

Research Network and Database System – a Netbased Form of Working Platform in the CRS 600

Thomas Burch and Ansgar Schmitz
University of Trier / Germany
burch@uni-trier.de / schmitza@uni-trier.de

Abstract

In this paper, we present the design and implementation of a completely Tcl/Tk-based client/server application, which is intended to support a large community of researchers within a collaborative research centre (CRS). The main purpose of this distributed system is to enable the participating users to cooperate via the internet in creating and annotating their data. On the client's side, it provides a graphical user interface, which comprises components for user accounting, data import, data classification, workflow management, group management, data retrieval and finally online publication. On the server side, it is connected to a relational database management system. The communication between client and server is done with the help of socket commands within Tcl/Tk. The advantage of this bidirectional and synchronized connection is that the user can be immediately informed about a change of data in his area of interest. This avoids data redundancies and enables a large group of users to effectively establish a commonly used pool of information. Besides Tcl/Tk, we use several additional libraries such as bwidgets, CoST, mysqltcl and tcllib.

1 Initial Situation

The collaborative research centre 600 “Strangers and Poor People. Changing Patterns of Inclusion and Exclusion from Classical Antiquity to the Present Day” is a central institution at the University of Trier and is financed by the DFG, the federal state of Rhineland-Palatinate as well as the University itself. This research association with History in a leading role was launched on 1st January 2002 and is momentarily in its second phase (2005-2008). Besides History, there are many other subjects participating in the CRS 600 such as German studies, Art History, Media Studies, Politics, Law or Theology.¹

¹ Cf. www.sfb600.uni-trier.de

The CRS 600 focuses on the attitudes and behaviour towards strangers and paupers in various types of society from classical antiquity up to the 20th century. Analysing the changing patterns of inclusion and exclusion of strangers and poor people, we would like to form a basis for a description of European and Mediterranean societies in terms of social and cultural history, which focuses mainly on problems connected to the organisation and limitation of social solidarity. By examining these phenomena from a wide historical and geographical perspective, it will be possible to identify and explain common characteristics and differences as well as continuities and fundamental changes in the ways European and Mediterranean societies treated and lived with strangers and paupers.

Our programme of research is subdivided into three areas: A) strangeness, B) poverty and poor relief and C) collective representations and historical semantic analysis of poverty and strangeness. In the 17 projects assigned to these three areas, a wide range of central questions is explored. Additionally, as a result of overlaps between the single questions and the use of sources essential to all three areas, there is a close interconnection of these. Because of the complexity of the field of work, the basis for the work on these questions is a heterogeneous amount of data such as texts, visual documents, archive records, personal data, atlas maps and so forth.

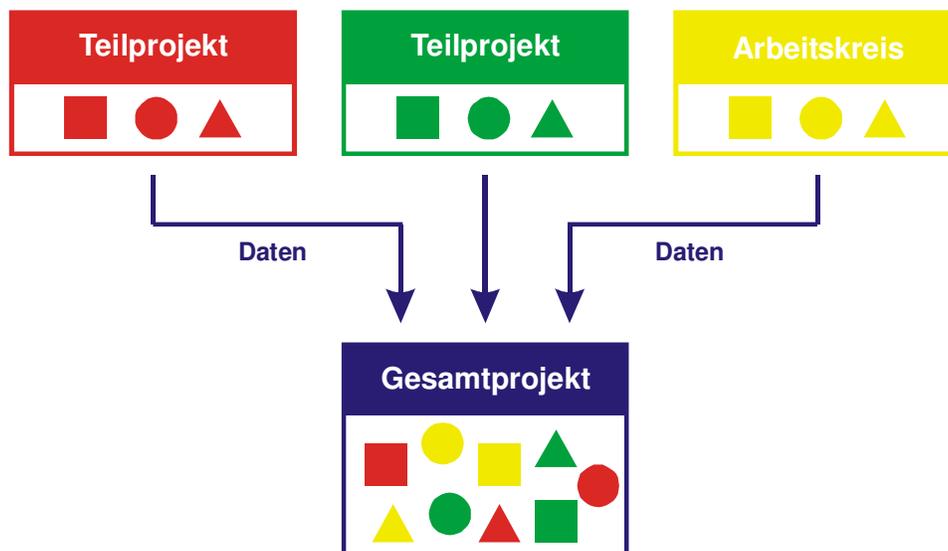


Fig.1: initial situation – heterogeneous data material in projects and research groups

As a rule, social scientists still work with a personal computer where they record and store data locally in a proprietary word-processing system such as Word or Access. In order to build up a corpus of sources and literature accessible for all colleagues for the work on common aspects of research, the separate files have to be put together into one file. With ongoing data entry, it is necessary to repeat this work regularly in order to build up a common basis of data. Consequently, the necessary effort for the update gets bigger with the increase in participating projects and the rising numbers and diversity of the recorded documents.²

2 Objective

Therefore, in cooperation with the projects, a decentrally working system is set up with which all colleagues of the CRS can record and manipulate their data as well as - later on - publish them on the internet. The aim is the development of a database application based on a client/server architecture, which offers the following prospects for the common work in the CRS 600:

This application provides the researcher with forms of various types in order to enable him to include a variety of documents into the system; moreover, it enables the development of an index to build up index-based enquiries and it supports the manipulation of the questions through the design and the building up of semantic schemata. Through this a working platform is developed to support the various steps within the process of research as, for example, the inventory of sources and literature or the collection, analysis and enquiry of source contents. Thus, the system also serves as a means for the preparation of publications of research results - be it in printed, be it in digital form.

The system is being developed within a separate project in the CRS (project Z) which for the time being is financed by the DFG and the University itself for two years (2005-2006).

² Cf. Tamara Stazic: Forschungsnetzwerk und Datenbanksystem. Erläuterungen zur Arbeit mit dem projektübergreifenden Datenbanksystem. Benutzerhandbuch, CRS 600-internal publication

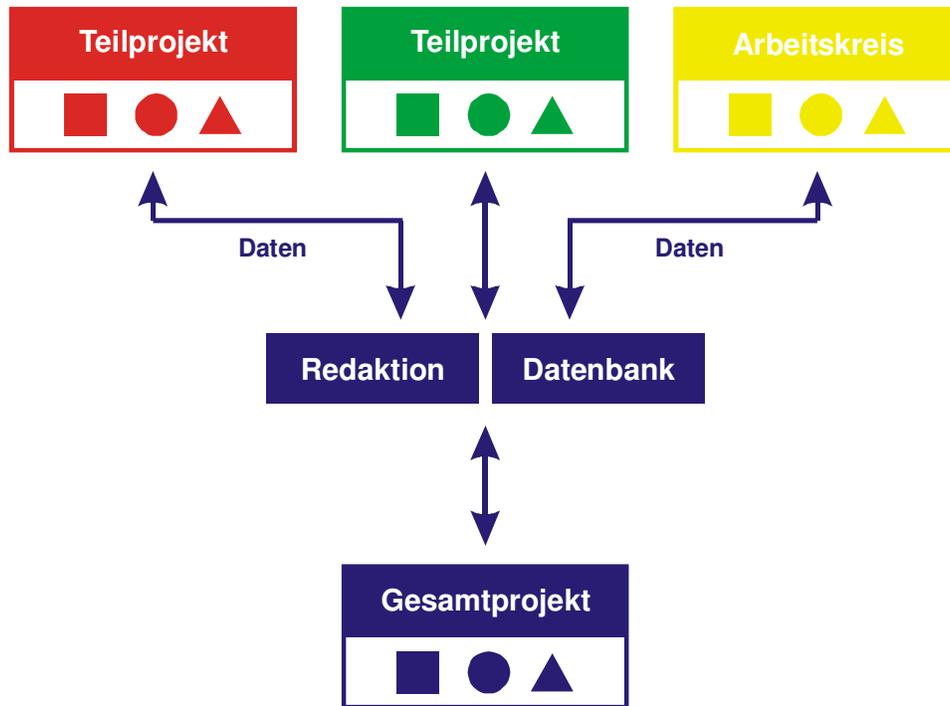


Fig.2: Objective – building up of a central database with objects encoded in a standardised way

3 Design of the System

The basic idea concerning the design of the system required is the division of the application into service provider and customer. This is what you call a client/server system whose basic architecture is depicted in figure 3. In this approach the working process is split into two parts.

The service provider (server) offers a number of services via a specified interface. The customer (client) sends requests to the server, manipulates the returned values and presents the results. The exchange of information between the two parties involved runs within the network according to a specified communication protocol.

In principle it is possible that several clients can use the services of the server simultaneously. For that reason the server has a control and synchronisation component which specifies the order of requests. Besides the synchronisation of client requests the server also takes charge of the communication with the databases and returns the results of the requests to the clients.

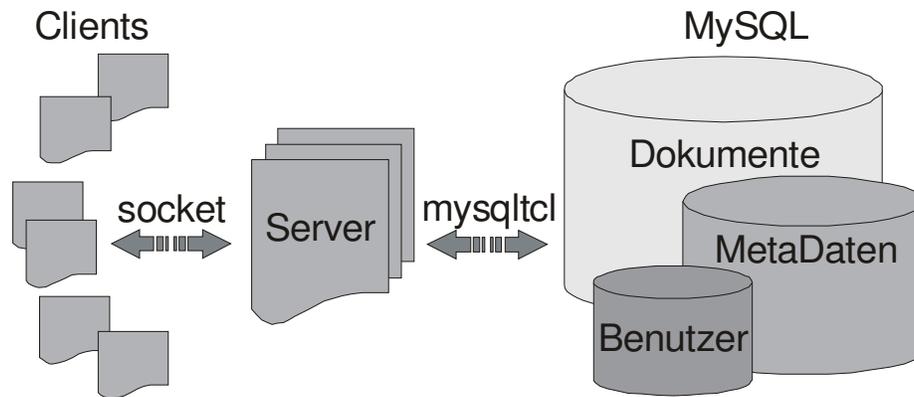


Fig.3: Client/server architecture of the FuD system

The application software of the client consists basically of two components. The application component sends requests to the server and passes them on to the network software. It sends the requests through the network to the server, waits for results and returns these to the application component which might process further operations on the basis of the returned results. Then, as second part of the client software, the presentation component manipulates the results by displaying it within the graphical user interface and prepares it for interaction with the user.

The server acts as intermediate between the different stations of work and the resources of information. Moreover it has to possess a network component which is able to interpret the communication protocol used by the client. At this point incoming requests are recognised and prepared for further processing. The resources of information are stored in the databases, which are responsible for the administration of application data.

The relational database constitutes the third level within the whole application. It is based on a corresponding data model. This model describes in a structured and formal manner the way in which data and relations between this data are displayed. As a rule the design of the data model is based on a data analysis, during which the parameters relevant to the modelling are specified and are classified according to their types. The data model of the research network is partly depicted in figure 4.

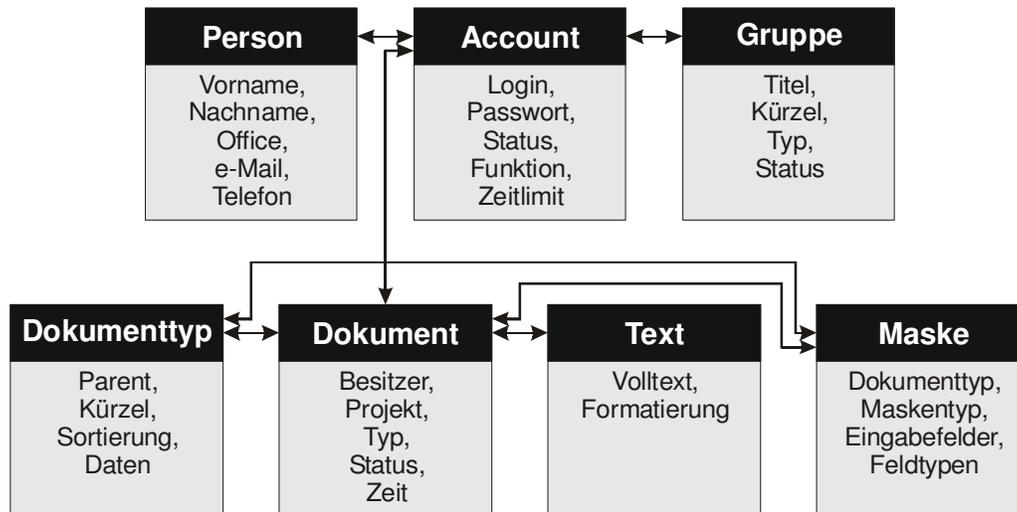


Fig.4: Central data model for the administration of user and document data

Figure 4 shows the levels for the modelling of the users and the documents proper in a schematic way. As can be seen, users are stored in a customary user database and have access to the system via one or more accounts. Each account can belong to one or more groups through which, for example, the projects of the CRS or any other research group can be specified. Additionally certain functions as for example “project director” or “guest” are assigned to every single account. These have effects on the disposable functions of this account. Thus a “guest” can, under certain circumstances, only have the right to read the documents, whereas a “project director” can not only make changes on his own documents but also on documents of his section’s colleagues. Thus, the user management of the system allows a distribution of rights, which ranges from looking at and recording documents to their indexing and linking with other documents. An owner of a document can pass on access to its contents and its header information to other users at any time.

The documents themselves are also modelled in the data scheme. Each document is mainly represented by its type by which also its most important characteristics are specified. The document type influences its storage as well as the access on its contents at a later point in time, for instance via an appropriate search engine. The document type also specifies the layout of its associated inventory form. The relation between “document” and “form” in the

data model specifies an amount of relevant fields as well as their respective types, which is automatically transferred into a corresponding form within the graphical interface. Via the table “text” the contents of the document is stored as full text. Additional formatting details provide for an appropriate display in a text window of the graphical user interface.

4 Graphical User Interface

The graphical user interface of the client offers functionalities in the main areas “user management”, “document inventory”, “document choice and its release”, “indexing and analysis of contents”, “database enquiries” as well as “data export”. In the following the areas of “document choice”, “inventory” and “analysis of contents” are going to be described exemplarily.

After user identification via a login form in which the user authorises himself via an account that has been entered for him in the database, the list of documents available for him is displayed (cf. figure 5).

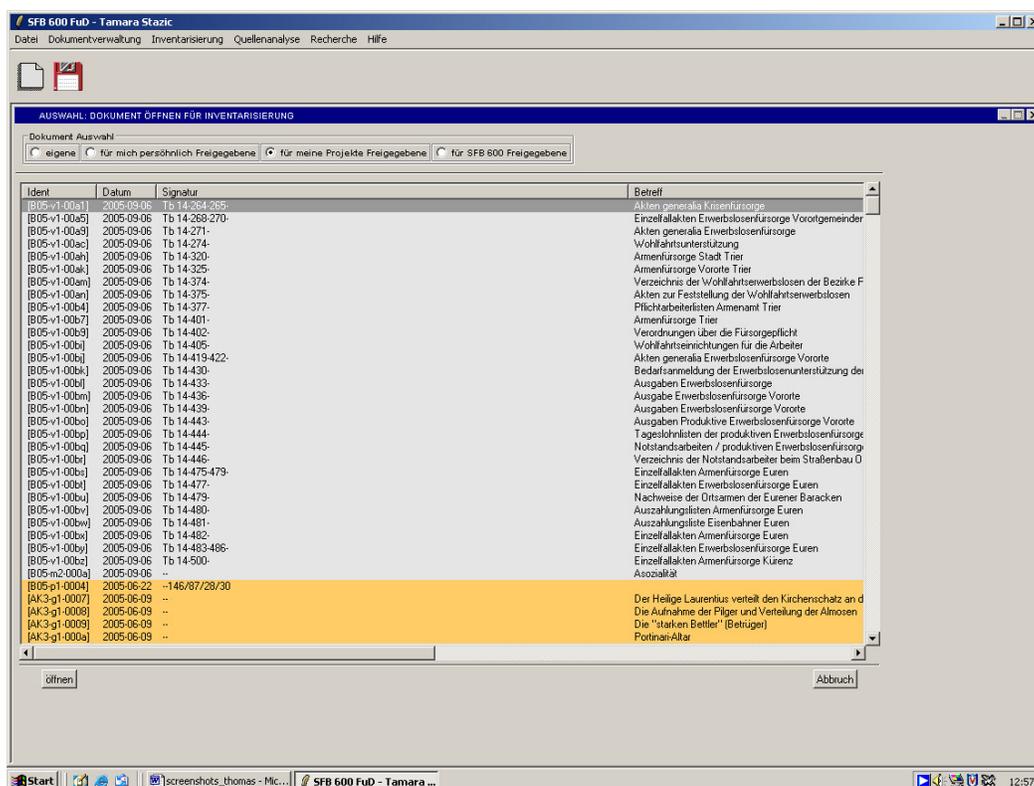


Fig.5: list of available documents

Within the system, documents are identified via a unique identification number which consists of an abbreviation for the project or the research group by which it has been put into the system, an abbreviation for the document type (e.g. "v1" for Verwaltungsschriftgut (administration documents)) and an auto-incremented four-figure number specified for these two criteria. All further assignments of information such as entries in the subject matter index are associated by this document identification. The other columns in figure 5 show the date of document entry, a call number as well as a short subject line through which the contents or the function of the document in question can be presented in short.

Via the menu above the list of documents the user can configure the displayed list by listing, as you can see in figure 5, only those documents "released for his own projects". Further restrictions are possible by reducing the list to the user's own documents or to documents personally released for the user. On the other hand he can also have himself composed a complete list of all available documents of the whole CRS 600.

Besides this choice via the inventory lists of the system which only consider the belonging of the documents to users, projects or research groups and a release for other participants connected to it, the system also possesses a comfortable search engine. With the help of this search engine, a choice can be made via the contents of the document, via the given categories of inventory or the index of the documents.

The inventory is done via document specific forms, as they are specified in the database depending on the document type (cf. figure 4). These forms are generated automatically from information contained in the database and then integrated in the interface as a Tcl/Tk code fragment. Figure 6 shows a form for a document of the type "archive / administration document" generated accordingly. The number of input fields, their labelling and input types (e.g. text, number value or date) are read from the database and automatically transferred into the shown layout. Through this way of form management, the system can easily be enlarged by additional documents.

The whole form is divided into seven logical sections. Apart from the field where you can enter inventory data, as it is shown in figure 6, the further units are interface components to insert a short description of the document, the

text proper (that is the contents of the document), a possible appendix and commentary as well as personal entries and particularly the global index of the document.

Fig.6: Inventory form for a document of the type "archive / administration document"

The values entered into the single fields can later be requested by means of the general search engine of the system. This is particularly helpful with requests concerning the index, since a kind of systematic catalogue of documents taking into account various criteria such as personal names, statements of place or subject index is created, by which a more exact choice of documents than via the pure inventory list as it is shown in figure 6 can be made.

An additional level of quality regarding the content analysis of the documents can be achieved via a further central interface component of the system. In the so-called RegIE environment (Rechner gestützte Index-Erstellung / computer supported creation of indexes) the user can systematically highlight the contents of the document and link the highlighted sections with various hierarchically organised indexes. An example for this is shown in figure 7. On

the right side of the screen the text with some highlighted passages is displayed. These passages correspond to the associations with the indexes shown on the left side. The different colours refer to different areas of the index (the yellow sections, for example, belong to the field “poverty”, the orange ones to “inclusion/exclusion”). The linking between text and index is bidirectional. In the index the linking with the documents can be seen from the document identification; a list of documents from the whole system connected with this entry being displayed for each index entry. Furthermore, as shown above, only the documents released for the user are displayed. The other direction of the linking results from the fact that the user can choose a highlighted passage and subsequently the area of index appears in the index window on the left side. By this he can determine all accessible documents, in which a passage is stored in the same index entry.

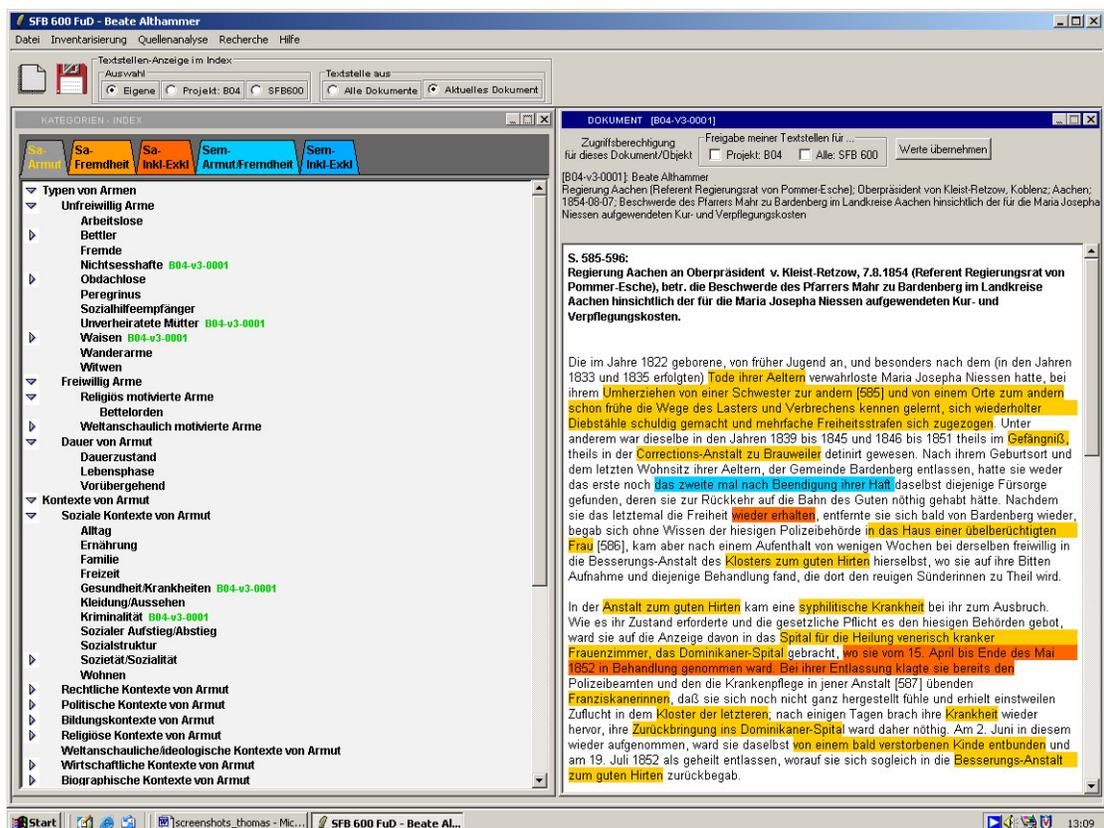


Fig.7: Content analysis of a document by means of the RegIE component

The advantage of the system results from the fact that, because of the decentral work on common data, all changes and extensions within the analysis of the document can be communicated immediately to the registered users. In this way, redundancies and inconsistencies within the data are largely avoidable.

5 Prospects

Since autumn 2004 the system has been used in the projects of the CRS 600, who are for the time being mainly responsible for its input. In the course of the ongoing development a first version which is stable and usable for the whole CRS is going to be ready by autumn this year and in a further step is going through a phase of intensive testing integrating this version into the regular working process. This phase lasts until summer 2006. The development of the system includes as a central part, among other things, a publication component through which the results can either be published online or in a printed version. The plans for this comprise a standardised XML-export which then can be prepared for publication via XSLT-applications for the internet or via typesetting (e.g. TUSTEP) for the print version.