Abstracts of Papers in Post-Proceedings

1. Michael Soffner, Norbert Siegmund, Marko Rosenmüller, Janet Feigenspan, Thomas Leich and Gunter Saake. **A VARIABILITY MODEL FOR QUERY OPTIMIZERS**

By adopting to more domains, database management systems (DBMSs) increase their functionality continuously. This leads to DBMSs that include often unnecessary functionality, which decreases performance. A result of this trend is that new specialized systems arise that focus only on a certain application scenario but often reimplement already existing functionality. To overcome overbloated DBMSs, we propose to introduce variability in DBMS implementations that allows users to select only needed functionality for a specific application scenario. In this paper, we focus on the query optimizer as it is a key component of DBMSs. We describe the potentials of tailoring query optimizers. Furthermore, we analyze common and differing functionality of three query optimizers of industrial DBMSs (SQLite, Oracle, and PostgreSQL) to create a variability model for query optimizers that can be used as a basis for future variability-aware implementations.

2. Igor Epimakhov, Abdelkader Hameurlain and Franck Morvan. **GEOLOC: ROBUST RESOURCE ALLOCATION METHOD FOR QUERY OPTIMIZATION IN DATA GRID SYSTEMS**

Resource allocation (RA) is one of the key stages of distributed query processing in the Data Grid environment. In the last decade were published a number of works in the field that deals with different aspects of the problem. We believe that in those studies was given insufficient attention to such important aspects as allocation space definition and criterion of parallelism degree determination. In this paper we propose our method of RA that extends existing solutions in those two points of interest and resolves the problem in the specific conditions of the large scale heterogeneous environment of Data Grid. Firstly, we propose to use a geographical proximity of nodes to data sources to define the AS. Secondly, we present the principle of execution time parity between read and join operations for determination of the parallelism degree and generation of load balanced query execution plan. We conducted an experiment that proved the superiority of our GeoLoc method in terms of response time over the RA method that we chose for the comparison. The present study provides also a brief description of existing methods and their qualitative comparison with the proposed method.
3. Boris Novikov, Elena Mikhaylova, Ekaterina Ivannikova and Alice Pigul. MINING LOGS FOR LONG-TERM PATTERNS

In this work we made an approach for data storage system optimization. Most high-capacity storage systems consist of several devices. These devices may have different performance. The goal is to control data placement in such way that data are moved to faster devices just before they are expected to be intensively used. To accomplish we would like to find long term data access patterns. However the high level application logic and schedules are not available at the storage system level. Our approach is to use log mining to identify data access patterns. If the system has information about data that will soon be required for processing, it is possible to prepare the data by transferring them to a faster storage parts. We analyze the database log files containing the history query executions and identify repeating query groups. Our hypothesis is that this query groups are closely related with meaningful business processes of the application. These groups are very likely related to the business process. Knowing the business processes, we can determine the data they need.

In this paper we offer the algorithm for query groups’ detection that and describe the parameters affecting algorithm efficiency. Also we describe the algorithm for periods identifying for detected query groups.

Testing the algorithm on real production data showed that the proposed algorithm identifies more than 60% of known business processes.

4. Benameur Ziani and Youcef Quinten. COMBINING DATA MINING TECHNIQUE AND QUERY FREQUENCIES FOR AUTOMATIC SELECTION OF INDEXES IN DATA WAREHOUSES

Index selection is an important part of physical database design. Its goal is to select an appropriate set of indexes to minimize the cost for a given workload under storage constraint. However, selecting a suitable configuration of indexes is a difficult problem to solve. The problem becomes more complex for indexes defined on multiple tables such as bitmap join indexes, since it requires the exploration of a much more search space. Studies dealing with the bitmap join indexes selection problem mainly focused on proposing pruning solutions of the search space by the means of data mining techniques or heuristic approaches. So far, the data mining based approaches have used closed frequent itemsets to reduce the search space for the selection process. These approaches have two notable shortcomings. Firstly, they generate a huge number of indexes with a lot of redundancy that it is very difficult to manage according to the system limitation (number of Indexes per table, storage space constraint). Secondly, when they construct the extraction context for mining frequent sets of attributes, they have used indexable attributes only once for each query in the workload which does not reflect the importance of a given query in the workload. Indeed, the queries in a workload are unlikely to have the same probability of being requested. To overcome these imitations, we propose to combine maximal frequent itemsets and query frequencies to improve the quality of generated indexes. This paper describes an approach that refines the index selection process, incorporating query frequencies in the extraction context for mining frequent set of attributes. We experimentally prove that our approach reduces the storage space and improves the quality of the recommended indexes.
5. Janari Põld, Tarmo Robal and Ahto Kalja. **ON PROVING THE CONCEPT OF AN ONTOLOGY AIDED SOFTWARE REFACTORIZING TOOL**

Through years more and more software is produced. The quality of software architecture however has an important role in systems exploitation, as it determines the maintainability and extensibility of an application. Recently more emphasis is put on quality of the design, so that new features can be added with ease. To preserve code readability and extensibility, software architecture must be refactored from time to time to cope with the modifications. Nevertheless, reviewing the whole source code is time consuming and does not return any surplus, thus it is often skipped, causing the software architecture to decay in time over several modifications and making it harder to add new functionality in the future. An automated method of recognizing "bad" code would help to solve some of the issues. In this article the authors propose a concept of a refactoring tool, which uses ontology to find "smelly" design and tackle the aforementioned problems. Several aspects of the tool are discussed – how it works and how it can be used to improve the software architecture, thus augment the quality.

6. Kārlis Čerāns, Renārs Liepiņš, Jūlija Ovčinnikova and Arturs Sprogis. **ADVANCED OWL 2.0 ONTOLOGY VISUALIZATION IN OWLGrEd**

Intuitive ontology visualization is a key for their learning, exchange, as well as their usage in conceptual modeling and semantic database schema design. OWLGrEd is a visual tool for compact graphical UML-style rendering and editing of OWL 2.0 ontologies. We describe here the extensibility features for OWLGrEd that allow tailoring the editor for specific ontology-based modeling needs, including custom entity annotation visualizations and description of integrity constraints for semantic database schemas. We discuss the application of concrete OWLGrEd extensions in the context of ontology-centered information system engineering.

7. Uldis Donins. **FORMAL ANALYSIS OF PROBLEM DOMAIN WORKFLOWS**

The formal foundation of Topological functioning model (TFM) makes it as a powerful tool to analyze the functioning of a problem domain and to formally relate problem domain artifacts with the artifacts that should exist in solution domain. TFM captures system functioning specification in the form of topological space consisting of functional features and cause-and-effect relations among them and is represented in a form of directed graph. The functional features together with topological relationships contain the necessary information to create diagrams of other type, e.g., Activity or Class diagrams. To specify the behavior of system execution a new artifact is added to the TFM – the logical relations. The presence of logical relations within TFM denotes forking, branching, decision making, and joining during the functioning of the system. Thus it is needed to identify and carefully analyze logical relations within TFM in order to have all the necessary information to transform it to diagrams of other type. This paper gives the formal method of transforming TFM into Activity diagram together with an example of such transformation.
8. Janis Barzdins, Edgars Rencis and Agris Sostaks. TOWARDS HUMAN-EXECUTABLE BUSINESS PROCESS MODELING

There are many organizations, whose everyday life involves lots of tasks performed, or let us say executed, by lots of different people. Since nowadays processes have become much more complex, a big challenge for humans is to even understand what, when and how have to be done in order to reach their goals. Business process models are frequently used in organizations to make the process understandable to performers and to alleviate their work by connecting the process to organization's information system thus making processes human-executable. However, while developing a solution, there are usually only two extremes to choose from – either we use an all-in-one solution for describing process steps or we develop a domain-specific process modeling language from scratch. In this paper we propose the golden mean – a good base for domain-specific process modeling languages and appropriate tooling to be used in a big portion of related organizations and relatively easily integrated into their information systems. We define, what is meant to be “good” by binding the process language base with the natural language generator. We also demonstrate the approach on a case study of a process modeling language for the University of Latvia.


There are several data silos containing information about business entities and people but are not semantically connected. If in integration process of data sources trust management is also employed than we can expect much higher success rate in relations discovery among entities. Majority of current mash-up approaches that deal with integration of information from several data sources omit or don't fully address the aspect of trust. In this paper we discuss semantic integration of personal and business information from various data sources coupled with trust layer. The resulting system has higher and more defined solidity while trust for single entity and also for data source is defined. The case study presented in the paper focuses on integration of personal information from data sources mainly maintained by government authorities which have higher trustability than information from social networks, but we also include other less trusted sources. The developed SocioLeaks system allows users traversal and further relation discovery in a graph based manner.

10. Erika Asnina, Janis Osis and Asnate Jansone. FORMAL SPECIFICATIONS OF TOPOLOGICAL RELATIONS

The paper discusses application of the topological functioning model (TFM) of the system for its automated transformation to behavioral specifications such as UML Activity Diagram, BPMN diagrams, scenarios, etc. The paper addresses a lack of formal specification of causal relations between functional features of the TFM by using inference means suggested by classical logic. The result is reduced human
participation in the transformation as well as additional check of analysis and specification of the system.

11. Elena Sivogolovko. **The Influence of Data Quality on Clustering Outcomes**

Relationship between Clustering and Data Quality has not been thoroughly established. It is usually assumed that input dataset does not contain any errors or contains some "noise", and this concept of "noise" is not related to any Data Quality concept. In this paper we focus on the four most commonly used data quality dimensions, namely accuracy, completeness, consistency and timeliness. We evaluate the impact of data quality on clustering outcomes using definitions and constructs of these quality dimensions. Four different clustering algorithms and five real datasets were selected to show the interaction between data quality and cluster validity.

12. Vitaly Zabiniako. **Visualization of Graph Structures with Magnetic-Spring Model and Color-Coded Interaction**

In this paper author provides description of the original approach for visual analysis of data represented with general graphs, based on modification of magnetic-spring model and color-coded cognitive manipulation with graph elements. The theoretical background of magnetic fields in application to graph drawing is presented along with discussion of appropriate visualization techniques for improved information analysis and comprehension. Usage of other existing graph layout strategies (e.g. hierarchical, circular) in conjunction with magnetic-spring approach are also considered for improved data representation capabilities. A concept of integrated virtual workshop for graph visualization is introduced which relies on aforementioned model and can be used in GVS (Graph Visualization Systems). A case study of application of proposed approach is presented along with conclusions of its usability and potential future work in this field.


Business Process Management Systems (BPM systems) are used to control, analyze and manage business processes in organizations. BPM systems help to reduce the amount of administrative effort and focus on the processes which add value. Nowadays, moving towards cloud-based Software-as-a-Service (SaaS) architecture, some additional requirements for successful BPM implementation are identified. One of the main challenges is how to integrate SaaS BPM systems with existing on-premises systems, data sources and devices. In this paper, mobile agents are proposed as the technology addressing this new challenge. A mobile agent is a composition of computer software and data which is able to migrate from one device to another autonomously and continue its execution on the destination device. The paper starts
with and overview of SaaS BPM and existing approaches to address SaaS integration challenges. Then, the concept of mobile agents is described, and the idea of how mobile agents may be used in SaaS BPM integration scenarios is presented. The paper is continued with a comparison of widely used integration approaches with proposed mobile agents based mechanism. Finally, a newly proposed architecture is presented in a prototype, outlining its advantages and proposing directions for future research.

14. Tarmo Robal and Ahto Kalja. APPLYING USER DOMAIN MODEL TO IMPROVE WEB RECOMMENDATIONS

The enormous amount of information available over the Internet has forced users to face information overload while browsing the World Wide Web. Alongside with search engines, recommender systems and web personalization are seen as a remedy to this problem, since users are browsing the web according to their informational expectations while having a sort of implicit conceptual model in their mind. The latter is partially shared with other site visitors. In this paper we apply ontological modeling of anonymous ad-hoc web users’ behavior to improve online user action prediction for web personalization via recommendations.

15. Riina Maigre, Pavel Grigorenko, Hele-Mai Haav and Ahto Kalja. A SEMANTIC METHOD OF AUTOMATIC COMPOSITION OF E-GOVERNMENT SERVICES

It is hard to automatically find a semantically meaningful web service composition over a huge collection of web services available on the web. However, recent results in semantic web service research and technology could be effectively used within some specific domains. E-government is one of the sectors that need horizontal integration. Therefore, semantic web services and their composition become necessary and applicable in this domain. The paper proposes a semantic method of automatic composition of e-government services. It uses domain ontologies presented in OWL, semantic web services described in SAWSDL, quality of service (QoS) characteristics, ontology reasoning and AI planner in order to automatically provide service plans that could be presented in BPEL for execution. The approach is motivated by a case study from the domain of the Estonian state information systems.


Communication channel choice is the use by enterprises of one media channel compared to another (Reddick & Turner, 2012). Channel choice has been studied in media in the use and gratification literature (Kaye & Johnson), and the question whether old media are driven out of existence by new media or the importance of choosing right media for communication has been a concern in academic and industrial research (Nguyen & Western, 2006; Lengel & Daft, 1989; Vassilakis, Lepoureas Halatsis, 2007). Despite of fast increase in the use of e-government services there still exists a need of enterprises to contact with government via traditional channels. The literature
on why enterprises initiate contact with government through different communication channels has not got much attention.

The aim of current article is to identify the factors influencing the enterprises’ choice of communication channels with government comparing e-government to traditional service delivery channels such as the phone, mail, fax or visiting a government office. The study examines factors that explain the choice of channels according to the reasons for communication with government as well as depend on the characteristics of enterprises (e.g. sector, size, ownership, location, strategic choices). When focusing on the online portals of government institutions the impact of external factors influencing the use of e-government services will be analysed. In addition, the enterprises opinion about their satisfactory experience with public service delivery and benefits as well as problems connected with the use (or not use) of e-government services will be used to determine their impact to the choice of communication channels.

The main research questions are: 1) What factors explain enterprises’ choice of communication channels with government; 2) What factors could impact the increase of the use of e-government services.

The article, through logistic regression of enterprises’ opinion survey in Estonia and Germany is assessing the most commonly used communication channels depend on the nature of enterprises’ interaction between government and other characteristics of enterprises, and their experience with using e-government services. The results of analysis should show the reasons for using multiple channels for conducting with government, and whether there will be possibilities for increasing the use of e-government services in enterprises.

17. Evari Koppel and Raimundas Matulevicius. AN EVALUATION FRAMEWORK FOR SOFTWARE TEST MANAGEMENT TOOLS

Software testing has proven its value for software development increasingly over the last decade. With the recognition of the benefits of software testing, several software test management tools (TMT) have emerged on the market. Although there exist different approaches, there is no method for a systematic TMT assessment. This is a problem because to our knowledge, evaluating TMT is rather the subjective task, heavily depending on the evaluators’ opinions rather than based on the objective approach. The same problem applies when test managers are asked to evaluate whether their currently used TMT meets the company’s expectations. In this paper based on the survey performed among Estonian testing practitioners, we deliver a TMT evaluation framework. The paper applies structured approach by performing a literature study on software testing processes, existing TMT market research, and mapping together the identified test activities and test artifacts. The results help formulate and design the online questionnaire and perform a TMT survey in the Estonian IT companies. Based on this survey results, a framework for evaluating TMT software is created. Such a framework could potentially help companies to measure the TMT suitability to company’s goals and to decrease subjectivity of the TMT assessment. The framework also provides test and project managers the understanding whether their current TMTs meet the company’s expectations.
18. Edgars Diebelis and Janis Bicevskis. SOFTWARE SELF-TESTING

The Paper presents an overview of the results of 5 years of research in the field of self-testing. In 2007, self-testing was defined as one direction of smart technologies, a common idea of which is the desire to fit software with features of living beings: abilities to adapt to changing external environment, to optimise themselves and to defend themselves against threats. The purpose of self-testing is to provide a possibility to verify that the software is working correctly at any point of its life cycle. The research was carried out in several stages: at first, the concept and functionality of self-testing and its applicability in various software operating environments were defined; it was followed by implementing the self-testing functionality by integrating testing support options into the software developed. After that, the self-testing concept was compared against the possibilities offered by traditional testing support tools and implemented in an actual banking information system, and the efficiency of self-testing options was evaluated. The final conclusions drawn are: self-testing offers a number of advantages in achieving the software quality at comparatively low costs, at the same time ensuring the same functionality as provided by conventional testing support tools.

19. Guntis Arnicans, Dainis Romans and Uldis Straujums. SEMI-AUTOMATIC GENERATION OF A SOFTWARE TESTING LIGHTWEIGHT ONTOLOGY FROM A GLOSSARY BASED ON THE ONTO6 METHODOLOGY

We propose a methodology of semi-automatic obtaining of a lightweight ontology for software testing domain based on the glossary “Standard glossary of terms used in Software Testing” created by ISTQB. From the same glossary many ontologies might be developed depending on the strategy for extracting concepts, categorizing them, and determining hierarchical and some other relationships. Initially we use the ONTO6 methodology that allows identification of the most important aspects in the given domain. These identified aspects serve as the most general concepts in taxonomy (roots in the concept hierarchy). By applying natural language processing techniques and analyzing the discovered relations between concepts, an intermediate representation of lightweight ontology is created. Afterwards the lightweight ontology is exported to OWL format, stored in the ontology editor Protégé, and analyzed and refined by OWLGrEd – an UML style graphical editor for OWL that interoperates with Protégé. The obtained lightweight ontology might be useful for building up a heavyweight software testing ontology.

20. Stanislovas Norgėla, Julius Andrikonis and Arūnas Stočkus. QUALITATIVE REASONING ABOUT SPACE WITH HYBRID LOGIC

This article describes the way to employ hybrid logic H(\@,↓) in the analysis of qualitative spatial information. Moreover, it shows how the complexity of model checking algorithm is derived using the Kripke structure of qualitative spatial information and the query, which is presented as a formula of hybrid logic.
21. Laura Savičienė. **MODELING OPERATIONALIZATION OF NORMATIVE RULES IN DECISION SUPPORT FOR AIRCRAFT APPROACH/DEPARTURE**

This research is focused on norm operationalization in aeronautics domain. The investigated paradigm can be described as: from legal norms to technical rules in the artifact. Normative requirements (norms) for the aircraft trajectories are extracted from the flight rules and airport procedures. These norms are operationalized in a decision support system (DSS). An example of a normative rule: keep 3 degree descent angle while landing and hold restrictions of altitude and geography which is depicted in the approach chart. The decision support is based on evaluation of risk to violate the normative requirement. The following risks are modeled: trajectories' conformance with the flight rules, safe distance between aircraft, wake vortex separation and avoidance of dangerous substances in the atmosphere. The DSS is for the air traffic controller (not pilot) and must respond in real time. A DSS system provides surveillance, evaluates and recommends, whereas the human controller takes a decision.


Deploying new versions of server-side software is similar to deploying new versions of desktop software, however it is considered more complex and time consuming. Therefore, if new versions are released frequently and they need to be deployed to many servers, doing the work manually may lead to several problems - errors due to incorrect deployments, misconfigurations and considerable amount of time spent on routine tasks. This paper is a study of methods used for desktop software versioning in order to apply them to server-side software needs. The main focus was set on server-side software that is based on PHP and Oracle technology, however solutions where sought that could be used for other serverside technologies as well, e.g., ASP.NET, Java and Ruby. As a result, a solution was created and applied in a real-world scenario that helps handling server-side software versioning by automating builds of new versions, deployment and validation processes.

23. Rudolfs Bundulis and Guntis Arnicans. **ARCHITECTURAL AND TECHNOLOGICAL ISSUES IN THE FIELD OF BUILDING HIGH-RESOLUTION DISPLAY WALLS**

Currently there is a rising need to lay out a vastly growing amount of information and supersize working areas for collaboration and presentation needs. The hardware side is not able to catch up with the needs – display surfaces are still limited either in size or resolution and are not capable to offer a homogenous large scaled display with a high resolution to present the needed amount of information. This issue is tackled by constructing a multiple display wall that has a tiled display surface where the resolution is high enough since it sums up the individual resolutions of each tile. But as this solution is also limited to the number of video cards in the computer and their ability to feed multiple display targets and different, there are ongoing studies to understand how to cope with the current limitations on bandwidths by altering the architecture of the
solution. This paper summarizes the current limitations and cost-effectiveness of display wall environments and proposes ideas for alternate solutions.

24. Arturs Sprogis and Janis Barzdins. SPECIFICATION, CONFIGURATION AND IMPLEMENTATION OF DSL TOOL

A new specification method for DSL and DSL tools is proposed. The method is based on an advanced stereotype mechanism. A special feature of the proposed method is a precise definition of the extension mechanism for realization of non-standard features of DSL tools. In conclusions the architecture of a DSL tool building framework based on the proposed specification method is described.

25. Inga Zilinskiene and Saulius Preidys. A MODEL FOR PERSONALIZED SELECTION OF A LEARNING SCENARIO DEPENDING ON LEARNING STYLES

This paper deals with one of Technology Enhanced Learning (TEL) problems - the personalized selection of a learning scenario. Personalization is treated here as appropriateness of a learning scenario to preferences of a particular student, mainly, his/her learning style. This paper proposes an extended approach to modelling learning scenario selection based on preferences of a student's learning style. An ant colony optimization algorithm is modified and applied. In order to give a theoretical background the main conceptions of personalization, learning scenario and learning style are briefly presented. The aim of this paper is twofold. First, data mining technique to obtain a student's learning style is presented, second, a model for personalized selection of a learning scenario is proposed.

26. Oskars Rasnacs and Maris Vītins. AN INFORMATION SYSTEM TO LEARN CHARACTERISTIC SETS OF WORDS AND TO EXAMINE KNOWLEDGE IN STATISTICS

The authors have found that many students in the fields of health care and the social sciences, as well as practicing specialists, have problems when writing bachelor’s or master’s theses or other scholarly publications when it comes to taking decisions on the most appropriate data processing methods in their work. The authors have studied and analysed the theses and papers that have been produced, as well as the data processing methods that are indicated therein. Aspects of statistics are discussed in various areas of specialisation and in various courses. This means that students usually obtain a lot of information that is useful, but very hard to remember; they do not learn about schemes related to how the information can be brought to bear. This paper is based on the question of what students and practicing specialists must remember if they hope to find the necessary information from various sources (the Internet, the literature) to make independent decisions about the acceptance of appropriate data processing methods and about the implementation of those methods. The authors have found that there are many situations in the area of data processing which can be classified in different ways, and course instructors have divergent views about the most appropriate method for each situation. At the same time, each situation is in line with several sets of
characteristic words. Because software package management teams, assistance systems and educational literature are all usually in English, it is recommended that students learn the terminology in English irrespective of the language of the course which they are taking. The authors led a working group to design an information system in which each course instructor can implement a classification of data processing methods which is acceptable to him or her, also coming up with characteristic sets of words which are in line with the situation, as well as appropriate examples of data files. There is no denying that it would be more useful for students to work with data from real patients, but legal acts make that impossible. That is why the authors have addressed the issue of generating data on the basis of statistical indicators from scholarly publications. Students and specialists can use this information system in the educational and the test regime. In the education regime, the generated data files and corresponding sets of characteristic words can be examined. The test regime examines knowledge about the sets of characteristic words. The proposed information system has been tested in a traditional educational process at the university level, as well as in individual training sessions. Participants in the tests were tested and surveyed via a questionnaire. The results proved the effectiveness of the approach and the system.

27. Svetlana Kubilinskienė and Valentina Dagienė. METHODOLOGICAL DIGITAL RESOURCES: HOW WE CAN HELP EDUCATORS TO FIND THEM MORE EFFECTIVELY

The paper deals with digital resources in education and mainly focus on an approbation of the extended metadata model of digital learning resources. The model has been developed by covering methodological resources and learning method objects in order to increase their accessibility and usage in teaching process. The key purpose of methodological resources is to render conditions for teachers to share the professional experience, to spread methodological novelties, to help students and their parents to join the training and learning process more effectively. Different ways of choosing and combining learning methods obligate teachers first of all to know and estimate them, in line with the requirements posed to the contemporary school. Effective learning resource search and browsing possibilities can be realized only if standardized metadata are used. The metadata are the essential part of information infrastructure which is necessary while establishing order in internet chaos by using descriptions, classifications and structures which are helpful in creating more power and useful information repositories. At the moment the extended metadata model is implemented in the Lithuanian learning object metadata repository prototype. The paper focuses mainly on the results of an experimental approbation of the metadata model.