따라 타 시스템으로 전달한다. 마지막 키 조건인 경우 추출된 키 값을 최종 결과로 리턴한다.



[그림 3] 프로그램 구성

식별 코드를 부여받은 콘텐츠를 식별코드를 이용하여 원래의 콘텐츠 위치를 해석하는 과정을 보인 것으로 프락시 서버, 포트 8000 번을 이용하여 식별 코드 해석을 요구하며 이를 해석하여 콘텐츠의 위치를 최종 결과로 리턴해 주는 과정을 보여준다. 시스템 간의 관계를 도식화 하면 다음과 같다.



[그림 4] 서브 시스템 간의 관계

URN:100.1300/ADO-2004-02-27-15285-BOF75E517633D6E050EAAE2536915AFA



[그림 5] 식별코드 분석과정

5. 결론 및 향후 연구방안

프락시 서버를 통해 콘텐츠 식별 코드를 입력받아 DDS 알고리즘을 활용하여 콘텐츠 위치 정보를 얻는 시스템을 설계하였다. 단일 시스템에서 구현이 가능하지만 시스템 운영 주체에 따라 분리할 필요가 있는 콘텐츠 유통 환경에서 콘텐츠 별 관리 주체에 따른 시스템 식별 체계를 통합하는 과정에서 DDS 알고리즘은 유용하게 활용될 것이다. 별도의 시스템으로 개발되었지만 DNS 와 연계, 통합하는 방안이 연구되어야 할 것이며 또한 사용자와의 인터페이스는 계속 보완되어야 할 과제이다.

참고문헌

- [1] 한국교육학술정보원, “디지털정보에 대한 식별자 부여 및 전자상거래용 메타데이터 모델에 대한 연구”, 1999.12
- [2] 디지털콘텐츠포럼, “URN 기반 식별체계 시스템 운영지침 및 운영정책 연구”, 2003.11
- [3] 한국정보통신기술협회(한국인터넷정보센터), “URI 표준화포럼구성 및 운영”, 2002
- [4] 한국정보과학회, “DOI 기반의 디지털 콘텐츠 유통 관리시스템 설계 및 구현”, 2000.10
- [5] Norman Paskin, “DOI: Current Status and Outlook”, D-Lib magazine, 1999.5