

# 정보통신 표준화 지식관리 시스템(SKMS)의 구현

이준섭\* · 송기평\* · 구경철\* · 손 홍\*

\*한국전자통신연구원 표준연구센터

## Development of Standards Knowledge Management System

Jun-seob Lee\* · Gi-pyeong Song\* · Kyung-cheol Koo\* · Hong Sohn\*

\*ETRI/Protocol Engineering Center

E-mail : {jslee, gpsong, kckoo, hsohn}@pec.etri.re.kr

### 요 약

우리는 정보의 홍수 속에서 살고 있다. 정보통신 표준화 정보도 예외는 아니어서 여러 웹사이트에서 정보통신 표준화 관련 정보를 검색하고 검색 결과에서 원하는 정보를 찾아 이용하는 일이 쉽지 않다. 예를 들어 표준원문의 경우 대부분의 표준 기구에서 서비스를 제공하고 있으나, 사용자가 표준 원문을 찾기 위해서는 표준의 번호나 제목을 알고 있어야 한다.

본 논문에서는 표준화 관련 웹사이트에서 관련 정보와 표준원문 등을 수집하여 미리 정의된 키워드에 따라 분류하고 사용자가 원하는 표준화 정보를 지식 맵을 통해 쉽게 검색하고 원하는 정보를 사용할 수 있도록 하는 정보통신 표준화 지식관리 시스템의 구현 사례를 설명한다.

### ABSTRACT

We live in the inundation of information. Information of telecommunication standardization is also overflowing. Accordingly it is not easy to search information from web sites scattered all over the world and to find information which we want to use from the result of searching. For instance, none the less most of standardization organization provides their own search facilities on specification, it is difficult to find specification without information about the specification such as number or title of that specification.

In this paper, we introduce Standards Knowledge Management System(SKMS) which provide the easy way of finding information using knowledge map. The SKMS collects information from the standardization-related web sites and categorizes information according to keyword. And users can easily find information with the pre-defined knowledge map.

### 키워드

표준화, 정보검색, 지능형 검색 엔진, Knowledge Map

### 1. 서 론

우리는 정보의 홍수 속에서 살고 있다. 매일같이 새로 쏟아져 나오는 정보들 중에서 필요한 정보를 선별하여 수용하는 것이 점차 어려워지고 있으며, 정보통신 표준화 정보도 예외는 아니어서 여러 웹사이트에서 정보통신 표준화와 관련된 정보를 검색하고 검색된 정보를 선별적으로 수용하는 일이 쉽지 않다. 각 정보통신 표준화 관련 기구는 자체적으로 웹사이트를 구축하여 관련 정보를 공개하고 있으며, 나아가 고유의 EDH(Electronic

Documents Handling) 시스템을 개발하여 표준화와 관련된 정보의 효율적인 유통을 꾀하고 있다.

사용자의 입장에서 이러한 정보통신 관련 정보의 효율적인 검색 및 선별적인 수용은 이제 별도의 도구를 이용하지 않으면 안될 정도로 관련 정보가 대량으로 생성, 유통되고 있다. 따라서, 표준화와 관련 정보를 효율적으로 검색하고 원하는 정보를 선별적으로 수용할 수 있도록 도와주는 도구들의 개발은 필수적이라 할 수 있다. 본 논문에서는 표준화 관련 웹사이트에서 관련 정보와 표준원문 등을 수집하여 미리 정의된 키워드에

따라 분류하고 사용자가 원하는 표준화 정보를 지식 맵을 통하여 쉽게 검색하고 원하는 정보를 선별적으로 수용할 수 있도록 하는 정보통신 표준화 지식관리 시스템(SKMS: Standards Knowledge Management System)의 구현 사례를 설명한다.

## II. SKMS의 개요

### 1. SKMS의 소개

SKMS는 Windows NT 4.0 Server 환경에서 DOCSEfulcrum™사의 Fulcrum Knowledge Server와 IIS(Internet Information Server)를 이용하여 개발되었다.

SKMS는 표준화 관련 웹사이트와 특정 파일 시스템에서 정보들을 수집하여 미리 정의된 분류 규칙에 따라 문서를 분류하고, 사용자의 요청에 따라 분류된 문서를 다양한 검색 인터페이스를 통해 제공하는 시스템이다. 웹사이트와 파일 시스템의 크롤링을 위해서는 Fulcrum Knowledge Server를 이용하고, 문서의 분류는 자체적으로 개발한 Windows NT 기반의 문서 자동 분류 시스템을 사용하며, IIS를 이용하여 검색 인터페이스를 제공한다.

SKMS는 효율적인 검색 기능의 제공을 위해 세 가지 검색 인터페이스를 제공한다. 첫째로, 키워드를 이용한 검색과 둘째로, 결과 내 검색 기능, 마지막으로 미리 정의된 정보통신 표준화 지식 맵을 이용한 단계별 검색 기능을 제공한다. 또한 이러한 세 가지 검색 기능이 서로 상호 보완적으로 사용될 수 있으므로 사용자는 원하는 정보를 쉽게 찾아 볼 수 있다.

### 2. SKMS의 구성 및 운영

SKMS는 그림 1.과 같이 세 개의 서버로 이루어진 분산환경에서 동작한다.

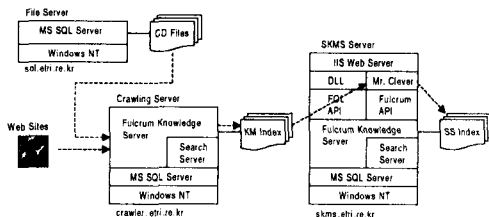


그림 102. SKMS의 구성

서버들의 구성을 자세히 살펴보면, 우선 문서 분류 작업을 수행하고 검색 인터페이스를 제공하는 SKMS 서버와 웹사이트 및 파일 시스템으로부터 정보를 수집하는 크롤링 서버, 그리고 표준원문과 같은 대용량 정보를 저장하기 위한 파일 서버가 있다.

파일 서버는 크롤링 서버와 네트워크를 통해

디스크를 공유하고 있으며, 크롤링 서버는 이 네트워크 공유를 통해 파일 시스템을 액세스 할 수 있다. 크롤링 서버는 SQL Server 4.0과 Fulcrum Knowledge Server를 이용하며, Fulcrum Knowledge Server의 웹 크롤러를 이용하여 각 웹사이트의 정보를 수집한다. 수집된 정보와 파일 시스템의 정보로부터 문서정보를 추출하여 인덱스 파일을 생성한다. SKMS 서버는 크롤링 서버가 생성한 인덱스 파일로부터 다시 한번 문서정보를 추출하여 웹사이트나 파일 시스템에서 실제로 문서가 존재하는지 확인하고 웹사이트의 경우에는 전문(Full text) 검색을 위해 문서 원본을 가져와 로컬 파일 시스템에 저장한다. 다음 미리 정의된 키워드를 포함하고 있는지의 여부에 따라 지식 맵으로 분류되며, 이 분류된 정보들은 다시 인덱스 파일로 저장된다. SKMS 서버에서 생성된 인덱스 파일은 사용자의 검색 요구 대상이 된다.

그림 1.에서 SKMS 서버의 문서 분류 기능은 Fulcrum Knowledge Server의 API를 이용하여 개발되었다.

## III. SKMS의 인터페이스

### 1. 사용자 인터페이스

SKMS의 사용자 인터페이스는 그림 2.와 같이 이루어져 있다. 키워드 검색을 위한 입력창과 지식 맵을 이용한 단계별 검색을 위한 정보통신 표준화 지식 맵, 그리고 어떤 검색 결과에서 재 검색을 수행하기 위한 입력 창으로 이루어져 있다. 키워드 검색에는 검색 대상(표준원문, 표준기구, 표준화 동향 등)의 지정 및 검색 범위(제목, 요약, 본문)의 설정, 검색 대상 문서의 생성 일자 등을 지정하여 보다 상세한 검색을 수행 할 수 있다.

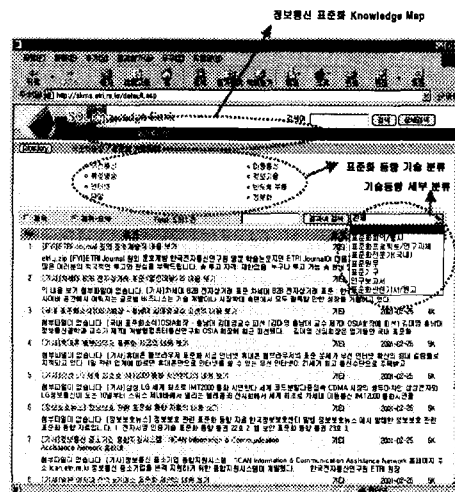


그림 103. 사용자 인터페이스

정보통신 표준화 지식 맵에 의한 검색에서는

경우에 따라 표준화 동향 기술 분류, 기술 동향 세부 분류, 표준화 기구 세부 분류와 같은 하위 지식 맵 들이 나타날 수 있다.

2. 관리자 인터페이스

SKMS의 관리자 인터페이스는 크게 그림 3과 같이 문서 분류의 규칙을 정의하는 화면과 그림 4와 같이 분류된 문서의 편집 및 수정을 위한 화면으로 이루어져 있다. 분류된 문서의 편집은 원본 문서를 변경하는 것이 아니라 검색 결과를 사용자에게 보여주는 화면에서 문서의 제목 및 요약 변경하는 것이다.



그림 104. 문서 분류 규칙 정의를 위한 관리자 인터페이스

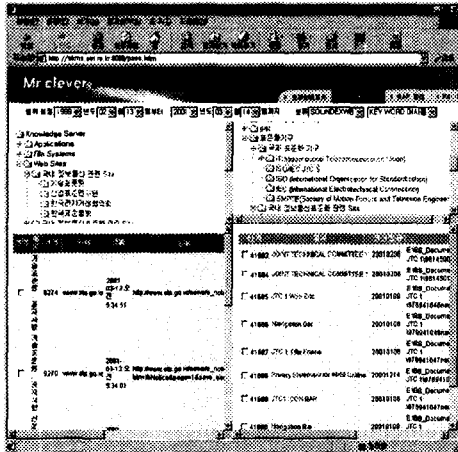


그림 105. 문서 편집을 위한 관리자 인터페이스

그림 3에서 왼쪽 상단은 크롤링 서버에서 크롤링을 위해 설정한 인덱스의 구조를 나타내며, 오른쪽 상단은 정보통신 표준화 지식 맵의 구조를 나타낸다. 왼쪽 하단은 각 웹사이트 및 파일

시스템에서 수집한 정보를 분류하기 위한 규칙 및 분류 대상을 설정하기 위한 화면이며, 오른쪽 하단은 각 지식 맵의 인덱스 파일 위치 등과 같은 정보를 수정하기 위한 화면이다. 관리자는 그림 4의 인터페이스를 통해서 이미 분류된 문서의 제목 및 요약 등을 수정할 수 있다.

IV. 문서 분류 시스템

문서 분류 시스템은 문서 분류를 위해 키워드를 사용하며 이러한 정보는 LOGICAL\_MAP\_TB 와 SITE\_CONTROL\_TB 데이터베이스 테이블에 저장되어 있다. 또한 각 문서들로부터 텍스트 정보를 추출하기 위해서는 각 문서에 맞는 Text Reader가 필요하며 이러한 정보는 TEXT\_READER\_STORE\_TB 데이터베이스 테이블에 저장된다. 표 1은 각 지식 맵의 분류 규칙을 정의하는 LOGICAL\_MAP\_TB 테이블의 스키마이다.

표 1. LOGICAL\_MAP\_TB 테이블 스키마

Column ID	의미	형태	길이
IN_SEQ	순차 번호	Int	4
VC_MAP_CODE	디렉토리 코드	Varchar	20
VC_MAP_NAME	디렉토리 명	Varchar	50
IN_BOTTOM_LEVEL	최종레벨 구분	Int	4
IN_MAP_LEVEL	디렉토리 레벨	Int	4
TX_REQUESTED_KEY	필수 키워드	Text	16
TX_OPTION_KEY	선택 키워드	Text	16
VC_EFFECT_CHECK	유효성 구분	Varchar	2
CH_WORK_OPT	분류작업 여부	Char	1
VC_MENU_CODE	MENU CODE	Char	20
VC_INDEX_TABLE	SSINDEX 명	Char	20
VC_BASE_PATH	침부 저장 경로	Varchar	200
VC_DISUSE	불용어	16	16

SKMS는 Fulcrum Knowledge Server에 포함된 ftmf text reader, nti, HwpTR, 그리고 HmmTR을 이용한다. ftmf text reader는 MS Word와 MS PowerPoint 등 200여 종류의 문서로부터 텍스트 정보를 추출할 수 있다[1]. HwpTR과 HmmTR은 각각 HWP와 훈민정음 문서에서 텍스트 정보를 추출하기 위한 text reader이다.

현재는 Fulcrum Knowledge Server가 한글 PDF 문서를 처리하지 못하기 때문에 크롤링 작업을 수행할 때 한글 PDF 파일은 제외하고 있다.

문서 분류 시스템은 우선 크롤링 서버가 웹사이트 및 파일 시스템을 이용하여 생성한 인덱스 파일로부터 문서 정보를 추출하여 원본 문서를 웹사이트에서 가져와 고유의 식별자를 붙여 로컬 파일 시스템에 저장한다. 다음 text reader를 이용하여 원본 문서로부터 텍스트 정보를 추출하고 이 문서가 지식 맵의 각 분류에 정의된 키워드를

포함하고 있는지 검사한다. 키워드를 포함하는 경우 각 분류에 이 문서의 식별자를 포함시킨다. 따라서 각 분류 항목은 각각 문서의 사본을 저장하고 있는 것이 아니라 문서의 식별자만을 포함하게 된다.

## V. 정보통신 표준화 지식 맵

정보통신 표준화 지식 맵은 최하위에 79개의 항목으로 구성되어 있으며 그 하부에 필요에 따라 표준화 동향 기술 분류, 그리고 표준화 기구 세부 분류를 포함하고 있다. 표준화 동향 기술 분류는 최하위에 126개의 기술 분류 항목을 포함하고 있으며, 표준화 동향 기술 분류의 일부 항목은 각 단계 별로 7개의 세부 분류를 포함하고 있다. 정보통신 표준화 지식 맵은 데이터베이스 형태로 관리되며 각각의 항목은 고유의 코드로써 식별된다.

정보통신 표준화 지식 맵은 최대 4단계까지 구성되며, 표준화 동향 기술 분류와 기술 동향 세부 분류를 조합할 경우 최대 6단계까지 구성될 수 있다.

표준화 동향 기술 분류는 한국전자통신연구원 에서 개발 및 운영 중인 ITFind[2]에서 사용하는 기술 분류를 기준으로 표준화 정보의 처리에 맞도록 수정하여 사용하고 있다.

## VI. 결 론

지금까지 표준화 정보의 효율적인 검색 및 활용을 위한 정보통신 표준화 지식관리 시스템(SKMS)의 구현 내용을 설명하였다. SKMS는 표준화 관련 웹사이트 및 표준원문 파일 시스템으로부터 정보를 수집, 가공하여 분류하고 이를 여러 가지 검색 인터페이스를 통해 사용자에게 제공하는 시스템이다. 효과적인 검색을 위해 SKMS는 키워드 검색, 결과 내 검색, 그리고 지식 맵을 이용한 단계별 검색을 지원하며, 서로 다른 검색 방법을 조합하여 사용함으로써 검색의 정확성 및 효율성을 향상시킬 수 있다.

SKMS는 한국전자통신연구원 표준연구센터에서 개발 및 운영 중인 정보통신 표준화 보탈 사이트인 종합 표준정보유통망(SOL: Standards On-Line)[3]의 일부 기능으로 개발되었으며, 향후 SOL의 모든 기능을 포함하는 종합적인 표준화 지식관리 시스템으로 발전할 것이다.

현재까지 SKMS를 시험 운영하면서 발견된 문제점으로는 웹사이트 크롤링 속도 및 대량의 문서 분류 시의 속도 저하 현상을 들 수 있다. 이를 해결하기 위하여 웹사이트 문서의 크롤링을 보다 효과적으로 수행하여 필요 없는 문서가 분류의 대상이 되지 않도록 하는 방법을 연구 중에 있으

며, 분류된 문서의 검색 정확성을 높이기 위한 연구를 진행 중에 있다. 또한 보다 정확한 분류를 위해 정보통신 표준화 지식 맵의 구성에 관한 연구를 진행 중에 있다.

SKMS의 구현에 관련된 연구가 성공적으로 완료된다면 국내 정보통신 표준화 활동의 지원을 위한 정보통신 표준화 보탈(Vortal) 사이트로서 그 역할을 충실히 수행 할 수 있을 것으로 예상된다.

## 참고문헌

- [1] PC DOCS, Inc, Serch Server Administration, PC DOCS/Fulcrum Education Services Department, 2-5, 1999.
- [2] ITFind, <http://www.itfind.or.kr>
- [3] SOL, <http://sol.etri.re.kr>