

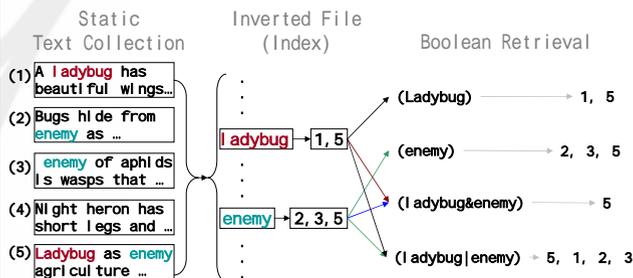
Introduction to KRISTAL-IRMS

Overview

- Introduction to KRISTAL-IRMS
- Background
- Features of KRISTAL-IRMS
- Applications
- Further Development Plans

- Installing KRISTAL-IRMS

Information Retrieval



However,

- Some documents are modified.
- New documents are created.
- Some documents are deleted.



KRISTAL-IRMS

Knowledge Retrieval In Science & Technology Affiliated Literatures

KRISTAL

Information Retrieval & Management System (IRMS) that combines the functions of an information retrieval system and a DB Management system (DBMS) developed by KISTI.

Information Retrieval

Full IRS

High-speed/Large-scale full-text retrieval
High-speed document indexing

Information Management

Partial DBMS

High-speed on-line document insert/delete/update
High-speed document loading

- KRISTAL, an IRS tightly-coupled with DBMS functions, supports
 - FULL functions of an Information Retrieval System,
 - SUBSET of data management functions of a DBMS, and
 - DOCUMENT MANAGEMENT & SERVICE **without DBMS software.**

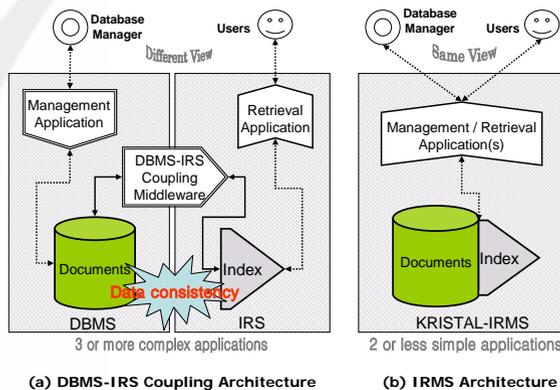
KRISTAL-IRMS History

- **KRISTAL-I**
 - 1991. 5 - 1996. 2 (Information Retrieval using BASIS+)
- **KRISTAL-II**
 - 1996. 03 – (Information Retrieval Engine)
- **KRISTAL-2000**
 - 2000. 03 – (Information Retrieval & Managmnt. System)
- **KRISTAL-2002**
 - 2002. 10 – (Information Retrieval & Managmnt. System)
- **KRISTAL-IRMS**
 - 2006. 01 – (Information Retrieval & Managmnt. System) : commercial product level

Background (1/2)

- **Motives for Development**
 - **Information Technologies based on native language/culture**
 - KRISTAL started with indexing and retrieval technologies for Korean and Chinese texts.
 - Asian languages differ from Westerns in the respect of language processing technologies as well.
 - **Complicated & Inefficient Information Management Systems**
 - Prevailed document management systems are based on application-based loose coupling of DBMS and IRS.
 - In document management & service systems, IRS is used for text retrieval and DBMS for document management. Applications are used to couple these two separate software packages.
 - These systems uses only a small subset of DBMS features to store and manage documents.
 - If this small subset of management functions is implemented on IRS, document management systems can be very simple since it can be implemented based on IRS only, without expensive DBMS.

Background (2/2)

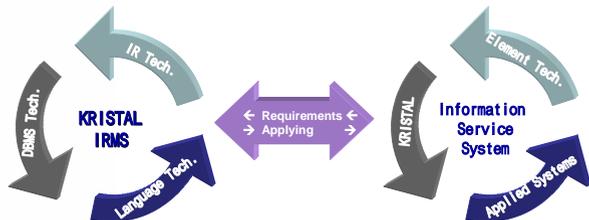


Current Trends in Document Management Systems

Systems	Examples	Merits	Demerits	Status
Application-based Loose Coupling of DBMS and IRS	DBMS for document storage system and IRS for text retrieval. Additional applications connect DBMS and IRS for text and index synch.	DBMS is strong for management and IRS is strong for text retrieval.	Two separate software packages (EXPENSIVE). Complex system architecture. Gap between a document (DBMS) and its index (IRS).	Prevailed
Loose coupling of IR features on DBMS system	Oracle Cartridge, IBM DB2's Text Extender	Extendable data types	Overhead for extended data types, SQL-based retrieval only, Slow search speed.	Almost Dead
Tight coupling of IR features on DBMS system	Odysseus (KAIST)	Simple document management	Slow search. Weak retrieval features.	Limited
Tight coupling of DBMS features on IR system	KRISTAL IRMS	Fast Search, Full IR features, Online data management	Still weak in document management in the respect of DB manager.	Will be prevailed ?

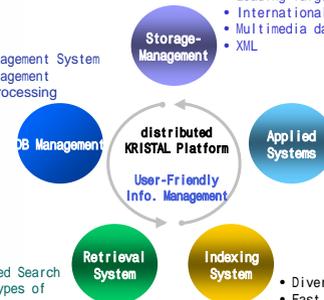
Strategic Focus on KRISTAL Development

- Focus on high-tech Information retrieval and service technologies
- Develop an extendible IRMS that combines a search engine and a DBMS
- Reflect requirements from IRMS-based information service systems



Features of KRISTAL-IRMS

- GUI-based Management System
- Simple DB Management
- Transaction Processing
- Loading large scale data at a high speed
- Internationalize through Unicode
- Multimedia data
- XML
- Applied systems run on various platforms
- Customization using APIs for each function
- Distributed Search
- Various types of retrieval model
- Compound noun extendable query processing
- Diverse Indexing Method
- Fast and accurate built-in Morphological analyzer
- Unicode-based Indexing



KRISTAL Features(1)

Document Storage and Management

- Fast uploading of large-scale data
- Stable structure not affected by the size of the document or DB
- Unicode-based documents and index storage
- XML storage and management
- Support various types of data (Text, Multimedia, BLOB, ..)

DB Management

- Guarantee online data management through transaction processing
- Provide a Primary Key for redundancy checking and management
- GUI DB management tool
- Easy DB uploading and back-up

KRISTAL Features(2)

Retrieval System

- Fast retrieval through multi-threaded database access
- Concurrent query processing through process-pool method
- High recall rate
- Vector/Boolean search model
- Similar documents retrieval and Retrieval In results

Indexing System

- Provide various types of indexing such as word/character-based indexing, morpheme analysis indexing, compound noun analysis indexing
- Provide more than one type of indexing for a single element
- Apply a Korean Morphological analyzer developed by KISTI
- Unicode-base

Application System

- Provide various types of libraries required for developing clients
- C/C++, JAVA APIs to access KRISTAL servers

Applications 4/5: Scientific Data Analysis

• Protein Sequence Analysis

- URL: <http://proses.kisti.re.kr>
- About 100 thousand of protein sequences
- Subcellular location(s) for a new protein sequence can be predicted.
- Indexing Protein Sequences
 - Overlapped Pentagram
 - Ex) ACDEFGHI → "ACDEF", "CDEFG", "DEFGH", "EFGHI"
- Automated Text Categorization



From Sequence
To Location

Applications 5/5: Other Sites (1/2)

• Scientific & Technical Information Services of KISTI

- <http://techtrend.kisti.re.kr> (Technical Trends Database)
- <http://next10.yeskisti.net> (Next Generation Technology Information Service)
- <http://www.nktech.net> (S&T Information of North Korea)
- And more...

• Full Text Search of Korean Books

- <http://www.booktopia.com> (Booktopia)

• News Photo Management Systems

- Korean Economy Daily, Kookmin Ilbo, etc. (For Intranet)

Applications 5/5: Other Sites (2/2)

• Retrieval Systems for Historical Literature

- <http://sjw.history.go.kr> (Seung-Jeong-Won Diary)
- <http://e-kyujanggak.snu.ac.kr> (Kyu-Jang-Gak)
- <http://www.minchu.or.kr> (Korean Classics Research Institute)
- And more to come.

• Retrieval System for Scientific Information

- <http://society.kisti.re.kr> (Portal for Korean Journal Contents)

• Photo Album with Full Text Search

- <http://www.animalpicturesarchive.com>

• And many more will be on-line sooner or later.

Further Development Plans

Information Retrieval Management System

- Support KNOWLEDGE CIRCULATION in Asian language texts
- Support SCIENTIFIC DATA ANALYSIS using data mining
- Do not need to buy an expensive RDBMS for document management

Asian Language Processing / Scientific Data Processing

SQL-like IMQL (Information Management Query Language)

Efficient Offline/Online Data Management

Distributed Information Management & Retrieval

Improvement of User Supporting Tools



Thank you for your attention!

<http://www.kristalinfo.com>