Information Systems Research at RWTH Aachen

Matthias Jarke
Informatik V (Information Systems),
RWTH Aachen,
Ahornstr. 55, 52056 Aachen, Germany

1 Overview

With about 8.000 researchers and 40.000 students, RWTH Aachen is the largest technical university in Europe. The science and engineering departments and their industrial collaborators offer a lot of challenges for database research.

The chair Informatik V (Information Systems) focuses on the theoretical analysis, prototypical development, and practical evaluation of *meta information systems*. Meta information systems, also called repositories, document and coordinate the distributed processes of producing, integrating, operating, and evolving database-intensive applications.

Our research approaches these problems from a technological and from an application perspective.

On the one hand, we pursue theory and system aspects of the integration of deductive and object-oriented technologies. One outcome of this work is a deductive object manager called ConceptBase which has been developed over the past eight years and is currently used by many research groups and industrial teams throughout the world.

On the other hand, a wide range of application-driven projects aims at building a sound basis of empirical knowledge about the demands on meta information systems, and about the quality of proposed solutions. They address application domains as diverse as requirements engineering, telecommunications, cooperative engineering, organization-wide quality management, evolution of chemical production processes, and medical knowledge management. They share the vision of supporting wide-area distributed cooperation not just by low-level interoperation technology but by exploiting conceptual product and process modeling.

Under the direction of M. Jarke, Informatik V comprises three research groups with a total of twenty senior researchers and doctoral students: distributed information systems (leader: Dr. Manfred Jeusfeld), process information systems (Dr. Klaus Pohl), and knowledge-based systems (Prof. Wolfgang Nejdl). Database-related activities also exist in the Software Engineering and Applied Mathematics groups.

2 Distributed Information Systems

M. Jarke, M. Jeusfeld

The globalization of business as well as science and engineering, together with organizational trends towards flexible alliances between smaller units, demand synchronous and asynchronous wide-area distributed teamwork. Increased availability of communication facilities via the Internet is just a first step towards an infrastructure for such collaborations. Our main interest is in the use of meta information systems for improved documentation and coordination of heterogeneous collaboration environments.

Our basic research hypothesis has been that a trader system exploiting deductive and object-oriented meta modeling offers both the structuring and the reasoning facilities necessary to organize, control, and evolve such collaborations. This hypothesis has been studied at several system levels and in various application fields, often using the ConceptBase system.

Formal foundations for the integration of deductive and object-oriented technologies have been studied in the COMPULOG Esprit project. Along with many smaller application projects, two large interdisciplinary consortia in the industrial engineering and medical domains have been exploiting the system integration and evolution facilities of ConceptBase for cooperative information systems applications. Additional infrastructural support is provided by a Sun Site grant in which our group operates a major Internet server donated by Sun Inc. Technical foundations of the underlying distributed database technology are investigated in the INDIA project, conducted jointly with the Philips Research Laboratories.

2.1 COMPULOG

M.Jarke, M.Jeusfeld, H.W.Nissen, M.Staudt

ConceptBase supports the conceptual modeling language Telos designed jointly with the group of John Mylopoulos (University of Toronto) in the second half of the 1980's [1, 2]. Research on the formal foundations of ConceptBase has been funded from 1988-1995 by the European Community under ESPRIT basic research project COMPULOG.

Telos is characterized by a deep integration of deductive and object-oriented aspects, with an emphasis on simplicity.

A Telos model can be seen as a logical theory expressed in Datalog with stratified negation whose object structure is enforced by a few facts, rules, and integrity constraints that must be part of any consistent object base. This perspective gives a clear and simple formal semantics to the language; it also allows the reuse of optimization techniques developed for deductive databases.

A Telos model can also be seen as an object-oriented database in which methods are declaratively specified (only) by deduction rules and integrity constraints, and implemented as triggers which can (therefore) be generated automatically from the specifications. This perspective provides a link to active (relational or object-oriented) databases as an implementation layer underneath.

The combination of both perspectives proves particularly helpful in supporting repositories with an unrestricted number of meta levels where each level constitutes a type system for the level below.

A first major result, described in the doctoral thesis of M.Jeusfeld, was the axiomatization of Telos as a deductive database, and the exploitation of this axiomatization for the optimization of rule evaluation, integrity checking, and partial evaluation of so-called meta constraints which span more than two classification levels in the object base. Such constraints, a necessary prerequisite for an extensible language, would lead to extremely inefficient evaluations with the traditional optimization techniques.

A second consequence of the combination of deductive and object-oriented technologies was the introduction of so-called query classes [3]. Query classes are parameterized views on Telos databases which have the same dual nature as Telos models, and have proven useful for many purposes, including user interface support, security management, version and configuration management, and object base analysis.

Recently, results from terminological reasoning have been investigated for the semantic optimization of query class evaluation [4]. Other current work includes an agent-oriented extension of Telos in which teams of ConceptBase users can co-evolve multiple viewpoints on some current or future reality.

[1] J.Mylopoulos, A.Borgida, M.Jarke, M.Koubarakis "Telos – representing knowledge about information systems", *ACM Trans. Information Systems* 8(4), 1990.
[2] M.Jarke, S.Eherer, R.Gallersdoerfer, M.Jeusfeld, M.Staudt "ConceptBase – a deductive object base for meta data management" *J. Intelligent Information Systems* 4(2), 1994.
[3] M.Staudt, H.W.Nissen, M.Jeusfeld "Query by class, rule, and concept", *Applied Intelligence* 4(2), 1994.
[4] M.Buchheit, M.Jeusfeld, W.Nutt, M.Staudt "Subsumption between queries to object-oriented databases", *Information Systems* 19(1), 1994.

2.2 Cooperative IS Applications

B.v.Buol, M.Jarke, M.Jeusfeld, P.Peters, S.Sklorz, P.Szczurko

Application projects in industrial engineering and medical systems investigate the scope of support offered by systems such as ConceptBase in an Internet environment.

Total quality management (TQM) is a business strategy that promotes customer-orientation and rapid adaptation to change throughout the organization [1]. As part of a five-year TQM programme funded by the German Federal Ministry of Research, WibQuS, a consortium of industrial engineering, ergonomics, and computer science researchers has investigated the role of (knowledge-based) information systems technology in important TQM methods such as Quality Function Deployment, Statistical Process Control, and Service Feedback Analysis. Our task in this consortium was to support the interaction of these methods in a wide-area cooperative information system based on conceptual modeling technology.

WibQuS has demonstrated the benefits of formal, computer supported conceptual models in three areas: cooperative modeling which bridges the terminological gap across departments; incremental generation of interoperability code which bridges the system gap; and workflow management which supports the organizational implementation of quality strategies that span multiple departments.

In the first step, industrial engineers in five German cities developed and interrelated formal conceptual models of the different quality methods via the Internet, using ConceptBase with a pre-agreed shared meta model of the quality cycle. The result was the first formal conceptual model of the complete quality cycle; despite the overhead of learning the Telos language and living with the technical problems of working in an Internet-based client-server architecture, a significant gain in quality and total efficiency of model development was reached in comparison to traditional CASE methods used by parallel projects in the research program.

A second advantage of the integrated conceptual model is that it could be mapped semi-automatically to the target runtime environment which consists of a network of different brands of relational databases linked by OmniSQL from Sybase. These databases reflect the data and services offered by the quality management subsystems in the individual departments linked to the system. Evolution support is ensured by keeping ConceptBase and OmniSQL in the runtime environment as a "trader".

In the third and last step, the conceptual model is further utilized for the planning, generation, and monitoring of workflows in this interoperable architecture. The pattern is the same as in the second step: workflow specifications are extracted from the conceptual model and their individual steps are represented and documented in the relational subsystems. A simulation model based on system dynamics is used for the planning step. Further research will study a combination of symbolic and sub-symbolic machine learning for process improvement in quality management.

In a related industry project called FAITH, some of the WibQuS ideas were applied to the development of a comprehensive quality management environment for the expert systems that control chemical production and sewage processes in a large German chemical firm [3]. This project confirmed the significant productivity and safety advantages of a specification-level approach to quality management.

MEDWIS is a large ten-year project in which about twenty medical research groups in Germany investigate the embedding of knowledge-based components in medical application systems. Since late 1994, an approach very similar to the one followed in WibQuS is being pursued in linking individual components into a cooperative information systems architecture.

[1] M.Jarke, K.Pohl "Information systems quality and quality information systems" *Proc. IFIP 8.2 Conf. Impact of Computer-Assisted Technology on Information Systems Development*, Minneapolis, Mi, June 1992.

[2] M.Jarke, M.Jeusfeld, P.Szczurko "Three aspects of intelligent cooperation in the quality cycle" *Intl. J. Intelligent and Cooperative Information Systems* 2(4), December 1993.

[3] K.Finke, M.Jarke, R. Soltysiak, P.Szczurko "Testing expert systems in process control" *IEEE Trans. Knowledge and Data Engineering*, to appear.

2.3 INDIA

R.Gallersdoerfer, M.Jarke

The main problem in the WibQuS and MEDWIS projects is semantic heterogeneity and distribution of expertise. In contrast, the INDIA project focuses on performance in a more homogeneous distributed database. The application problem is the interleaving of service creation and service operation in intelligent telecommunications networks (IN) with a potentially very large number of accesses in a short time.

INDIA pursues a solution strategy which combines constraint-oriented fragmentation and allocation design with a concurrency control strategy called Atomic Delayed Replication (ADR) [1]. In contrast to solutions currently offered by DBMS vendors, ADR offers a clearly defined and well-controlled relaxation of consistency across replicas to gain performance via scalability of commodity database systems. Our implementation exploits the active database facilities of the Sybase relational DBMS, at the same time extending them to the detached execution mode proposed, e.g., in the HiPAC project.

The project is conducted jointly with the Philips Research Labs in Aachen, and the implementation has been integrated in their IN environment. Current work focuses on performance evaluation by analytical models and system measurements, and on automated support for dynamic load-balancing in INDIA's distributed database design method.

[1] R.Gallersdoerfer, M.Jarke, K.Klabunde "Intelligent Networks as a data-intensive application (INDIA)" *Proc. Intl. Conf. Applications of Data Bases*, Vadstena, Sweden, 1994.

3 Process Information Systems

M.Jarke, K.Pohl

The term "process" means different things to different people. Our group is particularly concerned with understanding, supporting, and improving the informal and formal *modeling processes* taking place in the early phases of development (also in maintenance, reuse, or integration) of complex systems such as information systems, chemical plants and production processes, factory layout, and pharmaceutical research. In contrast to traditional workflow applications, these processes share the problem that very little is known about them. Nevertheless, they must be carefully traced, not only because they often produce important results whose rationale must be known for future reference but also to learn more about the processes themselves. Moreover, process models for these tasks must take care not to overly constrain creative work. In particular, it must avoid the frequent mistake to misuse descriptive process abstractions as guidance models.

3.1 NATURE

R.Doemges, P.Haumer, M.Jarke, H.W.Nissen, K.Pohl

Requirements engineering (RE) maintains the link between business evolution and information systems development and is therefore crucial for change management in information-intensive organizations [7]. Unfortunately, RE research has been fragmented, and a coherent RE community has just begun to form.

Extending results of the DAIDA project [1, 2], Esprit project NATURE explores Novel Approaches to Theories Underlying Requirements Engineering, especially concerning aspects of representation, domain, and process [3, 5, 6].

A major result to date has been the unification of RE research streams in a three-dimensional framework along which RE processes can be described and traced [4]. This framework has been mapped to a process-oriented and repository-centered architecture which we have demonstrated in the PRO-ART requirements engineering environment [8].

The PRO-ART environment contains a suite of CASE tools which support progress along the three dimensions of the framework, namely representation (from informal to formal), domain specification (from opaque to sufficiently complete coverage), and agreement (from individual viewpoints to consensus). These tools are process-integrated in a novel way such that they accept guidance from declarative process guidance models, and trace their behavior in terms of formal traceability models. Guidance and traceability models are specified in the PRO-ART repository (ConceptBase) under a common process meta model in order to allow experience-based process improvement. Due to the poorly understood nature of the requirements process, a decision-oriented and situation-based process meta model has been developed which allows the user great freedom and offers support only where actual knowledge is available.

Current research focuses on the empirical validation of the results and on improved support for very large distributed requirements engineering projects.

[1] M.Jarke, J.Mylopoulos, J.W.Schmidt, Y.Vassiliou "DAIDA – an environment for evolving information systems" *ACM Trans. Information Systems* 10(1), 1992.
[2] M.Jarke (ed.) *Database Application Development with DAIDA*, Springer-Verlag 1993.
[3] M.Jarke, J.Bubenko, C.Rolland, A.Sutcliffe, Y.Vassiliou "Theories underlying requirements engineering – an overview of NATURE at genesis" *Proc. IEEE Symp. Requirements Engineering*, San Diego, Ca, 1993.
[4] K.Pohl "The three dimensions of requirements engineering" *Proc. 5th Intl. Conf. Advanced Information Systems Engineering*, Paris 1993.
[5] M.Jarke et al. "Requirements engineering – an integrated perspective on representation, process, and domain" *Proc. 4th European Software Engineering Conf.*, Garmisch, Germany, 1993.
[6] M.Jarke, K.Pohl "Establishing visions in context: towards a model of requirements processes" *Proc. 16th Intl. Conf. Information Systems*, Orlando, Fl, 1993.
[7] M.Jarke, K.Pohl "Requirements engineering in 2001: (virtually) managing a changing reality" *IEE Software Engineering Journal*, November 1994.
[8] M.Jarke, R.Doemges, H.W.Nissen, K.Pohl "Requirements information management – the NATURE approach" *Technique et Science Informatiques*, 1995.

3.2 Chemical Process Engineering

M.Baumeister, M.Jarke, K.Pohl

Process simulation is an important activity in the planning and re-engineering of chemical production processes. Simulation models can become extremely complex, involving hundred thousands of differential equations. Current process simulation environments tend to be monolithic, proprietary, and very difficult to learn and use. In collaboration with the Process Engineering research group in Aachen (Prof. Marquardt) and with European companies, this project explores the use of conceptual product and process modeling technology in the development and maintenance of process simulation models. First results indicate that several lessons learned from our requirements engineering work can provide useful assistance in opening up the current environments, leading to improved technology transfer from engineering and OR research labs to industrial usage.

[1] M.Jarke, W.Marquardt "Computer-assisted process modeling" *Proc. Intl. Conf. Intelligent Systems in Process Engineering*, Snowmass, Co, July 1995.

3.3 Teamwork Support

S.Jacobs, M.Jarke, S.Kethers

Groupware and flexible workflow concepts are an essential ingredient of cooperation environments for science and engineering. In the CoNeX project (1987-1993), we developed a coherent approach which relates object evolution, task management, and decision support in teamwork settings [1, 2, 3].

The project *CoDecide* additionally focuses on team acquisition of requirements, considering multiple stakeholder goals and interests and adapting user interface ideas from the industrial engineering community. A group user interface toolkit for the confrontation of heterogeneous viewpoints has been developed and integrated with the PRO-ART requirements engineering environment [4] and with ASSIST, a knowledge-based factory layout tool developed by the Fraunhofer Institute for Logistics in Dortmund, Germany. The user interfaces generated by this toolkit can be seen as generalizations of the "house of quality", a popular presentation mechanism for quality planning.

Based on these experiences, two further projects are investigating collaboration support in industrial pharmaceutical research, and in the translation of a medical terminology standard book from English to German.

Relationships between this work and more technical aspects of interoperability and organizational modeling are pursued in European-Australian and European-Canadian projects conducted with Imperial College London, the Universities of Hagen, Hamburg, Milano, Queensland, Toronto, and British Columbia.

[1] U.Hahn, M.Jarke, T.Rose "Teamwork support in a knowledge-based software environment" *IEEE Trans. Software Engineering*, May 1991.
[2] T.Rose, M.Jarke, M.Gocek, C.Maltzahn, H.W.Nissen "A decision-oriented configuration process environment" *IEE Software Engineering Journal*, November 1991.
[3] M.Jarke, C.Maltzahn, T.Rose "Sharing processes: team support in design repositories" *Intl. J. Intelligent and Cooperative Information Systems* 1(1), March 1992.
[4] K.Pohl, S.Jacobs "Traceability between cross-functional teams" *Intl. J. Concurrent Engineering: Research and Applications*, October 1994.

4 Knowledge-Based Systems

W.Nejdl, J.Bachmeyer, J.Gamper

Research in the Knowledge-Based Systems group, led by Prof. Wolfgang Nejdl, focuses on second generation expert systems. These systems are based on a declarative system model which can be combined with specific algorithms for applications such as diagnosis, repair, and planning. Individual research activities concern

- the efficient implementation of systems based on logical and object-oriented modeling formalisms [1]
- the representation of temporal relationships in dynamic systems [4],
- specific applications in the diagnosis of hardware and software components [2]
- consistency checking of knowledge bases [3].

A belief revision system called DRUM is the technical basis for many of these applications.

In 1995, the group will move to the Technical University of Hannover where Prof. Nejdl will take over a chair in Technical Computer Science.

[1] W.Nejdl, S.Ceri, G.Wiederhold "Evaluating recursive queries in distributed databases" *IEEE Trans. Knowledge and Data Engineering* 5(1), 1993.
[2] T.Brunner, W.Nejdl, H.Schwarzjirg, M.Sturm "Online expert systems for power system diagnosis and repair" *J. Intelligent Systems Engineering*, Spring 1993.
[3] G.Friedrich, G.Gottlob, W.Nejdl "Formalizing the repair process - extended report" *Annals of Mathematics and Artificial Intelligence* 11, 1994.
[4] W.Nejdl, J.Gamper "Harnessing the power of temporal abstractions in model-based diagnosis of dynamic systems" *Proc. European Conf. Artificial Intelligence*, Amsterdam 1994.

5 Other Database Research in Aachen

In this section, we briefly sketch a few database projects conducted in other groups of the Informatics department at Aachen. Applied database research is, of course, also conducted in many of the engineering institutes.

5.1 GOM

G.Moerkotte, Informatik III

The group of Prof. Guido Moerkotte, established in late 1994, is concerned with the development of object base management systems for engineering applications. Special emphasis is put on performance enhancements. The research is carried out within the GOM project where groups from the University of Karlsruhe under Prof. Lockemann, the University of Passau under Prof. Kemper, and the RWTH Aachen under Prof. Moerkotte collaborate.

The special interest of the group in Aachen is the development of practical and theoretical results for building full fledged query optimizers for object bases as well as the construction of special purpose index structures. For query optimization there exists a loose cooperation with Dr. Cluet from INRIA.

[1] S.Cluet, G.Moerkotte "On the complexity of generating optimal left-deep processing trees with cross products" *Proc. Intl. Conf. Database Theory*, to appear 1995.
[2] A.Kemper, G.Moerkotte, K.Peithner "A blackboard architecture for query optimization in object

bases" *Proc. Intl. Conf. Very Large Data Bases*, Dublin 1993.
[3] A.Kemper, C.Kilger, G.Moerkotte "Materialization in object bases" *Proc. ACM-SIGMOD Conf.*, Denver 1991.
[4] C.Kilger, G.Moerkotte "Indexing multiple sets" *Proc. Intl. Conf. Very Large Data Bases*, Santiago de Chile 1994.

5.2 GRAS/PROGRES

A.Schuerr, B.Westfechtel, Informatik III

The Software Engineering group (Prof. M.Nagl) has designed and implemented a graph-oriented DBMS intended for the support of CASE environments and similar domains. In a multi-client/multi-server architecture, the GRAph-oriented database Server GRAS [1] supports incremental computation of derived data, undo/redo of data modifications, error recovery from system crashes, and version control.

The programming environment PROGRES [2] on top of GRAS supports the definition of graph-oriented database schemata, queries, and updates. It offers textual as well as graphical notations for the definition of graph schemata, and so-called graph rewrite rules as a declarative and visual formalism for the specification of queries and updates. It also supports deterministic and non-deterministic programming with built-in backtracking and cancelling arbitrary sequences of failing graph modifications.

[1] N.Kiesel, A.Schuerr, B.Westfechtel "Design and evaluation of GRAS, a graph-oriented database system for engineering applications" *Proc. 6th Intl. Workshop on Computer-Aided Software Engineering*, Montreal 1993
[2] G.Engels, C.Lewerentz, M.Nagl , W.Schaefer, A.Schuerr "Building integrated software development environments, part I: tool specification" *ACM Trans. Software Engineering and Methodology*, 1(2), 1994

5.3 DEDUCT

M.Meskes, Applied Mathematics and Informatics

DEDUCT is a deductive front-end for relational database systems such as ORACLE and INGRES, developed as part of a project to improve structuring and rewriting techniques for deductive databases. Results to date include: the identification of a large class of Datalog programs for which no runtime checks are needed, and an extended magic-set transformation which works on stratified Datalog and takes advantage of existing faster algorithms for more specialized classes where possible [1]. Current work concerns the inclusion of aggregate functions in deductive optimization methods.

[1] M.Meskes, J.Noack "The generalized supplementary magic-sets transformation for stratified Datalog" *Information Processing Letters* 47, 1993