



DOCU/MASTER® Search Engine for IBM Mainframe CICS Applications

Edward S. Hollshwandner and John H. Bachiochi
Document Systems, Inc.
January 2004 - revised

Summary

This document outlines the DOCU/MASTER search engine functionality and its ability to index and search both structured and unstructured data within multiple legacy CICS applications on the IBM zOS, OS/390, VSE and zVSE mainframe operating systems. It reviews the overall DOCU/MASTER search and retrieval capability, its underlying architecture and technology, and its use as a generalized mainframe search engine either embedded within legacy mainframe application system(s) or as a stand-alone application.

Product Overview

DOCU/MASTER is an interactive document search and retrieval system that provides for the fast, efficient retrieval of structured and unstructured data/documents based on specific word, phrase or concept searches. DOCU/MASTER features a complete full-text searching capability utilizing Boolean (AND/OR/NOT) operators; adjacency and proximity searching; field/record and free form searching; synonym tables; and artificial indexing. Horizontal, vertical, and bi-directional search strategies are supported. The product includes standard command-, menu-driven, and book metaphor user interfaces as well as support for custom interfaces. It is used for both free-form textual information access (i.e. documentation, abstracts, research documents, resumes) and combinations of structured and unstructured data (i.e. CRM applications, engineering specifications, product inventory) by a variety of industries (i.e. insurance, government, banking, public utilities, and manufacturing).

Created as a text retrieval product for IBM mainframe operating systems, it has a highly efficient indexing algorithm that provides consistent sub-second response times, whether searching hundreds or millions of documents/records. It allows for the automatic indexing of every word and number contained in a document or data set and updates/maintains the search index on a net change basis. It operates as either a CICS or batch application in the zOS, OS/390 and VSE operating systems.

Certain inherent design characteristics of the product, principally its Application Programming Interface (API) and indexing methodology, allow DOCU/MASTER to integrate within other mainframe applications and provide the technical groundwork for subsequent product extensions/enhancements.



Product Extensions/Enhancements

The original DOCU/MASTER product, introduced in the mid-1980's, required that a separate indexed DOCU/MASTER database be created, resulting in the possible duplication of data. In addition, the original API required an online CICS application component to receive and process results from the API. The product has been improved and expanded over the years and these product enhancements have effectively reengineered the creation and delivery of the DOCU/MASTER search results.

First, the DOCU/MASTER database was 'de-coupled' from the DOCU/MASTER engine and index by creating a DOCU/MASTER 'meta-file' that references the data or document in the legacy application. This was accomplished through a proprietary method of accessing/reading the data or documents in the application's persistent data store. Second, the API was expanded and enhanced to accommodate the initiation of and response to a DOCU/MASTER search request through a control block of data of virtually unlimited size. The results of an API request can be used either within an existing application or passed to other application modules for further processing.

The end result of both these developments has been the extension of the DOCU/MASTER search capability to allow: 1) indexing and searching of any data that may exist in legacy application databases (VSAM, Adabas, DB2, IMS, flat files, etc.) without having to maintain a separate copy of this data; and 2) invoking the DOCU/MASTER search engine from any other mainframe application or module. The combination of these capabilities provides the framework for DOCU/MASTER as a generalized mainframe search engine across the multitude of disparate, legacy applications and data within an enterprise.

Technology and Architecture

The DOCU/MASTER search engine and indexing algorithm is built using IBM Assembler language for speed and efficiency. It employs a proprietary non-positional indexing method that reduces the overall index size and results in consistently sub-second response times, regardless of the size of the index, number of documents/records, or the frequency of the searched elements. While filters and stop word tables are supported, they are not required due to the efficiency of the search engine.

The DOCU/MASTER API, index updating and maintenance modules are also built using IBM Assembler. The DOCU/MASTER API may be invoked using any high-level programming language that supports external calls. DOCU/MASTER pre-processing and presentation routines may be built using any high-level programming that supports external calls. Currently, we provide application software that is written in C++ and COBOL.

Exhibit A outlines the DOCU/MASTER system architecture within the context of a generalized mainframe search engine for legacy applications/data.



Application and Uses

The DOCU/MASTER search engine enables the search and retrieval of mainframe legacy application data within the confines of the following broad categories of usage:

- 1) *DOCU/MASTER embedded within a legacy application to provide expanded search capability for unstructured (free form) and structured (fielded) data.* In this use, the DOCU/MASTER engine and index are ‘under the covers’ of an existing application. The application’s UI or screens interface directly to the DOCU/MASTER index via the API for enhanced search and retrieval functionality. Control for application presentation and other functions are then passed back to the resident system for execution. For example, State Employment Security Agencies utilize DOCU/MASTER to enhance the capability of its legacy employment security systems by matching job applicants with available job openings based on free-form resume and interview data.
- 2) *Generalized search engine across multiple, disparate legacy databases.* DOCU/MASTER’s ‘Index-Only’ capability allows for multiple data stores, with multiple data formats and structures to be indexed and searched simultaneously from a search request processed through the API to the DOCU/MASTER index. For example, a consumer products company uses DOCU/MASTER to simultaneously search R&D, engineering, product catalog and customer complaint data to facilitate product failure response and remediation efforts.
- 3) *‘Virtual data miner’ for legacy databases.* This use represents another inherent benefit of indexing both structured and unstructured data in a legacy application. This approach uses the automated DOCU/MASTER rules-based searching capabilities to interrogate the legacy data set and produce a set of results for reporting or further analysis.
- 4) *Mainframe search and retrieval capability within the evolving Portal and Federated Search solution set.* In this use, DOCU/MASTER serves as the mainframe search component of an enterprise-wide knowledge management solution that utilizes a federated search methodology to simultaneously search across multiple applications and technology platforms.

Current Development

Two areas of current development will speed the implementation of the DOCU/MASTER mainframe search capability within a customer organization. An intelligent data descriptor module will allow IT resources to interrogate the legacy data sets, specify the data elements to index, and determine the indexing search method, resulting in a quicker deployment schedule. Additionally, an XML-based interface, in conjunction with technology provided by HostBridge Technology, allows for a standardized method to search, retrieve and present search results via server-initiated XML input. Exhibit B outlines the components of this XML-based interface method.



**Document
Systems, Inc.**

About the Company

Document Systems, Inc. (DSI) is a Connecticut corporation organized in 1986 to develop and market full-text retrieval and knowledge management software for IBM mainframe computing environments. DSI was divested by its former parent, AMS Services, Inc., in a management buy-out completed March 31, 1999 and is now run as a privately owned, independent software company. DSI currently services and supports direct sales and foreign resellers from its offices in Walnut Creek, California.

Exhibit A

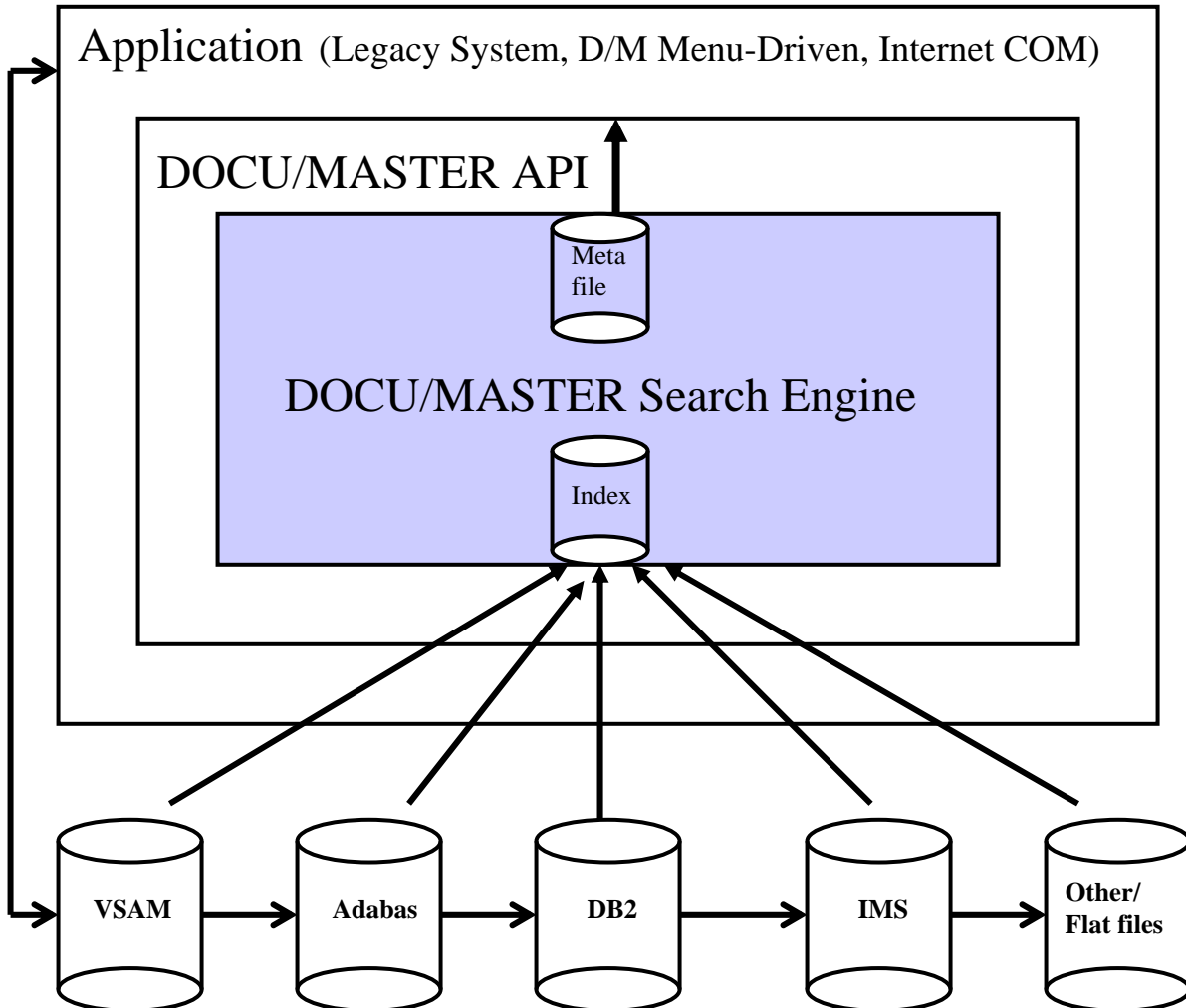


Exhibit B

Generalized DOCU/MASTER Mainframe Search and Access Capability

