J-ISIS Introductory Paper

I. Introduction

This document is an introductory paper on J-ISIS software suite. It describes J-ISIS software suite architecture and its main features. J-ISIS is a new multiplatform Free and Open Source Software (FOSS) that provides the same successful concepts and functionalities as the UNESCO ISIS suite while removing the usage restrictions. It uses a client-server model, supports UNICODE, and incorporates the latest software developments. J-ISIS has its own Persistence Manager based on Oracle Berkeley DB Java Edition for storing MARC records. It also includes a TCP/IP database server, an http web server and a web browser used for displaying records. You can easily create databases and populate them either through data entry or by importing data from external files using standard formats for information interchange such as ISO-2709, MARC21, UNIMARC, MARCXML, MODS and Dublin Core or from Z3950 servers. Records can be viewed, indexed, searched, edited and modified interactively or using batch processing. Selected, marked and search result records can be formatted, sorted and printed in HTML, PDF, etc. You can export data stored in the Persistence Manager into external files using standard formats for information interchange including ISO-2709, MARC21, UNIMARC, MARCXML, MODS and Dublin Core.

II. Background on J-ISIS (or CDS/ISIS) concepts

From the early development of the DOS version, CDS/ISIS was developed for *bibliographic information*, i.e. information about documents such as books, journal articles or conference proceedings. Usually each record in the database contains information about one document.

Many of the features of CDS/ISIS are different from those in relational database management system (RDBMS) which have been designed for general purposes.

Technically speaking, the major differentiating feature of CDS/ISIS when compared with other database management systems is in its use of variable length text fields and records. CDS/ISIS databases are close to **NoSQL** database management systems that differ from the classic model of the relational database management system (RDBMS) in some significant ways. No need to be an expert in relational databases and SQL language. These data stores does not require fixed table schemas, usually avoid join operations, and typically scale horizontally. Academic researchers typically refer to these databases as **structured storage.**

III.   J-ISIS Records/Fields/Occurrences/Subfields

J-ISIS structured storage format can accommodate many types of data, especially bibliographic description of all forms of materials and related data such as authority, holdings, circulation, etc. It describes a generalized structure, a framework, designed specifically for processing data.

Bibliographic data tend to be treated differently from other kinds of data, with less recourse to abbreviation and coding. Moreover, titles of books and other works which are contained in a

bibliographic record may be of any length, from one word to many. As a result, a method has to be found of allowing *variable length fields*, which many database packages do not permit.

Another feature of bibliographic data is the need for *repeatable fields*. One book may have a number of authors. Each author needs to be of equal status. Many bibliographic databases implemented on general database management systems have one field for 'authors'; all authors are entered in one field, but only the first is searchable. In J-ISIS, each attribute which has more than one value is entered in its own *field occurrence*. In other words, each field may be repeated.

Bibliographic data may make extensive use of *subfields*. This facility is available in the ISO 2709 format and J-ISIS has implemented it too. It is a very useful feature when a field is divided into different parts to be treated in different ways. An author's name may appear in an index as *Smith, John*, but sometimes it might be desirable to print it out as *John Smith*, for instance when producing data in different reference styles.

Subfields enable the different parts of the name to be separately manipulated. They are designated by a single character and when entering data in J-ISIS, you prefix this character with a control character referred to as *circumflex* *accent* in both ASCII[2] and Unicode[3] terminology (circumflex, **^**). Thus, you might enter the name Simon Maxwell as **^a**Maxwell**^b**Simon.

The system can treat both parts separately and may manipulate them out in either order, place punctuation between the parts, capitalize one part and not another, and so on. The exchange formats mentioned above include rules for the formulation of fields and subfields. A new feature introduced recently is that subfields can also be repeated.

IV.   Hyperlinks and HTML Data

Since the advent of the Internet, the range of information resources have considerably evolved. Metadata can be entered in fields or subfield as well as hypertext links. J-ISIS allows to enter hypertext links as data that will take users directly to electronic full-text when displayed. Any kind of data formatted in HTML5 can be stored in fields or subfields. It could be links to images, audio content, video material, electronic documents in pdf format or other formats. The J-ISIS embedded browser allows to process HTML5 code.

V. J-ISIS Client/Server Model and Components:

It uses a client-server model, J-ISIS uses its own persistence manager which is based on Berkeley Database Java Edition NoSQL database. Each record is assigned a unique ID called the Master File Number, or MFN. The MFN is used as the key and the structured record as the value in the key/value NoSQL storage.

Starting J-ISIS on a machine launch several components: i) the client application that provide the graphical user interface; ii) the database server that listen for requests from this machine or remote computers; iii) and a Web server that can serves static web pages. When J-ISIS is closed on a computer, the database and Web server threads are also closed.

J-ISIS database server listens on port 1111 by default. Requests are passed as messages to the database server and results are returned. Sockets are used to communicate between the client and the database server through TCP/IP. Thus, a client application can request data from a remote computer running J-ISIS.

The J-ISIS database server may be on the same machine or on a remote machine. The "Open Connection..." dialog allows to define the machine on which the J-ISIS database server is running (and also **where the databases are stored**).

VI. J-ISIS Databases Definition Files

Before a data base can be accessed for processing, it must be made known to J-ISIS by defining certain characteristics of its record structure and contents. The Data base definition editors allow to create and/or modify a *data base definition*.

A J-ISIS data base definition consists of the following components, each stored in a separate file:

Field Definition Table (FDT) : The FDT defines the fields which may be present in the records of the data base and their characteristics.

Data entry worksheet(s) (WKS) : One or more screen layouts used to create and/or update the records of the data base.

Display format(s) (PFT): Display formats define precise formatting requirements for either on-line display of records during browsing, searching or for the generation of printed output products such as catalogues and indexes. J-ISIS provides a powerful and comprehensive formatting language which allows you to display the contents of a record in any desired way. It also provides a specially designed editor to create these PFTs.

Field Select Table(s) (FST) : One FST defines the fields of the data base to be made searchable through the Inverted file. Additional FSTs define the most frequently used sorting requirements for the data base.

J-ISIS provides a set of specially designed editors to manage, create/edit the FDT, the worksheets, the FSTs. J-ISIS also offers a special PFT editor with syntax highlighting, Copy/Paste Undo/Redo, Search/Replace, Syntax checking, Verification of the output produced, PFT generation, etc...

A new database can either be:

i)   created from scratch: A "New Database" Wizard allows to create a J-ISIS database including the database definition files with a sequence of four dialogs in 10-15 minutes.

ii)   imported from existing ISIS-database: The Import Wizard offers a sequence of dialogs that can be filled quite fast.

iii)   or imported from MARC ISO2709, MARCXML, MODS, DC, etc. external files. Here again The Import Wizard offers a sequence of dialogs that can be filled quite fast.

VII. Data Entry/Worksheets

J-ISIS has a useful help facility which can display for any individual field on a data entry worksheet two/three lines of help message at the foot of the field Data Entry window. Pick lists can be associated to fields or subfields. They are defined either in the worksheet or on external files.

Subfields are of great benefit to the organization and manipulation of data. They enable repeated data elements to be associated correctly with other data elements. If forenames and surnames were entered in separate fields, the software would require a more complicated mechanism to associate each forename with the corresponding surname.

VIII. Indexing/Searching/Term suggestions

J-ISIS uses the open source high-performance Apache Lucene full-featured text search engine to enable faster searching of the database. The Lucene index can be considered as an inverted file. The expression refers to the fact that the records are turned inside out to bring different elements from the contents to the fore in a file.

It is possible to index each field in a number of ways using different *indexing techniques*: the complete field, each individual subfield, or each word. There are also two indexing techniques which index text enclosed by <...> or /.../. A stop words file can also be used to remove no-relevant search terms. This flexibility in indexing is not normally found in database management systems.

Indexed Terms

The indexed terms are normalized (i.e. converted to uppercase and stripped from diacritic characters). For example, " Jóború, Magda" is indexed and displayed as "JOBORU, MAGDA". But making a search on "JOBORU, MAGDA" will retrieve the record(s) with "Jóború, Magda"

Searching Methods

J-ISIS provides two searching methods; Guided Search and an Expert Search which allows all Lucene capabilities. These searching methods use the index (or inverted file) and thus you can only retrieve terms that are extracted through the FST entries and indexed.

Guided search is the simpler of the two and only supports Boolean terms. The structure of the search is constrained by the user interface, making it difficult to enter incorrect queries. Expert search permits a wider range of searching functions including proximity searching and searching for repeatable fields.

Fields that are indexed for searching are specified in the field selection table. Fields that are not indexed cannot be searched unless a free text search is used, which scans the entire contents of the records.

Suggestions

The Guided Search module uses **auto complete** user interface features that provide users with suggested queries or results as they type their query in the search box. This is also commonly called autosuggest or incremental search. J-ISIS auto complete implementation is very fast even on large indices in under a few milliseconds so that the user sees results pop up as he types them.

IX.   Print Formatting Language/ New Web Technologies

A further feature of the package is the flexibility of the so-called Print Formatting language (or PFT) for historical reasons. This is achieved by a sophisticated algebra-like *formatting language*. The J-ISIS formatting language follows CDS/ISIS formatting language syntax with some minor exceptions and has been extended with many new features. The main difference is that the display format output produced by interpreting a display format is formatted as HTML5 and displayed in the J-ISIS embedded browser or written on a file as HTML 5 in "*PrintSort*" module while it could also be printed as plain text. It's quite important to have in mind that the output produced for a display format is a text string (or file) formatted as a HTML5 document. This document is then loaded and rendered by the embedded browser in the modules that display records processed by display formats ("*Data Viewer*", "*Search*", "*Free Text Sea*rch", "*Search History*". Thus, the output is a mix of data extracted from a record and HTML5 elements. It means also that all the new Web technologies such as JavaScript, CSS3, Ajax, etc. can be used.

The formatting language serves a number of purposes in addition to its function of providing instructions for screen display and printout.

(i) It is used to specify the rules for the extraction of data from the database records to go into the index.

(ii) It is used for the extraction of data for export to another database or for converting records to a MARC format if data have not been entered precisely according to the rules of that MARC format.

(iii) It is used as the basis of the search language to provide the package with a very powerful free text searching feature, including the searching on values mentioned above.

(iv) It is used in the creation of statements in a validation file for data entry validation. This formatting language gives users of J-ISIS a higher level of control over their data than is found probably in any commercial library automation package.

X. Multilingual UNICODE Databases

J-ISIS is fully UNICODE for text storage, indexing, and database definitions files including pick list files which are read and stored using UTF-8. If you are unable to read some Unicode characters in your browser, it may be because your system is not properly configured

XI.   J-ISIS Digital Library

A special "DOC" field type can be used for storing the extracted plain text version of all electronic documents processed by Tika. The document can be selected in the data entry worksheet near the "DOC" data entry field. Once selected, the system will extract the plain text and store it in the "DOC" field, the text will be indexed word by word excluding stop words if a stop words file exists. The original document will be copied on the J-ISIS database server computer into the /idoc folder and a hyperlink will be created to reference the document on the server machine. If the record is displayed in the Data Viewer and the user

click on the hyperlink, this document will be served by the J-ISIS embedded Web server and the appropriate application will be used to open the document.

XII. Print/Sort

This feature allows to print the output of a given query and/or to print a selected range of records. You may sort the records by virtually any combination of fields and subfields. The field(s) by which the records are sorted may be used as headings in printing. When you select this command J-ISIS displays the **PrintSort Windows** which has two tabs: one to define the print parameters and another one to provide the specific sorting and page layout parameters you require for that particular print run. Printing can be done in HTLM, pdf or plain text with predefined PFTs

XIII. Free text searching

This module allows you to specify search requirements on fields which have not been inverted and/or to specify conditions which you would not otherwise be able to indicate through the type of search expressions described above, such as the comparison of fields or comparisons of the numerical value of fields.

XIV. Global Add/Delete/Replace

The Global Add command allows you to add a field, occurrence or subfield with a specified content to a range of records. The Global Delete command allows to delete a field, occurrence or subfield from a range of records. And the Global Replace command allows to perform global changes in the data base (e.g. replace a string by another). This can be done at the field, occurrence or subfield level.

XV. Groovy

Groovy programming language is available to write Format exits (Call from the PFT to external functions) so that PFT functions can be enriched

XVI. Client Z39.50

It works for MARC21 and UNIMARC, you can access Z3950 servers with User ID and Password. Parallel search on multi servers. Records are converted to UNICODE. You can select the records that you wish to export from the retrieved records. You can export to ISO2709, XML, MarcXML, Text and J-ISIS DB

Web-JISIS is a Rich Internet Application (RIA) whose goal is to develop a web top application that has the responsiveness, and look and feel of J-ISIS.

Web-JISIS

Web-JISIS is a Rich Internet Application (RIA) whose goal is to develop a web top application that has the responsiveness, and look and feel of J-ISIS.

Web-JISIS Web Application uses a Three-Tier Architecture

- **Tier 1**: At the client side, Web 2.0 technologies are responsible for the application's graphical user interface. AJAX is used to communicate between this tier and the application server in tier 2. Requests are sent with the URL using GET or POST data. Request and Result data are formatted using JSON, XML, or plaintext.

- **Tier 2**: the middle tier is also known as the application server, which provides the business processes logic and the data access. The server in the middle tier is a servlet container—a Web server capable of running Java-based Web applications (Tomcat, Jetty, etc). The J-ISIS services (the business rules) are coded in this tier using Java and the client part of jisis-core.jar library. Requests are passed as messages to the database server and results are returned. Sockets are used to communicate between this tier with the client and the database server through TCP/IP.

- **Tier 3**: The J-ISIS database server provides the business data. The J-ISIS database server is listening for request on port 1111, requests and results are passed as messages through TCP/IP.

The **JavaScript jQuery Library** is used to build interactivity to html pages. It automates common tasks and simplify complicated ones due to the library's foundation on standard HTML and CSS technologies. The Twitter Bootstrap Library is used as a front end framework for developing responsive web pages. It has a series of ready-made, great-looking user interface widgets and a comprehensive set of core interaction helpers designed to be implemented in a consistent and developer-friendly way.

The jQuery Select2 plugin is used for AutoComplete Searching and **jqGrid plugin** is used for CRUD (Create, Read, Update and Delete).

Figure 1: Record displayed in J-ISIS Data Viewer component

Data Viewer (DEF_HOME//bibliothek) ×

MFN: 10862 | Format: RAW | Record Count: 10865 | Next MFN to be Assigned: 10866 | Mark Menu | Mark This Record ☐

MFN: 10862
Autor: Bulitta, Erich [Verfasser]
Titel: Erinnerungs- und Gedenkkultur/Schritte zu einer Erinnerungs- und Gedenkkultur : Band II: Von der Erinnerung zu einer Erinnerungs- und Gedenkkultur - Analyse / Erich Bulitta ; Hildegard Bulitta.
Verleger: 1. Auflage -- Berlin : epubli ; 2017
Umfang/Format: 472 Seiten ; 19 cm x 12.5 cm, 547 g.
Series statement: Erinnerungs- und Gedenkkultur 2
ISBN/Einband/Preis: 9783745039863Broschur : EUR 29.50 (DE), EUR 29.50 (AT)
3745039866
EAN: 9783745039863
Subjects:
Index term: Gedenkkultur
erinnerungskultur
gedenkveranstaltungen
gedenken
volksbund
gedenkstunde
gedenkstätten
kulturformen
volkstrauertag
gedenktag
Additional authors:
Author: Bulitta, Hildegard Verfasserin
Corporate: epubli GmbH.
Verfügbare Exemplare
Standort: Geschäftsleitung -- Signatur: GR 01 / 18 -2
Online Inhaltstext
Inhaltstext
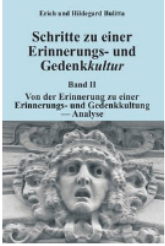
100%

Figure 2 : Worksheet for Data Entry

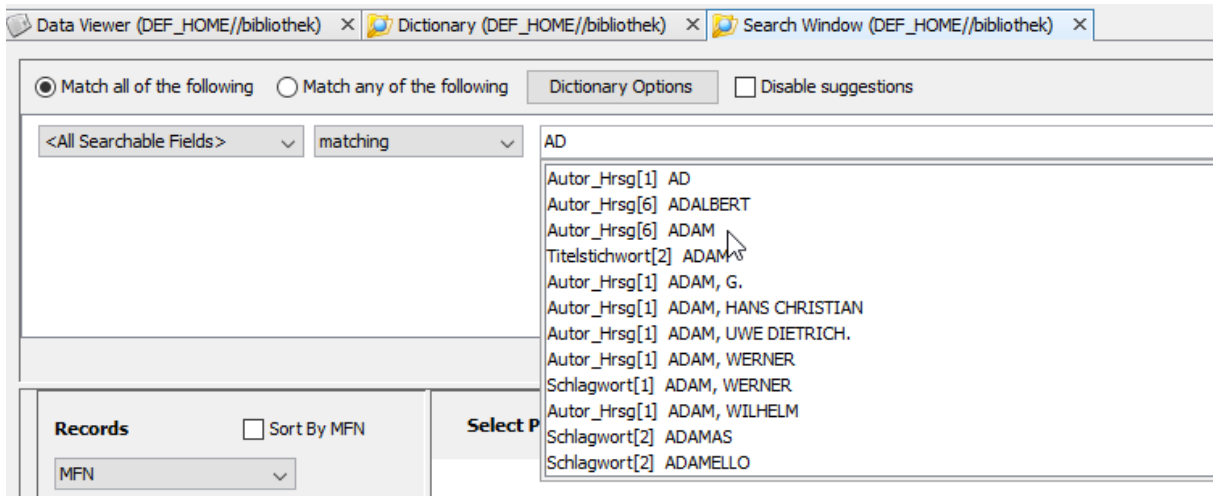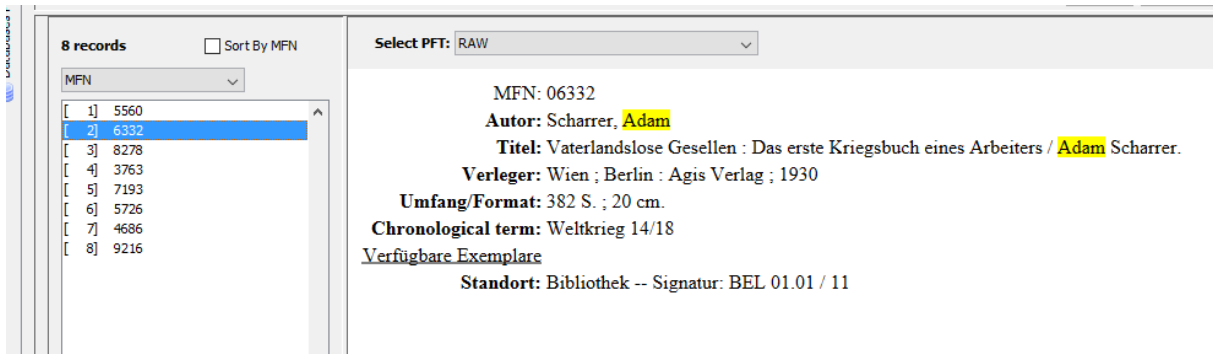Figure 3 : Worksheet - Pick list for language codes and verbose help text
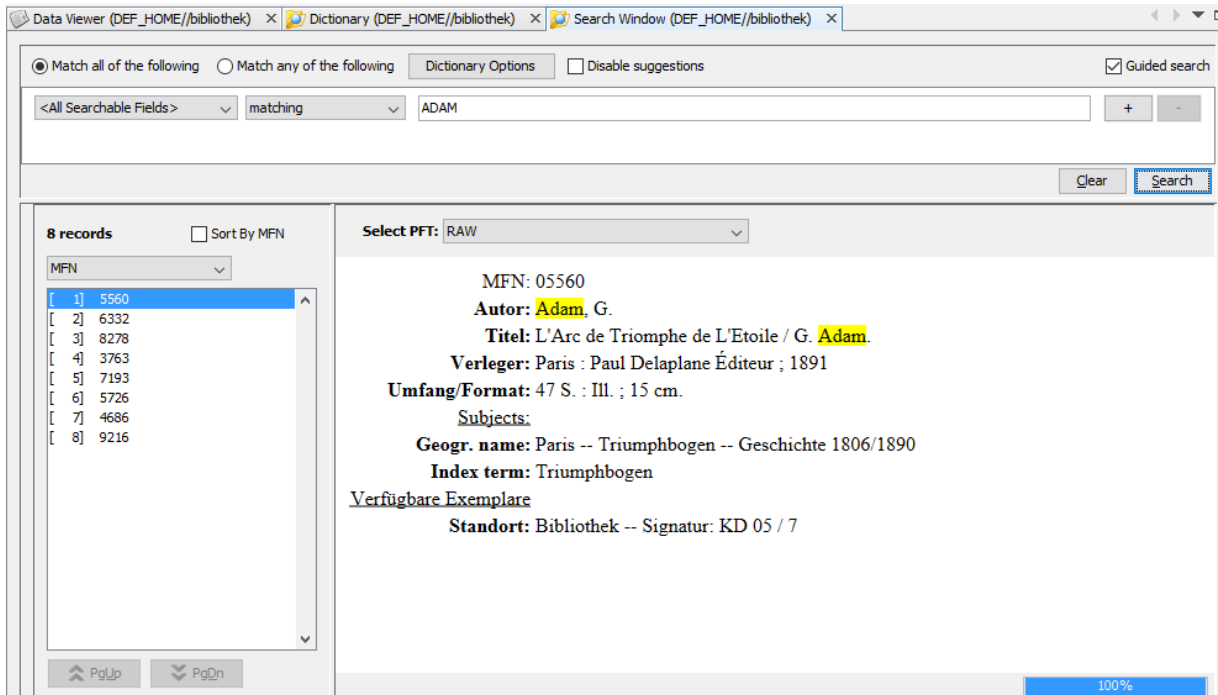
Figure 4 : Searching suggestions
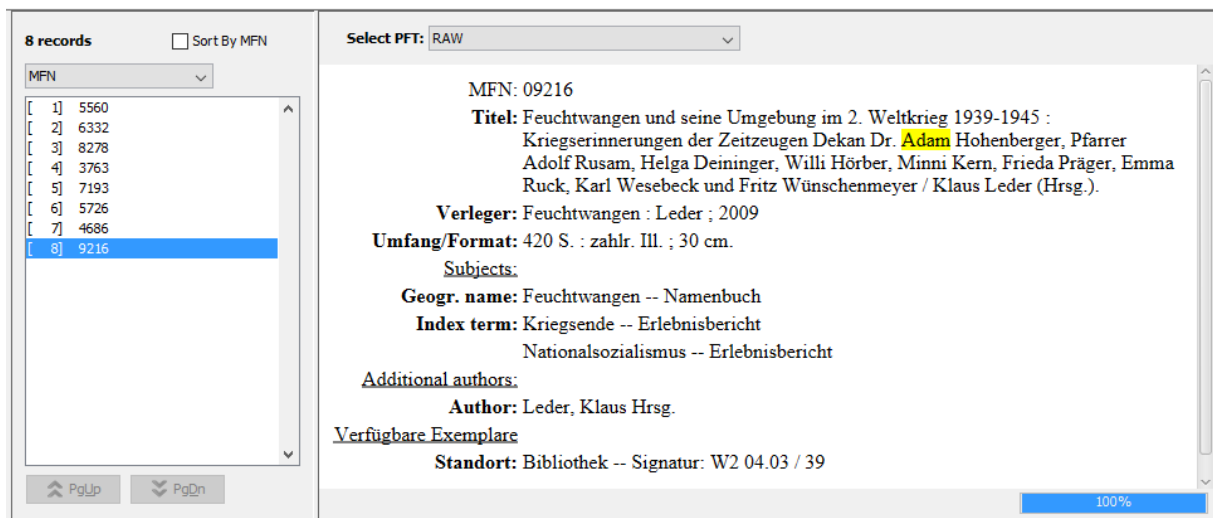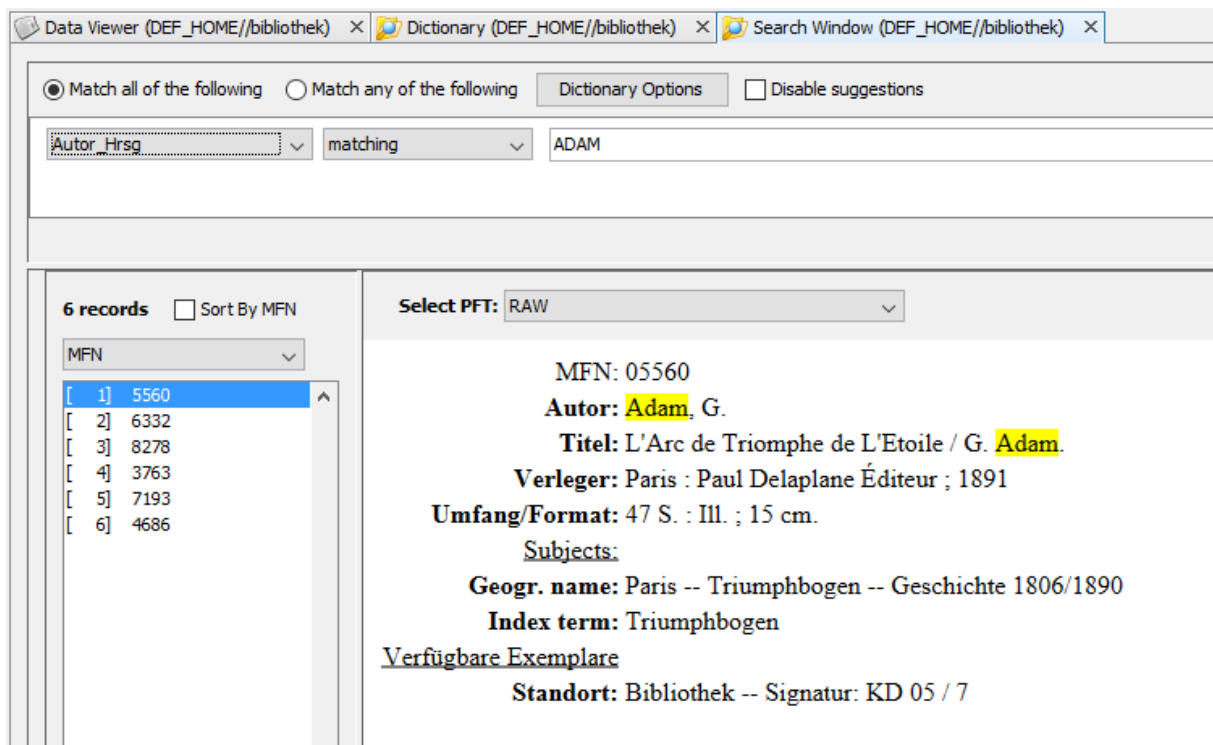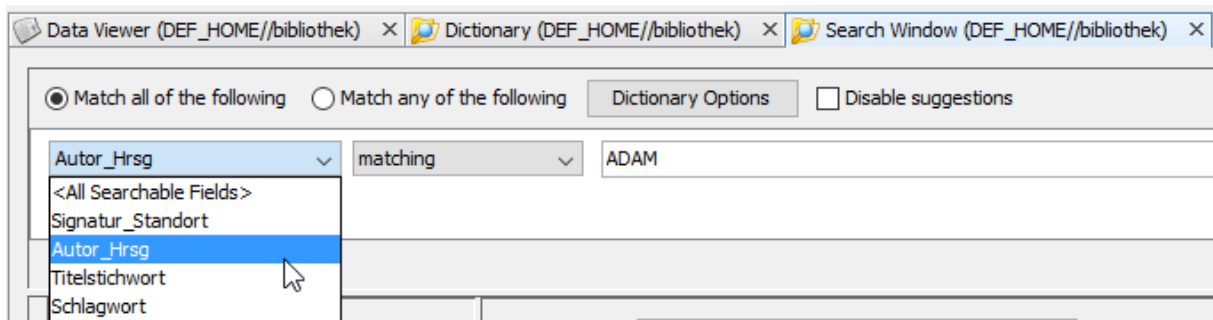

Figure 5 : Searching for term "Adam" in all fields

Figure 6 : Searching for "Adam" in field Auto_Hrsg only

MFN: 1 « ‹ › » Format: RAW

Record Count: 14   Next MFN to be Assigned: 15   Mark Menu Mark This Record ☐

RECORD(1)

Tag	Field/Occurrence
	<< Reference Manual (Version 1.5) UNESCO Information Society Division Sector of Communication and Information © UNESCO, June 2004 (This manual refers to Winisis 1.5 build 3) Revised by: Ben Winnubst, New Zealand (June 2004) Foreword CDS/ISIS is a menu-driven generalized Information Storage and Retrieval system designed specifically for the computerized management of structured non-numerical data bases. One of the major advantages offered by the generalized design of the system is that CDS/ISIS is able to manipulate an unlimited number of data bases each of which may consist of completely different data elements. Although some features of CDS/ISIS require knowledge of and experience with computerized information systems, once an application has been designed the system may be used by persons having had little or no prior computer experience. For advanced users, CDS/ISIS offers a wide range of programming facility allowing the development of specialized applications through the use of its powerful print formats. For real computer programmers, an external programming library, the ISIS_DLL1, provide all necessary tools for developing

100%

10:	<<WINISIS15rev.pdf>>

http://localhost:8585/DL-example/idocs/WINISIS15rev.pdf