

정보검색시스템 KRISTAL-II 구성 및 지원도구

장 푸 양(E-mail: lary@kortica.kr)

연구개발정보센터 정보시스템개발실 선임연구원

- I. 개요
- II. 시스템 구성
- III. 지원도구

I. 개요

연구개발정보센터(Korea R&D Information Center)는 과학과 기술에 관련된 정보의 검색 서비스를 제공하기 위해 1999년부터 현재까지 8년여 걸쳐 정보검색시스템을 개발하여 왔다.

KRISTAL-II(Korea Research Information of Science & Technology Access Line-II)라고 명명한 이 시스템은 기존의 외국산 정보검색시스템들이 지니고 있는 장점은 최대한 반영하고, 문제점인 한글 처리의 한계, 시스템 기능의 확장이나 변경의 한계, 고가의 도입 비용 등을 해결하기 위해 개발되었다.

KRISTAL-II는 부울 검색모형 및 자연어 검색모형을 기반으로 하여, 문서, 영상 등 다양한 형태의 정보를 관리할 수 있도록 설계되었다. 또한 한글 자동 색인을 제공함으로써 한글 및 영어 문서 모두를 검색할 수 있으며, 웹 인터페이스 구축을 위한 편리한 도구들을 제공한다.

II. 시스템 구성

● 커널(KRISTAL-Kernel)

- Raw Device 접근 방식의 독자적인 디스크 관리 수행
- LRU 방식의 버퍼 교환을 수행

함으로써 사용자 영역과 디스크 사이에 빠른 입출력 지원

● 저장 엔진(KRISTAL-ROSE)

- 검색 시스템의 가장 하부에 위치하는 저장 엔진으로, 비경형 텍스트 정보 저장 관리
- 카탈로그 관리기, 문서 관리기, 색인어 관리기로 구성

● 검색 엔진(KRISTAL-FIRE)

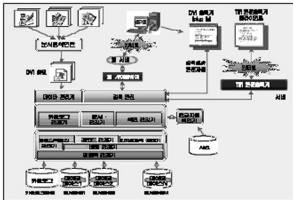
- 사용자 질의를 효율적으로 산산하는 검색 엔진
- NISO 2958을 기반으로 순수 부울 검색 모델, P-Norm 문서 우선 순위 결정 모델, P-Norm을 이용한 부울 검색 결과에 대한 문서 우선 순위 결정 모델, P-Norm과 단어간 커피 밀도를 이용한 검색 모델 구현

● 한글 자동 색인기(KRISTAL-IDX)

- 형태소 분석 기능을 이용하여 문서로부터 색인어 추출
- 불용어들을 제거하고 문서에서 의미를 갖는 단어들을 다양한 색인 방법을 이용하여 추출

● 데이터 관리기(KRISTAL-Admin)

- 기술인 스키마에 따라 카탈로그와 문서 블록 생성



●그림1 시스템 구성도

- 원시 정보를 문서 분류에 적재함으로써 데이터베이스 구축.

● 웹을 인테페이스(KRISTAL-WWW)

- 웹브라우저를 이용하는 사용자와 검색 엔진 사이의 게이트웨이 시스템.
- 필의를 입력받아 검색을 수행하고 검색된 결과를 웹서버를 강유하여 브라우저 화면에 출력.

● Tiff 원문 출력기(TifMover, TIFPrinter)

- WWW(World Wide Web)와 연동되어 클라이언트/서버 모델을 기반으로 구축.

Ⅲ. 지원 톨 :

가. 한글 자동 색인기.

한국어·어절의·색인을 위하여 서·본·시스템은 FST(Finite State Transducer)기반의·형태소 분석 모델을 구현하여 시스템화하고·효율적인 사전함조를 위하여 In-memory tree 구조의 사전을 설계하였다. 품사 태깅을 위해 어절구조를 반영한HMM(Hidden Markov Model)을 적용하였고, 분석모듈에 대한 재

학습을 통하여 신로어나·교유명사에 대한·미등록어 처리 기능을 구현하였다. 한글 자동 색인기의 주요기능 및 특징은 다음과 같다.

- 어절 분석을 위한 형태소 구성망(Morphem Transition Network) 생성.
- 10,000여개의·고등학교 국어 교과서의 용언 용례 및 2,000여 문장의·실데이터 분석을 통하여·확장·보안된·검색·규칙 및·음운현이·규칙·작성.
- 고속의 In-memory tree 구조의 품사사전 구성.
- Hybrid Hidden Markov Model에 기반한 품사·모호성·제거·시스템·개발 및·작성.
- 대응장·필문서·분석을·통한·수사·표현·어절·처리·시스템·작성.
- 형태소·분석기·품사·태깅·미등록어·처리·수사·처리·색인의·추출·기능들을·구조적으로·결합한·어절·분석·시스템.

나. 스카마 에디터(KSE : KRISTAL Scheme Editor)
KRISTAL-II 저장시스템에·데이터베이스를·구축할·때·필요한·데이터베이스·스카마·(데이터·크적 및

꼭 알아 필요한 각종 정보를 쉽고 빠르게 생성하기 위해 제공되는 도구이다. Java와 Java Swing 패키지를 사용하여 현재 2.0버전이 개발되어 있고, KSE의 주요 특징 및 기능은 다음과 같다.

- KSE의 주요 특징
 - GUI를 활용한 편리한 인터페이스
 - JAVA를 이용한 플랫폼 독립적 프로그램
 - 자동설치 도구를 활용한 쉽고 빠른 설치
- KSE의 주요기능
 - 스카라 모델 및 저장
 - 이전 버전의 KRISTAL 스카라를 2.0으로 자동 변환
 - 스카라 미리보기



■ 그림2 스카라에디션 2.0의 인터페이스 구성도

다 웹 인터페이스 자동 구축기

웹 인터페이스 자동 구축기는 사용자 인터페이스와 관련된 여러 가지 라이브러리와 구축환경을 제공하는 유틸리티로서 개발되었으며 웹 인터페이스를 개발할 때 필요한 기본적인 작업과 반복되는 작업을 자동으로 수행하기 위해 사용하고, 데이터베이스를 객체화하기 위해 사용자 스카라 파일과 사용자 입력을 통해 일련의 작업을 수행한다.

웹 인터페이스 자동 구축은 2단계로 나뉘어져 있

다. 1단계는 객체화 데이터베이스와 사용자 환경에 대한 정보를 데이터베이스 관리자가 입력하는 단계이다. 이 때 자바스크립트를 이용하여 관리자 정보가 입력하면, 입력된 정보와 관련된 정보를 자동으로 확장함으로써 입력의 편의성을 제공하였다.

2단계는 스카라 파일로부터 데이터베이스와 관련된 정보들을 분석한 후, 각종 간략 보기 상세 보기를 위한 기본적인 사용자들을 입력받아 데이터베이스에 대한 정보와 일과 기타 HTML 템플릿 파일들을 생성한다. 이러한 일련의 과정들은 신속하게 수행되므로, 데이터베이스 관리자 등 사용자 검색을 수행할 수 있는 환경을 제공한다.

자동구축기는 웹 인터페이스의 기본양식을 5개로 분류하였으며, 각각은 검색모듈, 간략 보기 모듈, 상세 보기 모듈, 검색자취 보기 모듈, 확인 및 조회 모듈이다. 웹 인터페이스 자동 구축기는 이 5개 모듈에 관한 HTML 템플릿, 프로그램 소스, 실행 파일을 제공하며, 이 이외의 기능을 확장하고자 할 경우에는 프로그램 소스를 수정 확장하여야 한다.



■ 그림3 웹 인터페이스 자동구축기 흐름도