Consiglio Nazionale delle Ricerche

Nota Interna

ETRDL Common User Interface Handout

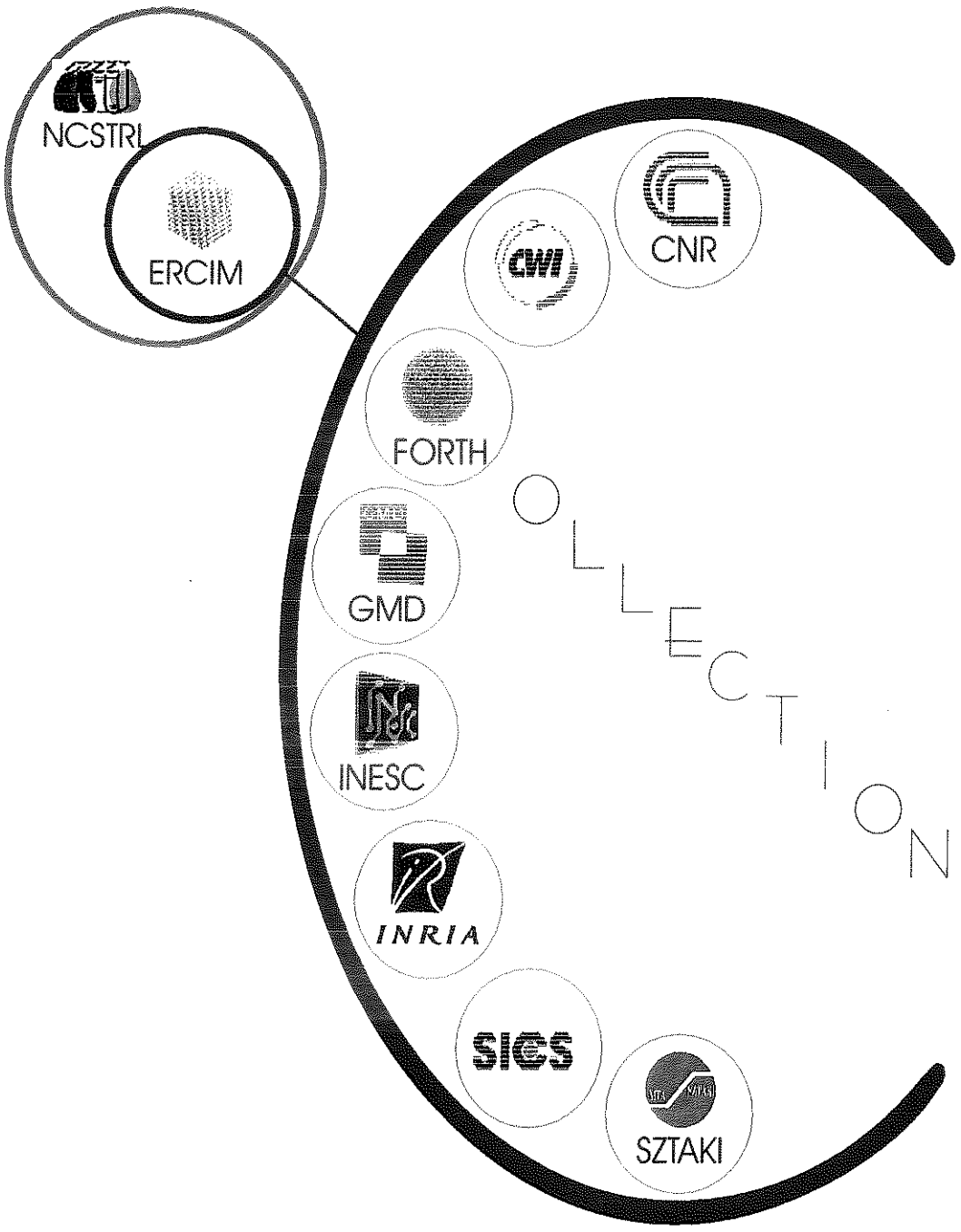
Stefania Biagioni, Carol Peters

B4-41
dic-1998

I.E.I.
ISTITUTO DI
ELABORAZIONE DELLA
INFORMAZIONE

Pisa

ETRDL ERCIM Technical Reference Digital Library



Edited by Stefania Biagioni and Carol Peters

Graphics by Massimo Martinelli

The ERCIM Technical Reference Digital Library

The European Research Consortium for Informatics and Mathematics (ERCIM) has decided to set up a digital collection of the technical literature produced by its 14 member institutions. The aim of the ERCIM Technical Reference Digital Library (ETRDL) is to assist ERCIM scientists to make their research results immediately available world-wide, and to provide them with appropriate on-line facilities to access the technical documentation of others working in the same field. Public access to this reference service is provided through Internet.

The ETRDL collection consists of all kinds of grey literature (technical reports, proceedings of conferences or workshops, theses, project deliverables, etc.) and is managed by a set of interoperating servers. Pilot server sites have already been set up at more than half of the 14 ERCIM national labs. Servers are expected to be installed at the other centres soon.

The system has been based on the Dienst system developed by a US consortium led by Cornell University and adopted by NCSTRL (Networked Computer Science Technical Reference Library). However, additional functionalities are being implemented in the ETRDL common user interface in order to meet the particular needs of the European IT scientific community.

The intention of this handout is to give the reader an idea of the basic system design and the Common User Interfaces (both for Search and Submission), and to describe the extensions that are now being implemented.

The system will be publicly available from September 1998 on the DELOS Web site:

http://www.iei.pi.cnr.it/DELOS/

The ETRDL System

The ETRDL system has two objectives: (i) to provide a digital library service for grey literature in computer science and applied mathematics; (ii) to provide a test-bed for ERCIM scientists working on digital library related research issues.

As can be seen from the cover page, the ERCIM collection is part of the NCSTRL collection. A user accessing the NCSTRL service can view and query any of the collections of the ERCIM institutions using the standard NCSTRL search functions. However, the ETRDL prototype represents an extension of the standard Dienst system as a set of additional browse and search functions and an extended metadata set has been added.

The building of the system is being carried out in two main stages.

The first stage has now been completed and has involved:

- the definition and setting up of the naming and regional scheme
- the development of a Common User Interface - CUI (for submission and search of documents)
- the adoption of a common classification scheme
- the definition and implementation of an Administrator Interface

The second stage is involving the study and development of a number of more advanced modules:

- a Z39.50 access to the Dienst system has been developed at FORTH and will be integrated in the ETRDL system
- a new advanced version of the User interface is being developed at SZTAKI.

Another important development is the implementation of a Multilingual Interface:

- each institution should install the CUI in its local language as well as English
- a multilingual classification schema has been studied by INESC and is now running for English/Portuguese computer science keywords; it is our intention to extend this schema with other languages and integrate it into the common ETRDL system.
- other methods for cross-language querying which enable the user of the DL service to retrieve texts composed or indexed in one language via a query formulated in another will be investigated.

The Metadata Set

In the ETRDL collection each document has a common metadata description associated. This description complies with the Dublin Core metadescription standard and represents an extension of the basic Dienst metadata set.

This description contains the following elements:
Title, **Author(s)** (the person(s) or organisation(s) primarily responsible for the intellectual content), **Subject** (a list of free keywords or selected document descriptors from the ACM and AMS categories), **Abstract** (an English abstract, and, optionally, a local language abstract), **Publisher** (the ERCIM institution), **Date** (the date of the intellectual content), **Type** (default value is Technical Report), **Format** (postcript, pdf, html, text, gif and tif), **Language** and **Identifier.**
Each field must be filled in when a new document is submitted to the collection.

Subject indexing/searching

The user can employ the ACM Computing Classification and/or the AMS Mathematics Subject Classification, and/or free keywords to represent subject terms for retrieval when querying the system or for document classification during the submission procedure. The ACM and AMS schemes are accessible on-line and can be browsed during both retrieval and submission; codes with associated descriptors can be selected and inserted in the appropriate fields. Authors must enter codes/descriptors from at least one classification; they can also use all three fields. Searches are performed on all three fields by default.

The Common User Interface

In the following pages of this handout, we will show the main features of the Common User and Administrator Interfaces. These interfaces have been designed in order to be highly user friendly: no special knowledge is needed by the user querying the system, the author submitting a new document, or the administrator/librarian responsible for verifying the submission and inserting it into the local collection. On-line helps can be called-up, in English or in the local language, when extra assistance is needed.

The Home Page

The Home Page provides the user with two main options: **search/browse** any collection or **submit/withdraw** a document from the local collection.

The search and browse functions can be activated over the entire NCSTRL collection, the ERCIM collection, or the collection of the local institution.

The Home Page is localised by institution. The logo of the institution appears in the top left hand corner and, at the bottom of the page, a button allows the user to switch between interfaces in English or the local language. In the figure on the next page, we see the Home Page of the Italian National Research Council (CNR) and the language choice is obviously Italian.

Browsing the ETRDL Collection

The user must select the collection that he/she wishes to browse. The collections can be viewed ordered by year or alphabetically by author or keywords. A given document is selected and displayed by clicking on it with the mouse. Documents in .ps format can currently be downloaded and then printed. Other formats will be included soon.

Searching the ETRDL Collection

The user can choose to search either selected collections or all collections.
Two kinds of search are available: **fielded search** or **simple search**.

Fielded Search
The user enters his/her query terms in the field(s) on which the search is to be performed. The value for document type and language are selected from a menu; the default value is All. By default, searches are performed in 'or'; the user has the option of activating the 'and' operator between the search fields.
Three different kinds of subject can be entered in the subject field: ACM Computing Classification descriptors, AMS Mathematics Subject Classification descriptors, or free keywords.

Simple Search
Query terms entered in the simple search field are searched throughout all the fields of the documents in the collection(s) selected.

In the next two pages we see screen dumps of the Browse and Search interfaces.

Results of a Search

The results of a search are first displayed in summary - the number of documents found is displayed for each authority.

The documents found are listed, authority for authority; the title and authors of each document is displayed. By clicking on a given document the user can view it. Documents may consist of just the title, authors, abstract and keywords; in this case the entire document is displayed immediately on the screen. If the whole text has been inserted, the user also has the choice of selecting to display an overview of the document (the whole document in thumbnail format, or page by page). He can also download and/or print out the whole document.

In the next three pages we see:

1. The first response from the system for a search requesting all documents in the ETRDL test collection containing in the abstract field any of the terms: *Information Search Retrieval* or in the subject field the code/descriptor: *H-3-3 Information Search and Retrieval*

2. The bibliographic record and abstract for one of the documents retrieved by the previous query and the instructions for how to view or download the entire document

3. Thumbnail view of the same document.

Inserting a New Document in the System

In order to insert a new document in the collection, the document submission form must be completed. All fields on this form are obligatory. For each field, on-line helps are available to assist the compiler if necessary. Subjects must be assigned to each document; these should be selected preferably either from the ACM or the AMS classification schemes, which are available on-line. However, the user also has the choice of adding to these or supplementing them by free keywords. If the compiler needs assistance in assigning the correct subjecs, he can contact his local librarian using the link at the bottom of the page.

An abstract in English is obligatory for all documents; for documents in a language other than English, an abstract in the local language should also be included.

The compiler indicates the name of the file containing the new document to be inserted in the system by using the Browse button to scan his/her system files and mark the appropriate one.

When the form has been completed, the compiler must click the Submit button. The system will display the completed form and request the compiler to confirm that it is correct. If corrections are to be made, the compiler can return to the original form, using the Go Back button. The corrected form is then resubmitted and sent to the system administrator.

In the following three pages, we see the submission form as it appears to the compiler who is preparing the bibliographic record for insertion in the collection and - on the third page - the completed form that has been returned by the system for verification and confirmation.

Administration Interface

Each institution will develop its own Administration Interface. This form is used by the Administrator (often the Librarian) for insertion of new or deletion of outdated documents from the collection(s).

The Administrator is generally responsible for assigning the Document Identification No. (DocId.) and verifying the formal correctness of submitted bibliographic records. If it is correct it is submitted directly to the system, otherwise the Administrator will contact the compiler asking for the necessary corrections to be made.

The screen dumps on the next two pages shows the CNR interface.

Multilingual Interface

The field of Multilingual Information Access is an important emerging area for research and development in the international scientific community and for this reason it was considered important that the ERCIM DL could provide a testbed for R&D activities. The ERCIM scientific community currently consists of 14 member Institutions, with 13 different major European languages.

The first activities of the ETRDL in this area are aimed at (i) implementing an interface capable of handling multiple languages and (ii) providing very basic functionalities for cross-language querying.

Multilingual Access and Browsing

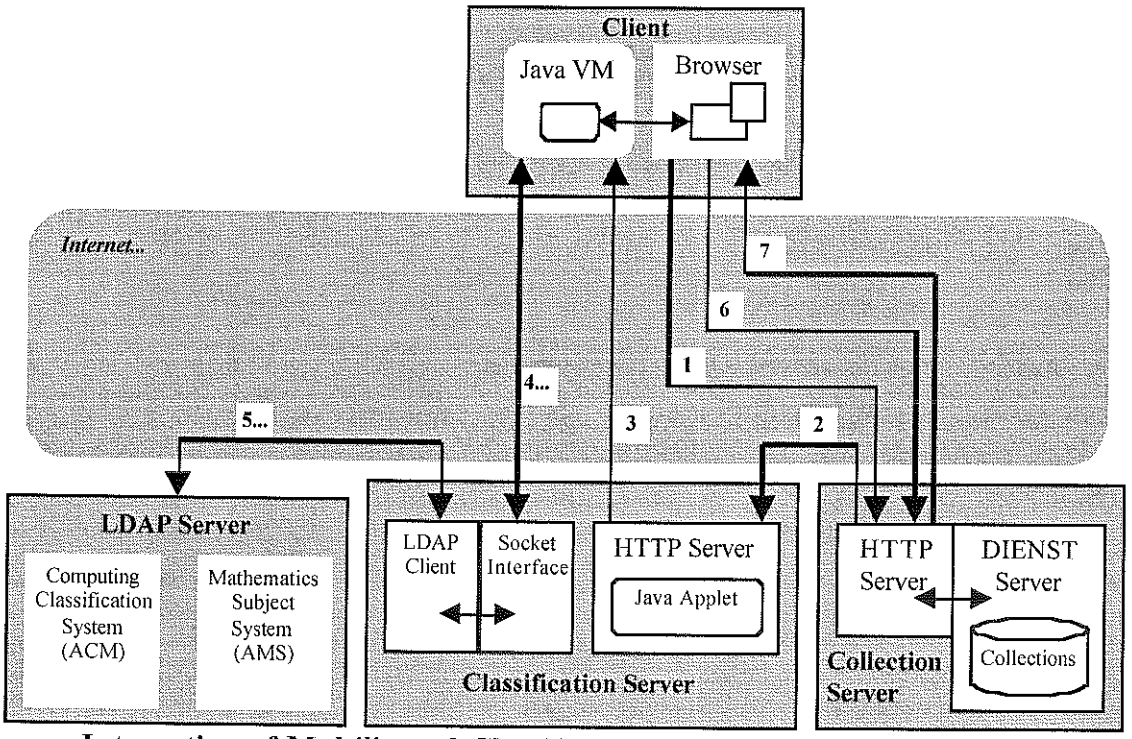
Each national site is responsible for localisation, i.e. implementation of local site user interfaces (also) in the national language as well as English (the screen dump on the previous page shows the Italian version of the Home Page). One of the tasks of the group will be to investigate problems involved in rendering the Dublin Core element set multilingual. Documents are tagged for language and character code switching mechanisms are now being provided for the local display and printing of non-Latin-1 languages (Greek in our collection). However, it is agreed that UNICODE must be adopted eventually in order to fully internationalise the system.

Cross-language Querying

In the short term a simple form of cross-language querying will use controlled (ACM/AMS) terms. All documents in the ETRDL, in whatever language, classified using this scheme, can thus be searched. Authors are also requested to include an abstract in English, which makes English free term searching over documents in any language also possible.
We are now working on integrating in the system a multilingual classification scheme developed by INESC. The languages covered are currently English and Portuguese but other languages should be added in the future.

The INESC LDAP Server

INESC has developed an LDAP service with a multilingual repository for the ACM and AMS classification systems, which is integrated in the ETRDL system. This multilingual service will make cross-language querying in local languages possible.



Integration of Multilingual Classification Systems with DIENST

Both users submitting new documents to the collection and those accessing the system to search/browse through the collection can use the Classification Server.

Interaction begins with a contact to the Collection Server, by HTTP, to select the operation desired (1). If the Classification Server is requested, the request is transferred to it (2), and it will reply by sending an applet (3) to the user.

With this applet the user can browse the classification systems available from the Classification Server (4). These classification systems are stored in a X.500 directory, accessible by LDAP (5). The LDAP server used in this system is provided by the University of Michigan.

The Classification Server currently contains the ACM Computing Classification System and the AMS Mathematics Subject System in English and translations in Portuguese. The directory has been designed to host other structured classification systems, and also their translations in multiple languages. The Java applet is completely independent of the contents of the directory, being configured according the information received from it.

When interacting with the Classification Server, the user can select the desired terms, in one or more of the languages, and return to the Collection Server with the terms selected (6) to be used in the task in course (7). The selected terms can be used in the Collection Server to classify a new document, or to search in the indexes. The actual Classification Server can maintain several indexes, depending on the metadata structure of the collection. In the current system we support one index for each classification schema, but the terms included in these indexes can also be used as generic keywords and indexed in the keyword index (in this sense, a user can perform a free searching in this index using terms from the classification systems).

A Z39.50 Gateway

Online libraries have established Z39.50 as a protocol for accessing their data. The Z39.50 protocol enables cooperation of libraries, so that users can issue one query and get answers from many libraries. Digital libraries are not just online cataloguing systems, but also hold and provide the material itself online, on a digital form, and provide advanced ways of searching and material retrieval and presentation.

We combine the two worlds, providing Z39.50 access to a digital library that is using Dienst, by adapting a Z39.50 server to use the digital library metadata and to provide links to all formats of the data. Our approach requires no changes to the Dienst server and to the part of the Z39.50 server that implements the Z39.50 protocol; it is implemented by extending the Z39.50 server. An experimental server that implements part of the functionality described in our mapping, including concurrent searching to multiple Z39.50 servers, is operational.

A set of tools that simplify the configuration of a Z39.50 server attached to a Dienst digital library has also been developed.

The ERCIM Technical Reference Digital Library (ETRDL) is one of the activities of the DELOS Working Group, funded by the ESPRIT Long Term Research programme (LTR No. 21057)

Project Partners

CNR - Consiglio Nazionale delle Ricerche - Italy
CWI - Centruum voor Wiskunde en Informatica - The Netherlands
FORTH - Foundation of Research and Technology - Hellas - Greece
GMD - German National Research Center for Information Technology
INESC - Instituto de Engenharia de Sistemas e Computadores - Portugal
INRIA - Institut National de Recherche en Informatique et en Automatique - France
MTA SZTAKI - Magyar Tudományos Akadémia - Számítástechnikai és Automatizálási Kutató Intézete - Hungary
SICS - Swedish Institute of Computer Science - Sweden

Main Contacts

Stefania Biagioni, IEI-CNR (Coordinator)
Josè Luis Borbinha, INESC
Reginald Ferber, GMD-IPSI
Preben Hansen. SICS
Sarantos Kapidakis, ICS-FORTH (Hariklia Tsalapata from Sept. 1998)
Laszlo Kovacs, SZTAKI
Frank Roos, CWI
Anne-Marie Vercoustre, INRIA

For further information on ETRDL, contact Stefania Biagioni (biagioni@iei.pi.cnr.it)

Acknowledgments

ETRDL is a collaborative effort between a number of ERCIM Institutes. Many ERCIM scientists and technicians have been involved in the setting up of the experimental technical reference service. In particular, we should mention: Barbara Lutes (GMD-IPSI); Jacob Mauroidis, Giorgos Sapunjis, Panagiotis Alexakos, Gregory Karvounarakis (ICS-FORTH); Maria Bruna Baldacci, Carlo Carlesi, Donatella Castelli, Carol Peters (IEI-CNR); Mario Loffredo and Giuseppe Romano (CNUCE-CNR); Paula Viana, Nuno Freire, João Fernandes and João Ferreira (INESC).