

PROJECT GOALS

Nowadays, in Digital Libraries, the non-textual documents are indexed and accessed based on textual meta data. This kind of meta data is expensive to obtain, and in many cases, the content cannot be described completely and free of ambiguities. PROBADO aims to overcome this limitation by developing content-based access methods for non-textual documents. This goal is important due to the ever increasing amount of non-textual documents and the demand for novel accessing methods by users in the R&D and academic domains. PROBADO addresses the following Digital Library challenges:

► **Content-based indexing.** Automatic content-based indexing is important to scale with the growing number and complexity of non-textual documents. This is particularly important in the light of tight library budgets, which prohibits manual mass indexing. PROBADO provides innovative and effective content-based indexing strategies for two exemplary document types: 3D architectural documents (3D shape representation), and music documents (audio waveforms, scores, and artwork). For both document types, state-of-the-art content-based access methods are developed.

► **Innovative user interfaces.** Content-based querying demands highly interactive, domain-specific user interfaces. Textual interfaces are no longer an option, as users like to describe the shape of a 3D model or a music document to be retrieved by a sketch or by a similar example piece, respectively. Within the PROBADO presentation layer, specialized visual-interactive user interfaces are developed.

► **Citability.** Specifically in scientific environments, documents need to be citable. PROBADO addresses this demand by supporting repository hosts with DOI and URN assignment.

► **Other added-value services.** As PROBADO consists of an integrated core layer, specific added-value services can be implemented in a scalable way. The PROBADO consortium researches several added-value services including user profiling and recommendation, provision of document annotations, and result-set visualization.

The PROBADO project is sponsored by the German Research Foundation DFG (Deutsche Forschungsgemeinschaft) within the program for Scientific Library Services and Information Systems (LIS).



PROJECT CONSORTIUM :

Technische Universität
Darmstadt,
Fachgebiet Graphisch-
interaktive Systeme

Universität Bonn,
Institut für Informatik III,
Multimedia-Signalverarbeitung

Universität Bonn,
Institut für Informatik II
Computer Graphik

Technische Universität Graz,
Institut für ComputerGraphik
und WissensVisualisierung

Technische Informations-
bibliothek Hannover

Bayerische Staatsbibliothek
München

CONTACT :

Coordinator

Dr. Tobias Schreck, Technische Universität Darmstadt
tobias.schreck@gris.informatik.tu-darmstadt.de

3D - Graphics

Dipl.- Ing. M.Sc. Ina Blümel, TIB
ina.bluemel@tib.uni-hannover.de

Music

Dipl.-Inf. David Damm, Universität Bonn
damm@informatik.uni-bonn.de

PROBADO Core

Dipl.-Inf. Maximilian Scherer, Technische Universität Darmstadt
maximilian.scherer@gris.informatik.tu-darmstadt.de



PROBADO

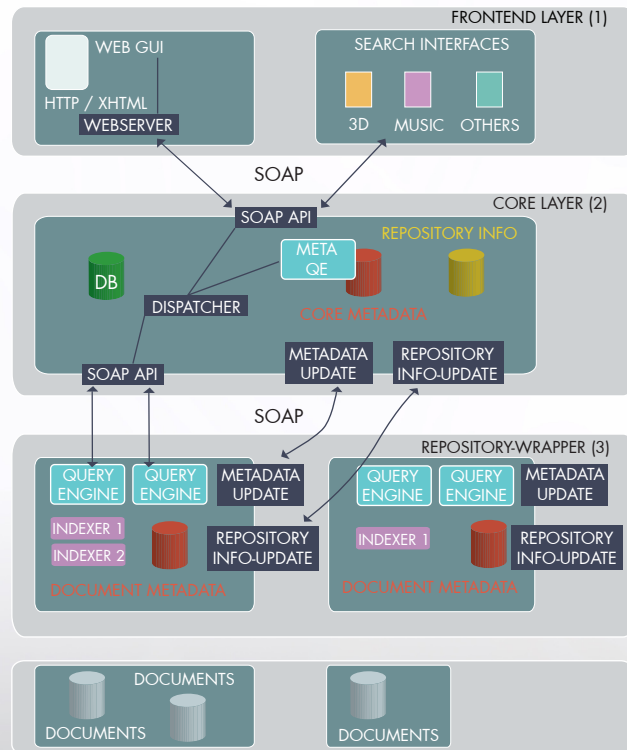
3D GRAPHICS MUSIC

Innovative
Digital Library
Services for Non-Textual
Documents



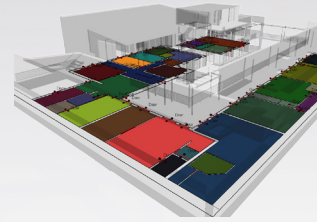
PROBADO SYSTEM ARCHITECTURE

The PROBADO framework allows integration of content-based indexing and retrieval methods for multimedia documents for Digital Library applications. The PROBADO architecture follows a three layer approach which consists of a repository layer, a core system layer, and a presentation layer (cf. the diagram below). Distributed local repositories implement document-type specific indexing and accessing techniques, including rich meta data models. The PROBADO core layer keeps track of all document repositories registered in the system. It maintains an integrated index of all documents. The presentation layer offers rich user access methods, including graphical query specification, and document visualization. PROBADO defines a system protocol based on Webservice technology. It allows dispatching content-based and metadata-based user queries to local repositories, which manage the primary documents. Synchronization methods allow the repositories to inform the core system about availability and updates of hosted contents.



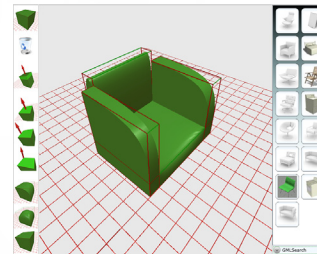
ADVANTAGES

The PROBADO approach is flexible in that it allows connecting any specialized document repository that implements the PROBADO protocol to the system. While the current reference implementation supports 3D architectural and music documents, other document types can be integrated seamlessly. A key advantage offered by this approach is that added-value services such as user recommendation, relevance feedback, right management, and result set visualization are provided as a system service that is available to any integrated repository. Thereby, the approach is highly scalable.



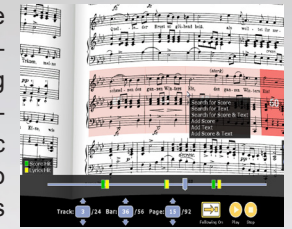
REFERENCE REPOSITORY TECHNOLOGY: 3D ARCHITECTURAL MODELS AND MUSIC

Two reference repository implementations are developed by the PROBADO consortium. These allow for automatic extraction of descriptors from 3D architectural model data and music. This special kind of meta data is integrated with possibly existing textual meta data in the respective repository data schemes.



In the 3D repository, methods for two kinds of descriptors are supported. Global shape descriptors capture information relating to the global shape of an architectural model, and support query-by-example style user queries. Structural building descriptors capture information on the configuration of rooms based on arrangement of rooms, and story levels. The latter descriptors support specific queries for building configurations. Both types of descriptors are extracted fully automatically based only on the model geometry. Graphical query interfaces include a sketch editor which allows easy entering of query shapes to be searched and a graph editor for room configuration queries.

In the music repository, the stored data includes audio concert streams, and corresponding score sheets and CD cover artwork. Specialized Optical Music Recognition algorithms help to convert the scanned score sheets to a symbolic representation. Together with appropriate audio analysis algorithms, the score sequences can be linked to audio streams. User query modalities in the music repository include query-by-example for similar subsequences in music pieces based on humming, selection, or specification by a virtual piano keyboard. A sophisticated music player was developed which in addition to playing a found music piece, allows following the piece in the score by highlighting of linked sections.



PROBADO WEB FRAMEWORK

PROBADO offers rich support for non-textual documents. Access relies these documents appropriately requires audio-visual interfaces for search formulation and results display. The PROBADO web framework consolidates the diverse access strategies offered by the presentation layer in a web application. The PROBADO web application supports repository-local or global search based on textual meta data, which is compatible with standard HTML/JavaScript. Content-based access is possible via rich web application modules which relies on plug-in technology such as Microsoft Silverlight. Both standard and content-based accessing approaches are integrated in the PROBADO web framework.

