ECP 2005 CULT 038097/Bernstein

BERNSTEIN

Deliverable no. 17, ref. D1.3 Workspace v.2

Deliverable number	*D1.3*
Dissemination level	*Public*
Delivery date	*31 July 2008*
Status	*Final*
Author(s)	*The Bernstein Consortium*

e*Content*plus

This project is funded under the *e*Content*plus* programme[1],
a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable.

[1] OJ L 79, 24.3.2005, p. 1.

The Bernstein Project

Deliverable D 1.3
Workspace demonstration version 2

Table of Contents

1 INTRODUCTION 3

2 WORKSPACE STRUCTURE 4

2.1 HEADER 4
2.2 NAVIGATION 4
2.3 INFORMATION FIELD 5
2.4 RESULT LIST OPTIONS 5

3 WORKSPACE FUNCTIONALITY 7

3.1 SIMPLE SEARCH 7
3.2 ADVANCED SEARCH 7
3.3 RESULT OPTIONS 10
3.4 COMPONENT SEARCH 12
3.5 BROWSE MOTIF 14

4 WORKSPACE COMMUNICATION 15

4.1 SRU GATEWAY 15

5 WORKSPACE TECHNOLOGY 17

5.1 SOFTWARE AND VERSIONS 17

LIST OF FIGURES 19

1 Introduction

In the following chapters the workspace demonstration version 2 will be described. This second prototype of the integrated workspace is available online under http://bernstein.iicm.tugraz.at:8080/BernsteinPortal and offers

- simple and advanced search in the watermark databases
- summary and detailed statistics over the results
- search and input within the component model prototype

As basis for the current design and the final functionality of the integrated workspace the third proposal for the design of the watermark catalog workspace (s. http://bernstein.iicm.tugraz.at/bernstein/mockups/index.html) was used. In the next months a new design will be developed by a web designer and as soon as the final design is available it will be implemented in the workspace.
The statistical functionalities of the integrated workspace were implemented as specified in "Deliverable no. 15, ref. D4.4 Statistical functionalities".
The first component model search prototype was enhanced with an input-interface which provides authentication as well as uploading, editing and deleting watermark components and thumbnails and is available online under http://bernstein.iicm.tugraz.at:8080/BernsteinSearchEngine.
The description of workspace communication and technology includes the implementation of the SRU-gateway and a specification of the software and versions which are used for the integrated workspace.

2 Workspace Structure

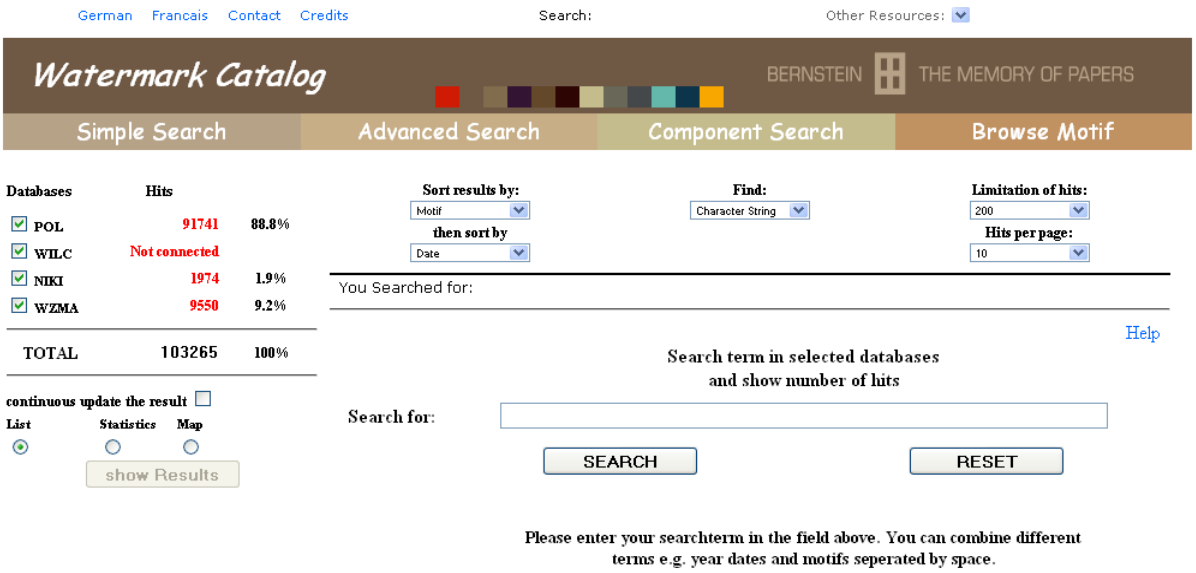

Figure 1: Overview

The start page of the "Watermark Catalog", which is one part of the integrated workspace, consists of the header, the navigation panel, the information field and the result list options. To guarantee an easy navigation through the pages these parts are always present while navigating.

2.1 Header

Figure 2: Workspace Header

In the header section the user can switch between the current available languages of the user interface (English, French and German), search the site or navigate to other resources and parts of the Bernstein-Homepage (e.g.: About Bernstein, Help-Section, Atlas, Bibliography …).

2.2 Navigation

Figure 3: Navigation Panel

Below the header there are the different search possibilities. The user is able to choose between the simple search mode, the advanced search mode, the component search, the browse motif or the user can combine the searches together. When the final design is available, the component search option will be moved to a special section, because it was decided at the general meeting in Stuttgart that this search option is too complicated and too less developed for this version of the portal.

2.3 Information Field

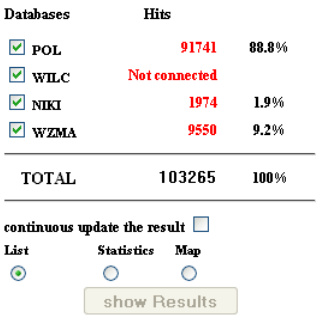

Figure 4: Information Field

The info field section on the left hand side is a main part of the search functionality. Here the user gets a continually live overview of the available databases and the number of hits after a search request. The checkboxes allow the user to select which databases to be searched. When the user gets on the page for the first time, he gets an overview of how many watermarks are searchable in the different databases and the percentage of watermarks per database of the total amount.
Below the user can see the total amount of records or hits in the selected databases. In this part of the workspace it is also possible to select which kind of result view the user wants. It is possible to choose between “List” (normal list view of the search results) or “Statistics” (statistical overview) or “Map” (shows the results on a map – this functionality is not implemented yet). If the “continuous update the result” box is checked, the user always gets directly the selected result view, without using the “show Results” button. Because this option is increasing the workload of the server and therefore is slowing down the search if many users use this option, the continually update box is not checked as default.
All the fields described above are static and always visible on the workspace.

2.4 Result List Options

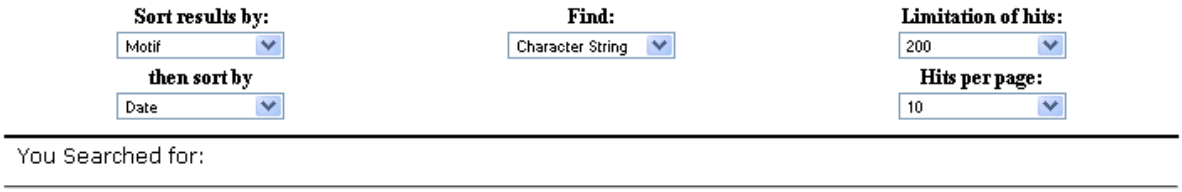

Figure 5: Result List Options

In the result list option fields above the search input field the user can manipulate the result behaviour. On the left it is possible to change the sort order of the results. It is possible to sort by height, date, reference number, motif, place of use and database and it is also possible to enter a second sort criterion, but the second criterion is not mandatory. The default sort order is first by motif and second by date.
In the mid of the result list option fields the user can choose between “character string” search and “exact” search. “Exact” means that the term the user types into the input field is searched as stand alone word; “Character String” means that the input term is searched also as substring of a word. So if the user enters “bull” and sets this option to “exact” only the watermarks with “bull” in the description will be found, but no “bullhead” or any word containing “bull”. With the “character string” search every word containing just “bull” or compounds of this will be found in the databases. “Character String” is the default setting in the workspace.

On the right hand side the user is able to set the “Limitation of hits” and the “Hits per page”. If the number of hits in the result set is above the limitation, the user is not able to click on the “show Results” button; he only gets the number of hits in the info field. The user can specify more search terms or deselect databases to constrain the number of hits or increase the “Limitation of hits”. In the case the number of hits is below the adjusted limitation of hits, the “show Results” button is activated and the user is now able to get detailed results.
The limitation of hits is also a functionality that avoids a too big workload for the workspace if many users are working with it at the same time. The default value is set to 200, but it is also possible to get all results, with the restriction that the system would become very slow if many users are online and are using this simultaneously.
Below this option field is a short overview of the search history the user made and a help button, where the user gets a detailed description of the functionality (will be implemented with explanations and screenshots as soon as the design is ready).

3 Workspace Functionality

In this chapter the search and statistic functionalities and the result options of the workspace will be described.

3.1 Simple Search

When the user selects “Watermark Catalog” within the workspace the simple search appears (see Figure 1). In this area the user gets a simple text input field for his terms and the (already described) option field above to change the result options.
In the input field the user can type in different kind of terms (e.g.: dates, motifs …) and search for it using the search button or reset his inputs with the reset button. The simple search is performed within all fields which are available in each of the databases (i.e. motif, place of use, depository, date, height, distance of chainlines and reference number).

3.2 Advanced Search

The advanced search allows the user to search much more specific in the databases. If the user clicks on “Advanced Search” he has nearly the same view as in simple search with the info field and result list options, but instead of a simple input field the user can now choose between a general or a specific field in the databases he would like to search in. There are several different types of fields and options the user can choose and combine.

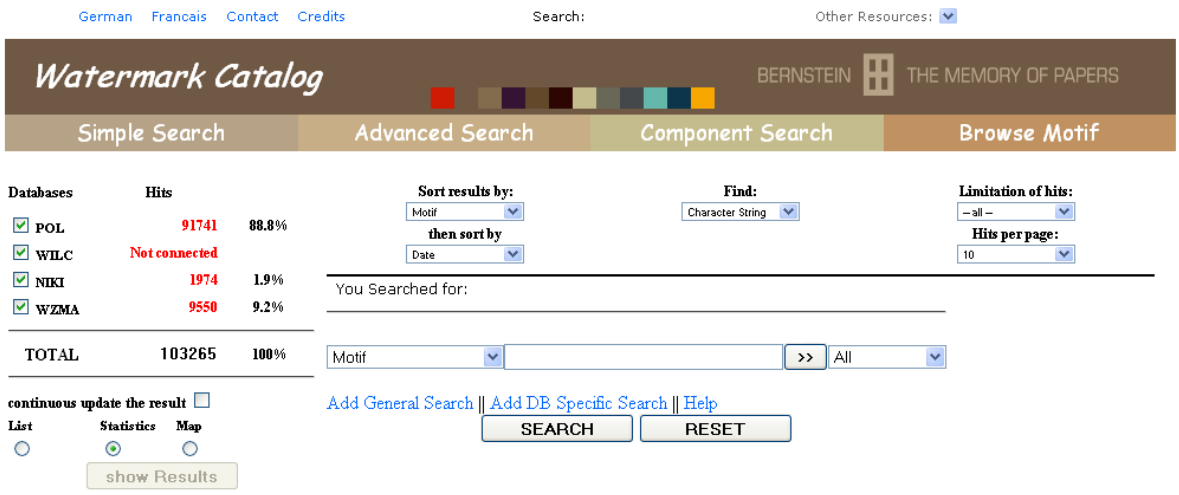

Figure 6: Advanced Search

3.2.1 General Search

With this search the user can search in all databases, because all selectable fields (motif, place of use, depository, date, height, distance of chainlines and reference number) are available for search in each of the databases.

3.2.1.1 Search in “All Fields”

If “All Fields” is selected the functionality is nearly the same as with the simple search. The user can type in terms, dates … and search for them. In the result list options the user has the same possibilities as in simple search, but additionally it is possible to choose “All”, “Any” or “Phrase” on the right side of the input field.

As an example the user choose “All Fields” and types in “bull 1486 57”. With the option “All” the searched watermark must contain all of the entered terms, so there must be “bull” and “1486” and “57” in it. If the option “Any” is selected it searches for all watermarks with at least one of the terms in it. So normally the user gets much more results with the “Any” option. If the “Phrase” option is selected, only watermarks with the exact term ”bull 1486 57” will appear. In this case here we wouldn’t get a result with phrase search, because there is no field in the databases that contains exactly this term.

3.2.1.2 Search in Date Field

If the user chooses “Date” he only has to enter a date and select “YYYY” if he wants a specific date. The user can search for a range of dates if he enters two dates separated by a hyphen and selects “YYYY-YYYY”.

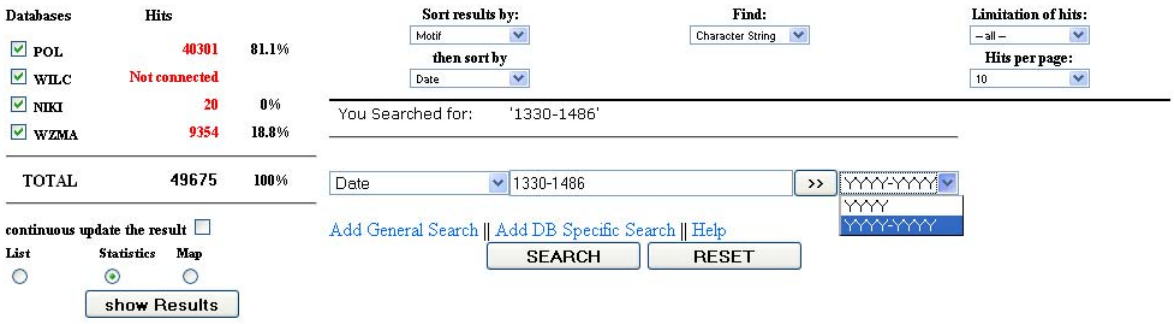

Figure 7: Search in Date Field

3.2.2 Specific Search

If the user wants to search for a field that is not available in each of the databases – but at least in 2 of them – he can choose the “Add specific search” option. From each of the databases where this field is not available the user gets back no results.
In specific search the fields “width”, “twin mark”, “laid line”, “creator”, “Briquet Nr.” and “shelf mark” are currently available.

3.2.3 Search in Text Fields

Search for a string is possible if the user chooses motif, place of use, depository, reference number, twin mark, creator, Briquet Nr. or shelf mark. In all of this cases the user has the same options with “All”, “Any” or “Phrase” with the same effect as described above, with the distinction that the workspace searches only in the selected field. For instance if the user choose “Motif” and enters the same term “bull 1486 57” as above, the workspace searches only in the motifs of the databases. With the “All” or “ Phrase” option we would get no results, because there is no watermark with motif “bull”,“1486” and “57” in it and also no watermark with motif “bull 1486 57”. If we choose “Any”, only one of these terms is required in the motifs to get results because “bull” is (part of) a motif.

3.2.4 Search in Numeric Fields

Numeric search is possible with the options height, distance of chainlines, width or laid line.
At a numeric value field there is the possibility to choose the range of the search (e.g.: 57 ±3), so the options “All”, “Any” or “Phrase will be replaced by a range field.

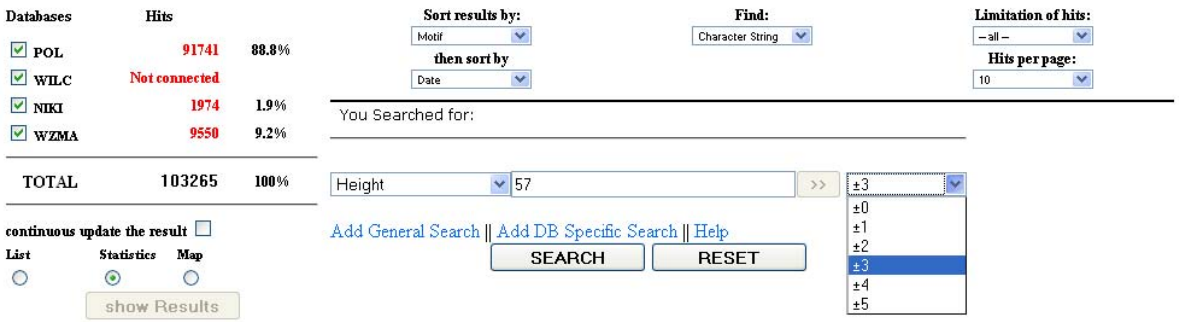

Figure 8: Search in Numeric Fields

3.2.5 Boolean Operators

It is also possible to combine different fields with the Boolean operators “AND”, “OR” or “NOT”. The user can add an additional field by clicking on “Add General Search” to search in all databases or “Add DB Specific Search” to search only in database specific fields. With the “X” button on the right side the added fields could be removed (like a back function).

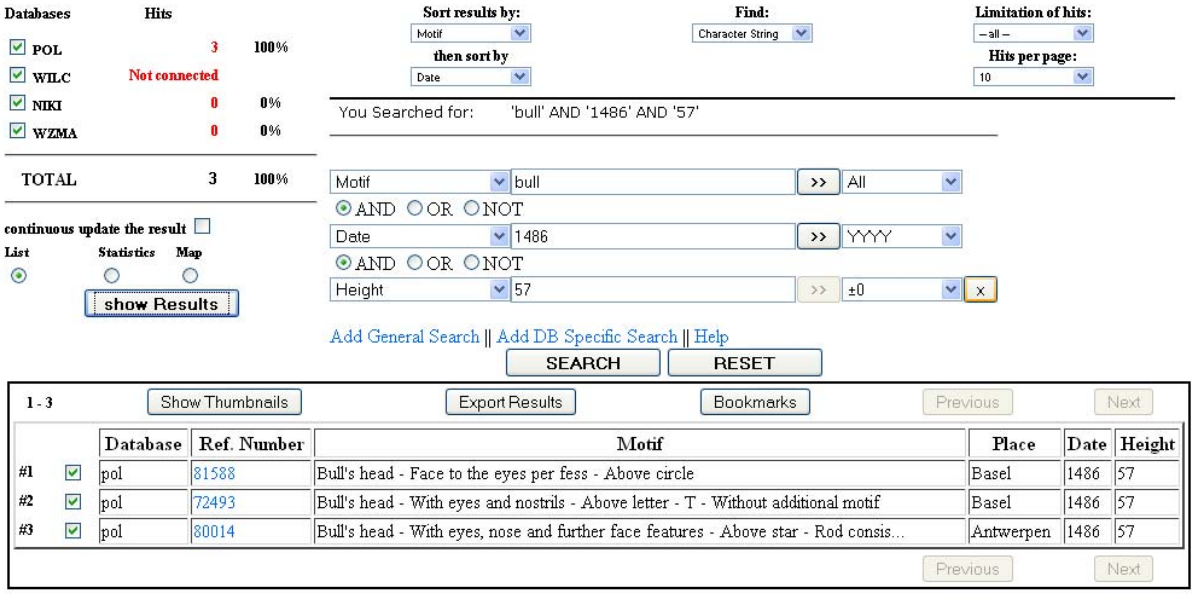

Figure 9: Boolean Operators

Note: If a specific field is combined with “AND” the workspace only gets results from the databases in which both of this fields exist!

3.2.6 Index

In the advanced search section the user also has the option to get an index of motif, place of use, depository, date and reference numbers. The user has to click on the “>>” button to open the index window.

There are several possibilities to search for an index. The user can open the index without entering something in the input field. Then the user gets an overview of all possible values of the selected field with the facility to browse through the index by clicking “next>>” or “<<previous”, or to click on a specific character on top of the view to get the index for this character.

On the right hand side beside the index entries the total number of occurrences in the databases is listed for each value.

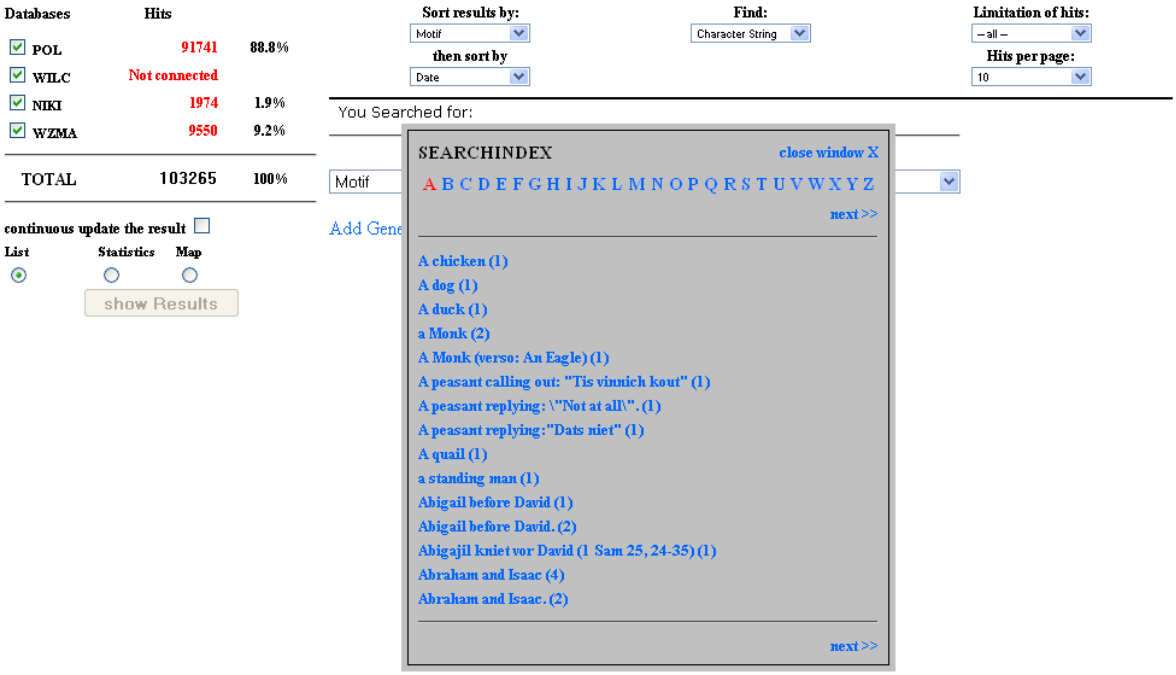

Figure 10: Index

If the user enters a term in the input field, he will only get the values beginning with this term. For instance if the user selects motif and enters “bu” in the input field and clicks on index, he gets all entries beginning with “bu”.

3.3 Result Options

In the info field section the user can change between three representation forms of the result.

3.3.1 Result List

In the list view the user gets an overview of the search results in table form. The list size depends on the selected number of hits per page.
On top of the list there are several options:

- With “Show Thumbnails” the thumbnails for the watermarks will be loaded from the databases. The user gets a small picture preview of the found watermarks.
- With “Export Results” the list will be exported into a file which can be downloaded (not implemented yet).
- With “Bookmarks” the list entries which the user has bookmarked will be shown (not implemented yet).

The table containing the results consists of the database name, the reference number, a short description of the motif, the place of use, the date and the height. The user is able to browse through the results page by page with “next” and “previous”.
In both views (the normal show result view and the view with the additional thumbnails) the user can click on the reference number to get all information of this single watermark.
In this detailed view the user can see all information available for the watermark, these are for instance the reference number, the date, the distance of chainlines, the height, the motif, the database, the depository, the place of use, a reference link and detailed date information.

When the user clicks on the reference link, he quits the integrated workspace and goes directly to the watermark in the specified database in a separate window.

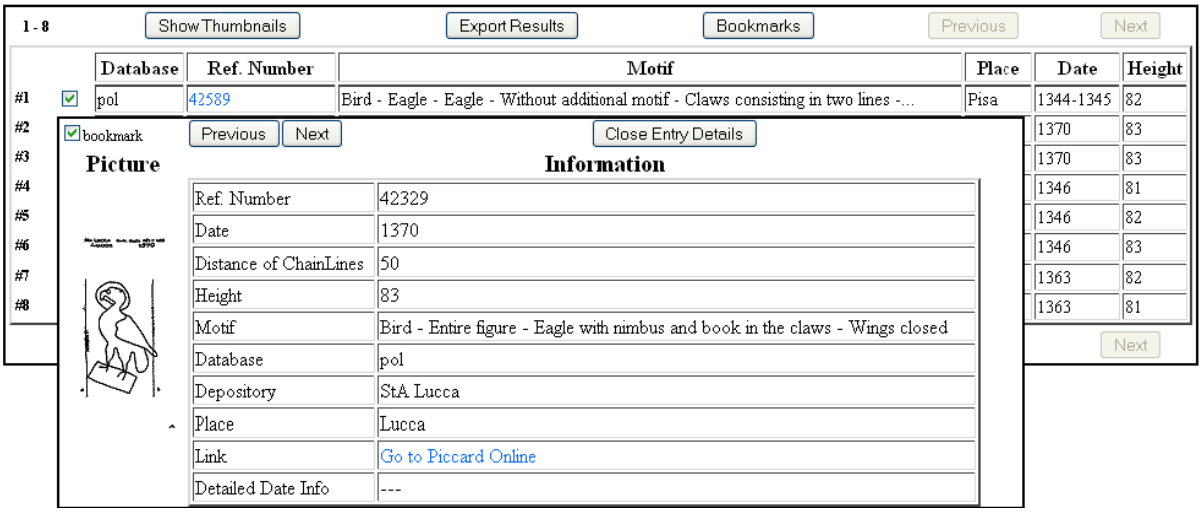

Figure 11: Result List with Detailed View

3.3.2 Statistic View

If the user chooses the statistic view, he first gets a statistical overview of the result set. This view consists of motif, place of use, depository, date, height and distance of chainlines (see Figure 12).
On the top of the page the result list option field is replaced by the combinational matrix for the statistical functions. With a click on “Hide” the result list option field is shown again.

3.3.3 Map View

The map view is not implemented yet, because currently there are no geographical data available in the databases.

Figure 12: Statistic View

3.4 Component Search

The entry page of the component search lists all components of the available motifs and allows the user to get a quick overview of the functionality. Left hand side on the top the databases can be selected and below the search history is shown and the motifs are listed. If the user selects one of the motifs the components are restricted to that motif.

The “Chain Lines Features”, the “Global Shape Attributes” and the “Morphological Components” with the available values are arranged in the middle of the workspace, all available additional elements are listed on the right hand side.

The search results are shown at the bottom and the user can choose between summary (total number of hits), list (reference numbers) and thumbnails (reference numbers and images).

This version of the component search was adapted to the current design of the workspace (s. Figure 13).

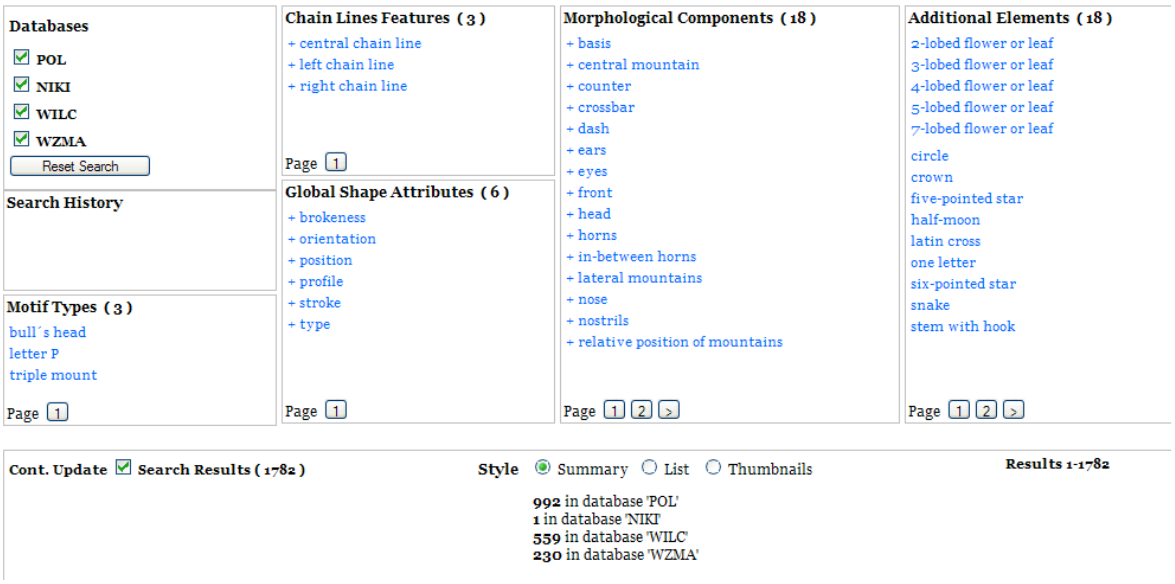

Figure 13: Search Components

Additionally we build an input area where new watermark components and thumbnails can be uploaded to the component database and existing ones can be edited or deleted (s. Figure 14).

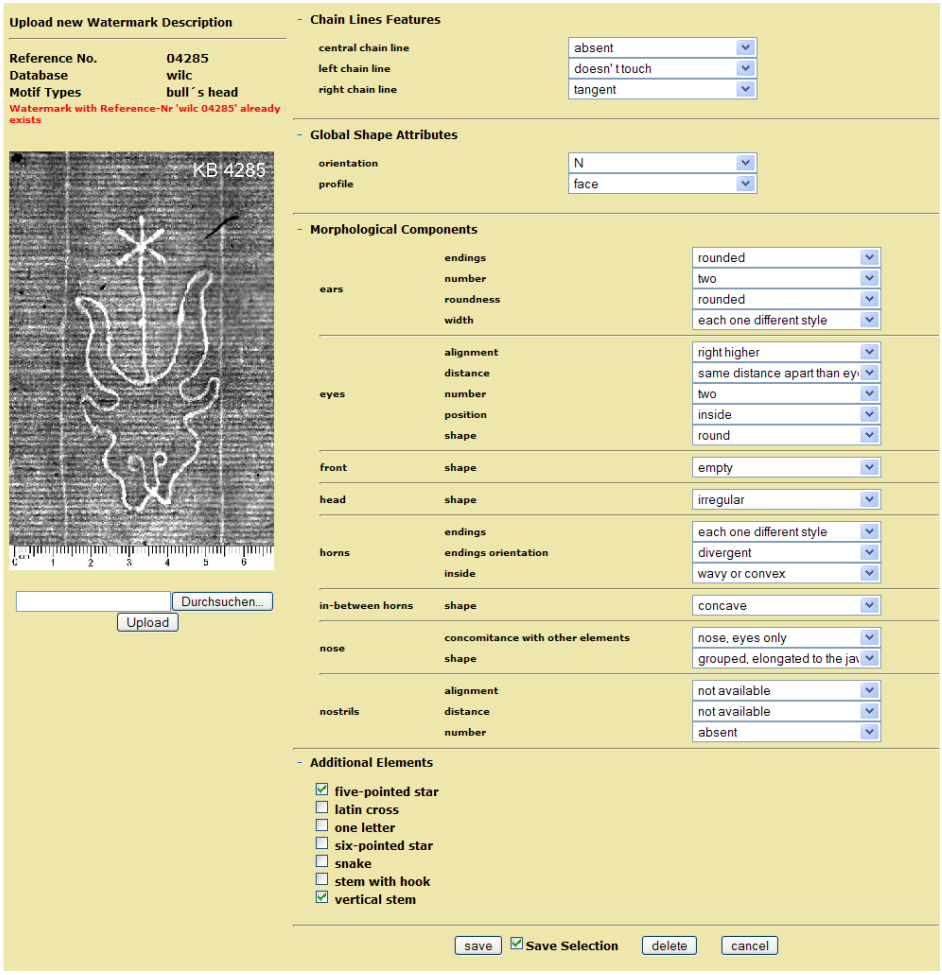

Figure 14: Edit Components

3.5 Browse Motif

The integrated workspace also should contain a motif browser. In this view the user should get an overview of the main watermark motifs and all sub motifs on the next layers.
For the realisation of this feature a commonly agreed “Bernstein hierarchy” of motifs over all existing watermark databases and a mapping of the existing hierarchies to this “Bernstein hierarchy” has to be specified. Because the hierarchy and the mapping are not available until now, this function is not implemented yet.

4 Workspace Communication

The workspace communicates with the databases via SRU (Search/Retrieval via URL) which is a standard XML-focused search protocol for Internet search queries, utilizing CQL (Contextual Query Language), a standard syntax for representing queries.

4.1 SRU Gateway

The basic idea of this stand alone module is to provide a search mechanism which implements the SRU protocol version 1.2 (see http://www.loc.gov/standards/sru).
The module is developed to run on a Tomcat server (version ≥ 5.5) and is implemented as a Java web-service. The supported communication methods are GET/POST and SOAP via HTTP.
Some of the basic characteristics and functionalities:

- A single configuration file based on XML is used to configure all properties of the individual databases:
  - mapping of search field names from local name to common “Bernstein name”
  - mapping of units by using scaling factors
  - parsing of search fields (e.g. which delimiters are used in database to separate motif layers)
  - configuration of connection data (URL to database, port and type of database)
  - possibility to define extra data fields (e.g. to access thumbnails in the remote database)
- The following operations according to the SRU specification are implemented:
  - searchRetrieve
  - Scan (for indices)
  - Explain

The gateway receives an SRU request and checks the elements according to the protocol. Next phase is the query conversion from CQL to SQL by using the configuration file. Then the gateway creates a connection to the database and sends the translated query. After the database responses with the result set the module creates a standardized XML response according to the request (again by using the configuration file) and sends it back.

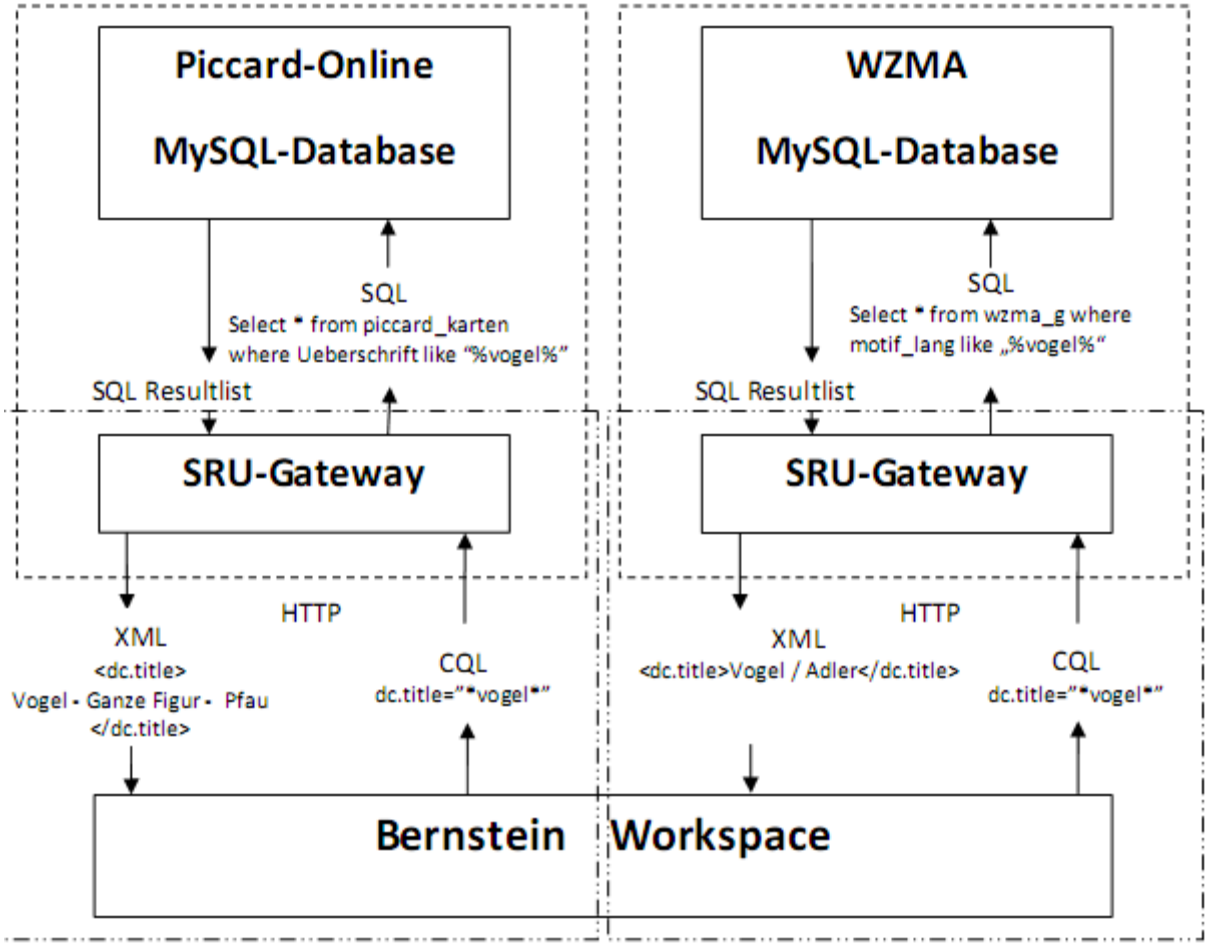

Figure 15: Workspace Workflow

4.1.1 SRU Request

The search query consists of following parts:

- URL of the gateway (hostname, port, path)
- Operation (searchRetrieve, Scan, Explain)
- Version (1.2)
- CQL search query (search terms, boolean operators, wildcards, …)
- Parameters (startRecord, maximumRecords, sortBy, …)
- Extra request data (language)

Example:
http://bernstein.iicm.tugraz.at:8080/GateWayPOL/sru?operation=searchRetrieve&
version=1.2&query=(dc.title="*vogel*" AND dc.title="*krone*")&
startRecord=1&maximumRecords=0&extraRequestData=de
Here we search for a watermark in the POL database with language “German” that contains a title with “vogel” and “krone” in it. Because in this example the “maximumRecords” is set to zero, we only get back the number of hits.

5 Workspace Technology

At the workspace the responses from all SRU gateways are combined and the correct combined result set is generated and displayed.

5.1 Software and Versions

Operating System	Linux	
Programming language	Java	JDK Version 1.6.0 Update 02 JRE Version 1.6.0 Update 02
Framework	Spring MVC	Version 2.1 M3
Front-End	Standard Tag Library (JSTL)	JSTL Version 1.1.2
	JFreeChart	Version 1.0.10
Direct Web Remoting	DWR	Version 2.0
Servlet Container	Apache Tomcat	Version 5.5.20
Database	MySQL	JDBC Version 3.0 including MySQL-Connector Version 5.0.5

5.1.1 Java Platform

“Java” was developed by Sun Microsystems in 1995 and is a cross-platform, object orientated programming language. The Java 2 Standard Edition (J2SE) is a bundle of programs which allow developing and running applications independent from operating system and type of processor.

5.1.2 Spring MVC Framework

The “Spring Framework” is an open source application framework primarily for the Java programming language but is not exclusively linked to.
The framework is extremely comprehensive and modular designed and promotes good programming practice. The layered architecture offers the programmer the possibility to implement and test the components of the application (e.g. flow control, logic, data access and view) isolated from each other and brings a consistent structure to entire application. Spring provides elegant integration of standard and quasi-standard interfaces like web-services or JDBC (Java DataBase Connectivity).
The additional web module called “MVC” (Model-View-Controller) offers an interface for a layered web-application design. This architectural pattern of a software design isolates the logic from the user interface. The advantage of the technique is that the programmer can generate and modify visual presentation of an application (e.g. an HTML-page) independent from underlying rules and logic.

5.1.3 Java Server Pages and Standard Tag Library

In the Bernstein workspace “Java Server Pages” (JSP’s) are used to generate the individual HTML-pages dynamically depending on the response of the web server. This technology allows embedding Java code, actions and directives in the statically content of a website. This

JSP's are compiled into Java servlets which are responsible for the view of the MVC-pattern described above.
The "Java Server Pages Standard Tag Library" (JSTL) is an extension to the JSP specification and provides tags for JSP tasks, loops, conditions and internationalization. It is a way to embed logic in a JSP without including the Java code directly.

5.1.4 JFreeChart

The free Java chart library "JFreeChart" is used to create the complex statistical charts in the workspace. It has a flexible design and is easy to extend to fulfil particular demands. JFreeChart supports various types of output like images, vector graphics or text.

5.1.5 Direct Web Remoting

In the workspace "Direct Web Remoting" (DWR) is used to call Java function (running on the server) directly from the client (browser). DWR is an open source library and has its own engine to generate JavaScript code based on the Java classes which are called. In the workspace this created code is used to update the website dynamically by using the retrieved data. Each call is event driven – this means that dependent on a user action (e.g. by clicking on "show Results") the remote function will be called asynchronously. When data are passed back an extra JavaScript function named "Callback Function" on the client is called which can be used to evaluate received data.

5.1.6 Apache Tomcat

"Apache Tomcat" is the servlet container which is used in the current project. This container provides an http-server environment for Java servlets and JSP's. It is a free, open source implementation and was developed by the Apache Software Foundation (Jakarta Project).

5.1.7 MySQL

"MySQL" is a relational database management system and one of the most popular SQL databases. It provides a very fast, multi-threaded, multi-user, and robust SQL database server.

List of Figures

Figure 1: Overview 4
Figure 2: Workspace Header 4
Figure 3: Navigation Panel 4
Figure 4: Information Field 5
Figure 5: Result List Options 5
Figure 6: Advanced Search 7
Figure 7: Search in Date Field 8
Figure 8: Search in Numeric Fields 9
Figure 9: Boolean Operators 9
Figure 10: Index 10
Figure 11: Result List with Detailed View 11
Figure 12: Statistic View 12
Figure 13: Search Components 13
Figure 14: Edit Components 13
Figure 15: Workspace Workflow 16