

510035

CACAO

User requirements for advanced features

Deliverable number	<i>D7.4</i>
Dissemination level	<i>Public</i>
Delivery date	<i>28 February 2009</i>
Status	<i>Final</i>
Author(s)	<i>Cassia Trojahn, Luigi Siciliano</i>



eContentplus

This project is funded under the *eContentplus* programme¹, a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable.

¹ OJ L 79, 24.3.2005, p. 1.

1	GOAL	3
2	FEATURES FOR THE ADMINISTRATION OPTIONS.....	4
2.1	SYSTEM SETTINGS	4
2.1.1	Basic administration routines	4
2.1.2	Web Service customization.....	5
2.1.3	Boolean customization	5
2.1.4	Boosting	5
2.2	ADMINISTRATION TASKS.....	5
2.2.1	Customization of the interface	5
2.2.2	Translations	6
2.2.3	Fields in the advanced search.....	6
2.2.4	Displayed fields in search results.....	6
2.2.5	"Null results" policy setup.....	6
2.2.6	External resources	6
2.2.7	Search Information customization	6
2.2.8	Search terms highlighting	7
2.2.9	Default sorting criteria	7
2.2.10	Default parameters for search results display	7
2.2.11	Text clouds.....	7
2.2.12	Duplicate detection.....	8
2.2.13	Facets	8
2.2.14	User feedback administration	8
2.2.15	RSS feed.....	8
2.2.16	User profile administration	8
3	FEATURES FOR THE ADVANCED USER INTERFACE	9
3.1	FACETED SEARCH NAVIGATION	9
3.2	TAG CLOUD	10
3.3	USER PROFILE	11
3.4	USER FEEDBACK	12

WP7 Business activities

Work package number :	7	Start date:	0	End date:	24				
Work package title:	Business activities								
Applicants involved:	1	2	3	4	5	6	8		

1 Goal

A deliverable providing a first overview of the user requirements for CACAO was produced at the beginning of the project. As a starting point, previous *D7.1 User Requirements* was mainly focusing on gathering information about the current situation of the partner libraries in order to support decisions about bibliographic standards and technological protocols to be considered in developing the system.

Since basic administration routines have been provided for M12 in D3.3 and D1.4 and an online demo for a basic user interface has been implemented¹, this deliverable provides a quite different approach. According to our *Description of Work* this deliverable will focus both on further *desiderata* which emerged during the period of actual test of the system at libraries premises and on the new functionalities which were not considered before.

Therefore it takes advantage of several weeks of testing of the demo prototype and also of the indications arose from its presentation at the Mid Term review of the project². The Simple User Interface may also take advantage of these suggestions in order to be kept aligned with the guidelines of the Advanced User Interface.

The goal is to list and to describe the features that are required by the different users of the system. On one hand, a user of CACAO will be the librarian who will be in charge to act as administrator of the system and should be enabled to set up and fine tune the needed functionalities. Exploiting this setup of the system, the end user should be able to enjoy the possibilities offered by an Advanced User Interface (due on M18).

This deliverable lists and shortly describes all the *desiderata* raised by the library partners³ and proposes them as suggestions for the further development the Administration Panel and the Advanced User Interface.

Although features for system administration and advanced user interface are strongly interdependent, they are logically separated and consecutive. Therefore, in order to improve its readability, this deliverable is divided in two different sections. The first with the requirements with the features for the Administration Panel, the second with the requirements for the Advanced User interface (AUI).

¹ See <http://demo.celi.it:8095/cacaoUI/>.

² Luxembourg, January 31st 2009. Reviewers Seth Van Hooland and Sofia Stamou.

³ Suggestions for the list of requirements provided in this deliverable have also been kindly provided by Sally Chambers and Romain Wenz of *The European Library (TEL)*. See: www.theeuropeanlibrary.org.

2 Features for the administration options

The Administration Panel of the CACAO platform will be administrated through a web based Graphic User Interface (GUI) and will enable the administrator of the system, to set up and fine-tune the behavior of the system according to the needs of the Library. As a matter of fact, public or university libraries, full text digital libraries and OPACs of more traditional libraries, single library or federation or thematic portals may require from the platform customization of some aspects of the search engine behavior and setting up some default values for the visualization of search results. According to the needs of the adopter, administration tasks may be in charge of only one person or more (e.g. a system administrator for the System settings and a librarian for the other Administration tasks), therefore in this deliverable we will generally speak of admin/librarian.

Basically, the admin/librarian should be able to set: a) parameters for the search engine b) which features will the user be able to choose in the AUI, and c) what the **default** values will be. Not all the possible features in the AUI are supposed to be enabled in the platform (e.g. because of metadata issues or because of the different type of users in a specific Library), and the admin/librarian should be enabled to turn on and off each single feature.

Since administration activity is crucial, contextual help in any language supported by the interface and comprehensive documentation should be provided as well, together with settings for a default, most common installation.

2.1 System settings

Configuration files have already been implemented for M12¹, but the Administration Panel will allow to set up them via a web based GUI.

2.1.1 Basic administration routines

According to *Description of Work*, the basic settings could be set up by the system administrator²:

- Supported languages for cross-lingual searching.
- Supported languages for the interface (with customization of each label).
- Configuration of the MainScheduler and of the HarvestingAgent of the Harvester³.
- Parameters of connection to one or more databases.
- Parameters of the indexing procedures.
- Parameters of the search engine.
- Parameters of the presentation of the results (via XSLT).

¹ See D1.4 Fully integrated CLIR system.

² See *Description of Work*, p. 31.

³ See D3.3 Implementation of access interfaces and basic administration routines, esp. chapters 5 and 6.

- Possibility of connection to different Natural Language Servers (or integration of different libraries).
- Possibility of integration of different multilingual resources.

Other settings are, which can be directly chosen by the librarian in charge of the system:

- Selection of user-based suggestions for translation.
- Selection of user-based suggestions for query expansion.
- Basic dictionary and thesaurus management.
- Monitoring of user activity.

2.1.2 Web Service customization

Use of each Web Service (WS) may be easily turned on and off. Particularly:

- Enable or disable use of Word Sense Disambiguation (WSD).
- Enable or disable use of available expansions (Semantic Vectors, WordNet, etc.). Set limits for the level of the query expansion.
- Set up Named Entity recognition on or off.

2.1.3 Boolean customization

Use of Boolean connecting search terms should be customizable. In particular:

- Choose between AND or OR as **default** Boolean for the query.
- Allow/ do not allow use of "+" for AND and "-" for NOT
- Allow use of NOT Boolean in the advanced search
- Allow use of double quotes for entering also exact match/phrase searches

2.1.4 Boosting

Allow boosting up of some specific metadata rather than others in calculating relevance ranking (e.g. increase the weight for matches on subject rather than on title etc.).

2.2 Administration Tasks

Other important tasks are connected with monitoring and interacting with the end users, the manual improvement of the query, set-up of the default values for the search results and the AUI.

2.2.1 Customization of the interface

In general, each label and caption in the search interface and in the search results should be easily customizable in any language. Different labels may in fact be required according to the different kind of users of each library (Research Library; National Library, University Library etc.). To this purpose an interface translation tool would be useful, showing captions in original language plus translations completed and translations awaiting completion.

2.2.2 Translations

Allow librarian to manually improve translations automatically provided by the system and enable / disable machine learning procedure.

Allow librarian to review and accept translations suggested by the end user via the feedback module¹.

2.2.3 Fields in the advanced search

The administrator should be able to turn on and off the availability to the user of the Advanced User Interface. If the AUI is turned on, the administrator should be able to restrict the values the user should be allowed to use to restrict her/his search:

- allow selection of languages of search results
- allow selection of a from... to ... range (for instance, for date of publication, where using facets, the range of values can be specified). If the index contains normalized values, restricting by date should be allowed. If the values of date of publication are not normalized or reliable, it should be possible to disable this feature.

2.2.4 Displayed fields in search results

Choose what field / metadata can be displayed in search results. For example allow to:

- Display contributor or not
- Display only the first creator or all the creators
- Display the publisher or not

Choose a separator: that is choose "/" or ":" for separating Author and Title in search results, according to different national cataloguing traditions.

2.2.5 "Null results" policy setup

Choose a policy in case of "Null results" to a end user query:

- Just a "No results for the query" message
 - Allow customization of this message
- Ask whether to search with further extension of the terms in the original query
- Seamlessly search with further extension of the terms in the original query (providing a customizable warning message).

2.2.6 External resources

A special section of the administration panel will enable to turn on and off third party plug-ins made available for the system as web services.

2.2.7 Search Information customization

Since the user will be able in the AUI to get more information about the query processing, clicking on a "search info" link, the admin/librarian should be able to customize the quality of this information and to checkbox each of the available information (that is e.g. allow view of recognized Named Entities, etc).

¹ See later, chapter 3.4.

2.2.8 Search terms highlighting

According to the IFLA's *Guidelines For Online Public Access Catalogue (Opac) Displays*, query terms should be marked-up in a different way in search results¹. The XML output used for search results should allow it easily. However a checkbox for enabling/disabling this behavior and a dropdown list of possible markup (different colors, highlight etc.) should be delivered.

2.2.9 Default sorting criteria

In normal circumstances the default sorting criterion should be relevancy. Further criteria should be offered to the user. A list of different criteria should be available in the Administration Panel (sort by relevancy, title, author, date). The admin/librarian should be also be enabled to provide brief documentation about each criterion. This text will be visualized by the end user in the AUI.

The availability of each sorting criterion in the system should be easily turned on or off by the admin/librarian.

2.2.10 Default parameters for search results display

Many aspects of the default search results display should be administered:

- default **amount of displayed search results** (10, 20 etc.). The user will be able to select this value according to her/his preferences. This parameter just defines what the default value will be, i.e. the amount of hits the end user gets in search results at its first access, without selecting any preference.
- checkbox of available **navigation buttons** for search results
 - Previous / Next
 - Page numbers (that is how many page numbers will be displayed)
 - First page / Last page
- **clustering** according to language yes/no
 - yes = results clustered by language (as in current Simple User Interface)
 - no = results sorted together by relevancy disregarding languages
- number of **visualizations style** and drop-down for choosing the templates of each visualization style

2.2.11 Text clouds

The admin should be able to set up text clouds in the AUI:

- enable or disable the feature.
- choose the kind of cloud that should be displayed.
- define default amount of displayed search results (10, 20 etc. - default value).
- enable or disable width of used font related to weight.
- set min. and max number of words in the cloud.
- select source for the text cloud (words may be derived from the index, from the labels of the notations in the Classification System used by the Library or, in a

¹ "Show in context and emphasize the search terms sought in the resultant displays, e.g., by highlighting." See <http://www.ifla.org/VII/s13/guide/opacguide03.pdf>. IFLA is the *International Federation of Library Associations and Institutions*.

federation of Libraries, from native Classification System notations mapped to DDC¹).

2.2.12 Duplicate detection

The admin/librarian should be able to turn on/off the duplicate detection feature in the system, selecting fields on which this module should work in order to detect candidate duplicates. Also, he/she can specify how candidate duplicates should be displayed.

Help about the behavior of the duplicate detection procedure should be customizable by the admin/librarian and displayed to the end user in the AUI.

2.2.13 Facets

Regarding the use of an approach based on faceted search navigation, the Administration Panel must make available to the admin/librarian the options:

- choose which metadata/fields will be available as facet.
- select a behavior in case of null values.
- set up threshold for displaying facet (e.g.: display facet for Date only if 80% of bibliographic records do provide a Date).
- selecting a policy for records without values for selected facet:
 - include records without a value for selected facet (but provide a customizable warning message);
 - exclude records without a value for selected facet (but provide a customizable warning message);
 - create automatically an "Undefined value" facet.

2.2.14 User feedback administration

- Enable / Disable the module for allowing feedback from the users on the quality of query processing.
- Allow visualization, clustering and analysis of feedback provided by users on provided query processing and on search results.

2.2.15 RSS feed

Allow setting up of RSS feeds² for the saved searches of the user.

2.2.16 User profile administration

Enable / disable availability of user profiles and turn on / off options for the user profile settings, according to the level of knowledge of the users. The user profile will be used to store information about the user preferences, such as selected documents, query history, configuration of the facets, and so on.

¹ DDC is the acronym for *Dewey Decimal Classification*. See D3.4 Final Report On Standards Compliance, chapt. 2, for information about the use of the Classification Systems.

² RSS is a family formats used to publish frequently list of updated works in a standardized format. See [http://en.wikipedia.org/wiki/RSS_\(file_format\)](http://en.wikipedia.org/wiki/RSS_(file_format)).

3 Features for the Advanced User Interface

The Advanced User Interface aims to provide end-users with advanced features for searching documents. Such interface must support the following main functionalities. As described above, most of them are configurable by the admin/librarian:

- Select the language to be used in the query.
- Select the libraries to be considered in the query.
- Select the languages for the results (e.g. just English and French)¹.
- Allow advanced query -- enabling query with different input fields (for example, using the facets).
- Submit a query.
- Navigate in the search results (zoom in the query, faceted and tag cloud navigation, select amount of displayed search results, e.g. 10, 20, etc.).
- Select the “n” most relevant documents using filter criteria (for instance, top ten more relevant documents – see section 2.2.9).
- Sort documents, for instance by language, year, or publisher (the facets can also be used).
- Turn on/off Natural Language Processing (NLP) option for Word Sense Disambiguation, query expansion, and Name Entity recognition.
- Change the style of visualization of the results (list or word cloud).
- Login – Logout (user profile).
- Favorite support/kind of media.
- View and combine terms in the query history.
- Print records or set of records.
- Send record or set of records to somebody (for instance, by e-mail).
- Save record or set of records in the personal account.
- Support user feedback (the user should be able to tell whether the translation was good or bad, being able to provide the translation he was expecting).
- Provide a general help session and contextual help in each page.
- Contact module

Regarding the help, a detailed description of the following issues must be provided: (a) how to submit a query using single or compound words; (b) how the relevance is computed; (c) what each NLP option means (i.e., sense disambiguation, expansion, etc); (d) the steps of the translation process; (e) how word clouds are computed; and (f) how to deal with multi-facets.

3.1 Faceted Search Navigation

A state-of-the-art way to provide queries and the manipulation of the results in a coherent manner is based on facets². A facet is a certain classifiable characteristic of the resource -- a way to classify something. Using facets, it is possible to navigate through relatively large

¹ In the Simple User Interface described in D4.1, in order to keep the language selection as simple as possible, only the query language has to be specified, while the results are always available in all languages. In the proposed AUI instead a group of languages for search results can be selected by the user for optimization purposes.

² http://en.wikipedia.org/wiki/Faceted_search

datasets, providing the users different views of the same dataset (i.e., allowing users to “zoom” into the data and explore the documents by filtering available information).

The AUI will present the user with a set of facets, through which the user can filter the search results. A multi-value facet selection will be available, where the user can select a set of values corresponding to each facet. As commented in Section 2.2.13, the admin/librarian will be responsible for specifying the metadata/fields that will be available as facets. It is important to point out that the facets can be applied after the keyword-based search, in order to allow the user to refine the set of search results.

Figure 1 shows an example of facets configuration, where Library, Language, and Subject are selected as facets. When the user selects the values (multi-values selection) of facets, the documents satisfying the values restriction are then presented (for instance, the list of documents is updated). Regarding the values of the *subject* facet, they can be derived from the Classification Explorer module, which is also being developed in the project.



The image shows a vertical stack of three facet selection panels. Each panel has a title and a dropdown arrow on the right. The first panel is titled 'Library' and contains seven items: 'Library 1' (unchecked), 'Library 2' (checked), 'Library 3' (checked), 'Library 4' (checked), 'Library 5' (unchecked), 'Library 6' (unchecked), and 'Library 7' (unchecked). The second panel is titled 'Language' and contains five items: 'English' (checked), 'German' (checked), 'French' (unchecked), 'Italian' (unchecked), and 'Polish' (unchecked). The third panel is titled 'Subject' and contains three items: 'Art' (unchecked), 'History' (unchecked), and 'Sciences' (unchecked).

Figure 1: Multi-value faceted navigation.

3.2 Tag Cloud

Regarding the visualization of the search results, the AUI provides two main visualization styles:

- List-based, where the user sees the list of documents (Figure 2, for a single example); and
- Tag-cloud based (word cloud or entity cloud).

The tag clouds can display all languages together or one specific language. The user will be able to specify the kind of tag cloud he wants, selecting, for instance, the option “all languages” or just the option “English”.



Figure 2: Example of document list.

Tag clouds¹ enable visual browsing, allowing the user to recognize the information he needs by scanning the interface. An advantage of tag clouds lies in their ability to highlight the most important or/and popular words dynamically, which is not the case in conventional navigation menus.

The AUI will provide two kinds of tag clouds: word cloud and entity cloud. The first one will contain the set of more frequent words occurring in the set of results. If the name entity recognition option is on, then an entity cloud can be also presented to the user.

For both kinds of tag clouds, by clicking on one tag, the corresponding list of documents containing such tag is presented.

3.3 User Profile

Different user profiles can be specified, according to the level of knowledge of the users and the interface should provide different services and presentation according to the different user profiles. For example, for expert end users, the options to manipulate the facets can be presented, such as removing facets, changing the order of facets, inverse selection.

Moreover, the user profile will be used to store information about the user preferences, such as selected documents, query history, configuration of the facets, and so on. Also, the AUI is able to store the languages more frequently selected by the user, in order to use this

¹ [Http://en.wikipedia.org/wiki/Tag_cloud](http://en.wikipedia.org/wiki/Tag_cloud)

information to configure, for instance, the default language of the AUI or the faceted navigation.

3.4 User Feedback

Ways of getting end-users feedback about query processing and search results will be considered. User feedback module should include:

- Evaluation of the translation;
- Possibility to suggest a better translation.