Fachhochschule Köln
Cologne University of Applied Sciences

Campus Gummersbach

SPT SEVEN

CONTENT

- VISION 3
- METHODS 4
  - 1 Define 4
  - 2 Data Acquisition and Measurement 4
  - 3 Modeling and Analysis 5
  - 4 Optimization 5
  - 5 Integration and Deployment 6
  - 6 Control 6
  - 7 Meta Evaluation 6
- PROJECTS 7
  - Applying Design of Experiments in Plastics Industry 7
  - Applying Genetic Programming for Steinmüller Engineering GmbH 8
  - Time-series Analysis for GreenPocket GmbH 9
  - Biogas Plant Optimization for GECO 10
- TEAM 11
  - Members 11
  - About SpotSeven 15
  - Publications / References 15

Picture credits: Title and background illustration: DrHitch; 7: Rainer Plendl; 8: Imaginis and Markus Kleine; 10: snapfoto105 and Andrei Merkulov; 13: Alexandr Mitiuc; 15: senticus – fotolia.com for all pictures.

VISION

Our mission is to provide scientific solutions to problems in industrial optimization and process analytics. We are passionate about technology, science, and critical reasoning. We are excited to promote modern data analysis and optimization technology into industry and improve life by empowering process control and product development. We are proud that every success in our consulting projects brings a strong new relationship with industrial partners, very often drives our research to new directions, and motivates our students to pursue education in computational intelligence.

The following, well-known allegory, originally written by Loren Eiseley, illustrates one key element of our approach.

An elderly gentleman had a habit of early morning walks on the beach. One day, after a storm, he looked down the beach and saw a human figure moving like a dancer. He smiled to himself at the thought of someone who would dance to the day. As he came closer he realized that it was a young woman and she was not dancing but was reaching down to the sand, picking up small objects and very gently throwing them into the ocean.

"Good morning! May I ask what it is that you are doing?"
The woman paused, looked up, and replied

"Throwing starfish into the ocean."
He asked,

"Why are you throwing starfish into the ocean?"

"The sun is up and the tide is going out. If I do not throw them in, they will die."

"Do you not realize that there are miles and miles of beach and there are starfish all along every mile? You cannot possibly make a difference."

The young woman listened politely, paused and then bent down, picked up another starfish and threw it into the sea, past the breaking waves, saying,

"It made a difference for that one." [Eise12a]

The starfish story reflects one essential aspect of our work: we focus on your individual task and work for your individual benefit. We work sequentially, in a stepwise manner, and pick up the urgent problems first, before we tackle the less important issues. Moreover, we do not claim to deliver a perfect solution in a single iteration, but we deliver measurable improvements in each step. We rely on interaction with and feedback from our clients, resulting in a sequential optimization process.

We apply cutting edge research results and continue exploring new scientific frontiers. Excellence in teaching, data sciences, optimization, and statistical analysis are standard for each member of our unique team. Close cooperation between industry and academia is a hallmark of our approach. We are well connected to the leading data science and optimization research groups in the world. Our goal is to discover solutions that integrate perfectly into your existing infrastructure. Our recent projects are focused on solving engineering problems in the domains: energy, water, steel, and plastic industry. We seek useable results that are of strategic value to your company. We develop and maintain individual solutions. We work independently, i.e., we do not sell one specific product or method. Customers get the best from several worlds.

Our team consists of passionate data detectives thrilled by knowledge discovery. SpotSeven members have experience from different projects and form a unique team, combining complementary skills. We are all different, have different backgrounds, have historically worked in different domains, but we all get a kick from having a problem solved.

The SpotSeven Process Model

SpotSeven defines a process model, which enables systematic optimization of complex real-world problems. It can be seen as an extension and adaptation of the well-known SixSigma procedure. Although steps of the SpotSeven process are clearly specified, they will not be used in an algorithmic manner. SpotSeven leaves room for individual procedures. The seven steps can be characterized as follows:

1 DEFINE

Customer and business requirements are specified from a high level point of view. The current situation, i.e., states and processes, is described in detail. Considering quality, time, and costs, the necessary steps are defined. Critical parameters and statistics are identified. The project team with clearly specified responsibilities is build.

As a result from the definition step, a tentative project charter with the following elements is available:

- Project description
- Project goals
- Benefits (value)
- Market analysis
- Project scope
- Responsibilities
- Start and end date

The project charter can be developed during a workshop, in which decision makers from the company participate. The following steps are planed with the SpotSeven team. In general, each of the following steps is initialized with a one-day team workshop.

2 DATA ACQUISITION AND MEASUREMENT

The second phase is devoted to the selection of suitable methods, instruments, and processes to collect data from the key process variables, which were defined during the first step. Data, which are already available, are collected, prepared, and integrated at the same time. This step comprehends also a systematic gathering of measurement units, frequencies, and accuracies. Important input-/output relations are defined. Based on the current value of the output variable the aimed value is specified.

3 MODELING AND ANALYSIS

Functional relationships between in-/output parameters are determined. Causes and effects are prioritized. SpotSeven uses a hierarchical approach, i.e., models with different complexity are used in parallel. High-quality open source software such as R, as well as commercial products such as JMP or Data Modeler, are used during the whole process. We distinguish process- and data analysis. Process analysis visualizes qualitative relationships, e.g., by integrating cause-effect diagrams.

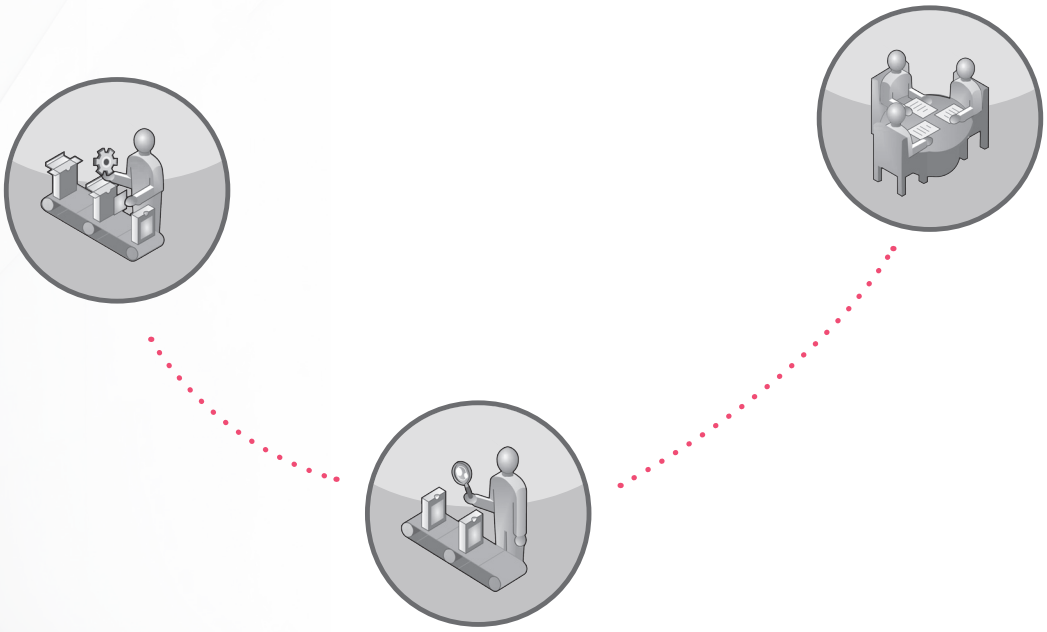
Data analysis is based on mathematical and statistical methods and relies on quantifiable measurement values. Methods from classical design of experiments are used to formulate statistical hypotheses about assumed functional relationships, e.g., $y = f(x_1, x_2...) + e$.

Statistical tests are applied to analyze these hypotheses. In addition to this classical approach from statistics, which define the baseline of our modeling and analysis, we use interactive visualizations [Cox10a], which represent an extended version of exploratory data analysis [Tuke77a]. With respect to the aimed value of the output variable, the current value is analyzed.

4 OPTIMIZATION

Sensitivity analysis is applied in order to determine suitable parameters for the optimization process. Based on this parameter set, which might include categorical, ordinal or numerical parameter values, an optimization technique has to be chosen. A broad variety of optimization methods is available. Besides the well-known classical approaches from mathematical optimization, SpotSeven team members have a strong background in computational intelligence. Evolutionary algorithms, namely evolution strategies and genetic programming, are valuable complements to standard approaches. Moreover, SpotSeven includes methods for multi-objective optimization, because many real-world problems have to handle several, conflicting objectives, e.g., minimization of production costs while maximizing product quality.

Our goal is to generate an understandable, simple solution. Interpretability is an important factor in the SpotSeven process. In many real-world settings, simple solutions are more robust than complex, high level solutions. Furthermore, simple solutions extrapolate better and protect against overfitting. We have developed several tools which allow an individual selection between exactness and robustness of the solution. After the model-based solution was determined and tested on the model hierarchy, a suitable solution is selected for implementation. A corresponding plan is developed and communicated with the practitioners.



5 INTEGRATION AND DEPLOYMENT

The optimal process parameters, which were determined during optimization on the model, are implemented in the real-world system. The process improvement is verified on the real system. If the real system does not allow these tests, model-based verification procedures will be used.

Operating and performance figures are defined in order to guarantee sustainability. The new settings are discussed with project leaders, technical experts and users to reach a high acceptance rate. If the new setting does not meet expectations, further analyses (step 3) might be useful. ●

6 CONTROL

A plan, which describes actions in time to keep the system under control, is developed. Methods from statistical quality control are tools of choice [Mont08a]. The documented solutions are delivered to the customer. Workshops and training courses can be organized to discuss results and to train practitioners. ●

7 META EVALUATION

SpotSeven itself contains a continuous improvement process. We ask for feedback from the customers and keep the contact alive, even after the project is finished. It is our passion to deliver the best results and we are interested to learn from the process. ●

Applying Design of Experiments for VOSS Automotive GmbH

This case study nicely illustrates a typical situation in which industry can directly benefit from academia – and the other way round. It is also an example that demonstrates how elementary Design of Experiment (DoE) techniques can yield good results in a very efficient manner. The company was facing severe quality issues from an injection molding process. Several million parts of this product were delivered to automobile manufacturers every year.

The project duration was 20 weeks. One student worked full time. Weekly meetings were held with project leaders, technical experts and users. Open Source Software, i.e., R, was used to perform the statistical analysis and to set up experimental designs. The process owner has basic statistical knowledge, the student acquired elementary expertise during the project. During brainstorming with process owners and technical experts the problem was defined. More than 20 parameters were considered as key-process variables. Parameters such as ambient temperature or air humidity were included in this setting.

During the "Data Acquisition and Measurement" phase, an experimental design with two stages was considered. Cause-effect diagrams played an important role during this stage of the project. The first experiments were performed to determine important parameters (screening). A second set of experiments were performed to determine an improved setting (optimization). The output variables and statistics were specified. Simple linear regression models and tree based models were used during the modeling and analysis phase in parallel. The resulting regression trees are easy to understand and supported results from the regression models. A fractional factorial design for the selected eight parameters was chosen during optimization. Response surface methodology was used as the driving method. Regression trees, response surface plots, interaction plots were used to visualize results. Further screening indicated that the behavior of the system can be explained basically with three parameters only.

During the "Integration and Deployment" phase, not the best configuration was selected to keep deterioration low. Settings with reduced pressure and velocity were used instead. The customer asks for a robust solution which performs slightly worse than the absolute best solution. Experiments were performed with settings which were found on the model. These experiments confirmed the results from the statistical analysis. An improved configuration was determined. During the "Control" phase, it was confirmed that the process was in a stable state. Further steps to keep the process stable were discussed.

Applying Genetic Programming for Steinmüller Engineering GmbH

Reduction of noxious emissions plays a major role in the planning and the operation of coal-fired power plants. Even small changes in the parameters for design and operation of such plants can reduce the emission of climate damaging gases like nitrogen oxides significantly. The central task of the project is to find a parameter setting optimizing the trade-off between efficiency of the power plant and noxious emissions.

To predict the emission profile, i.e., different concentrations of toxic gases in the emission, simulation models are required. Computational fluid dynamics methods offer highly accurate models that usually are used for this task. However, such models are very complex, hard to generate and the computational effort of these for the considered burning processes are extremely high. In contrast to computational fluid dynamics models, the projects aims for developing cost-sensitive, interpretable models by using symbolic regression via genetic programming.

Genetic programming belongs to the class of computational intelligence methods that mimic natural processes, e.g., natural evolution, to guide optimization processes. To this end, data-driven models are generated by genetic programming to adequately simulate the properties and behavior of the process under investigation, i.e., the burning processes in coal-fired power plants. This leads to insights into the processes themselves and a highly accurate simulation tool. This simulation tool is later on used for the optimization of the power plant's parameter setting by using multi-criteria evolutionary algorithms. ●

Time-series Analysis for GreenPocket GmbH

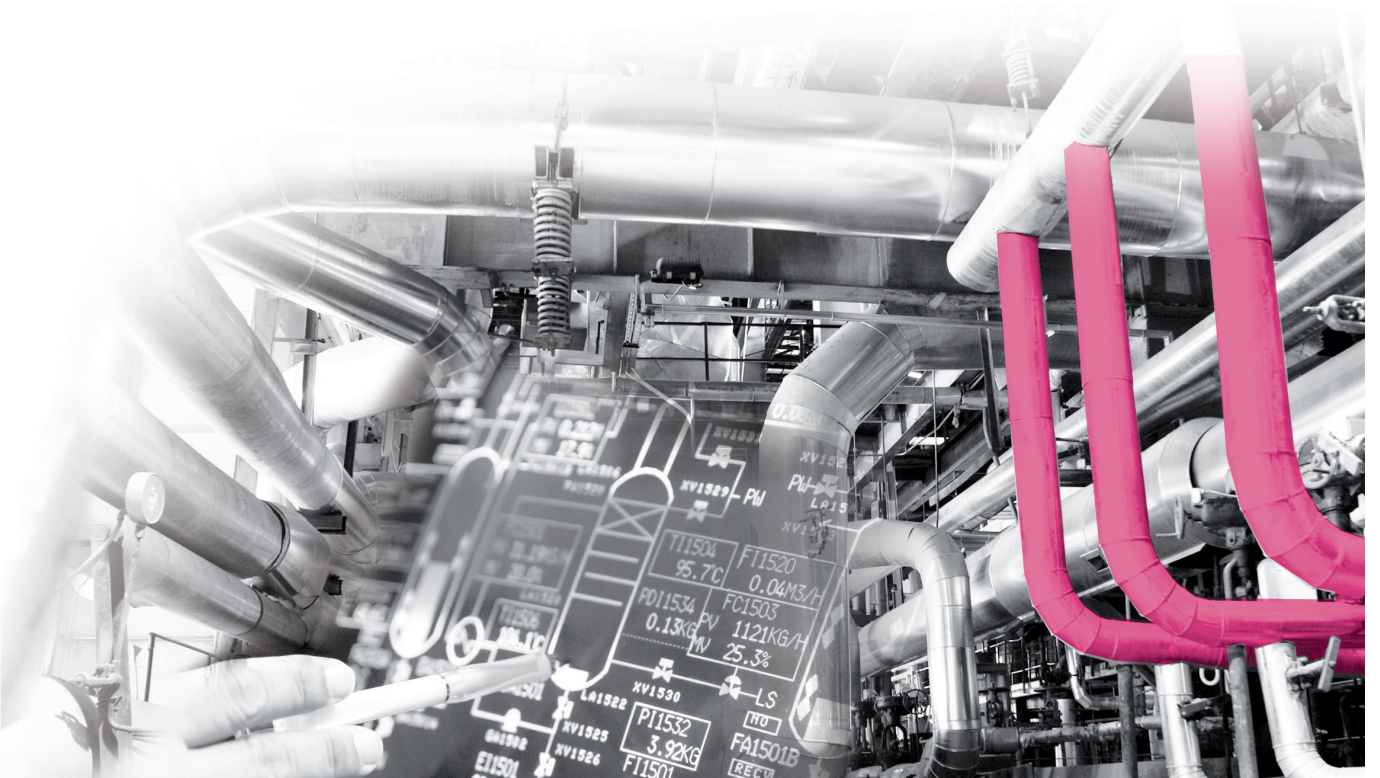
In times of accelerating climate change and rising energy costs, increasing energy efficiency becomes a high priority goal for businesses and private households alike. Smart metering equipment records energy consumption data in regular intervals multiple times an hour, reporting this data to a central system, usually located at a local public utility company. Here, consumption data from thousands businesses and households is correlated and analyzed to detect anomalies, discover patterns, and give important hints for improving energy efficiency.

In an ongoing project with GreenPocket GmbH, a pioneering provider of smart-metering software, the SpotSeven team provides state-of-the-art statistical methods for energy consumption time series analysis and forecasting. Challenges of this project include an exceptionally high data volume, necessitating methods and algorithms that are not only accurate, but also scalable and of acceptable computational complexity. Solutions have to integrate into GreenPocket's highly scalable computational infrastructure. In this project, SpotSeven chose a unique approach to discover a set of optimal solutions for GreenPocket's time series analysis problem by tapping into the joint creativity of the international research community. The problem was announced as a challenge. The GECCO Industrial Challenge 2012 is a competition featuring challenging real-world industry problems and is held at the Genetic and Evolutionary Computation Conference (GECCO), the largest and most renowned conference in the field of genetic and evolutionary computation. Participants of the Industrial Challenge will provide scalable time series predictions methods directly applicable by GreenPocket. On behalf of SpotSeven, the project is administered by two PhD students with support of one student assistant and has a scheduled runtime of six months.

Biogas Plant Optimization for GECO

In a collaboration with the Gummersbach Environmental Computing (GECO) group at CUAS, the substrate mixture infeed of a biogas plant was optimized. The GECO group provided a calibrated simulation model, representing a specific plant of an industrial partner. Additionally, the required background knowledge of the application was provided.

Previous approaches used classical approaches, i.e., first-principles models, for this optimization task. This work illustrates how statistical methods can complement the technical background provided by experts. The resulting application of state of the art methods with Matlab reduced the required number of model evaluations to one tenth and showed some interesting observation of the applied algorithms behavior. The project was tied to a diploma thesis, where one student spent three months, receiving regular assistance provided by supervisors of the SpotSeven Team and supporting members of GECO. Results from this diploma thesis were awarded with the 'Erzquell Förderpreis' for outstanding diploma theses.



Prof. Dr. Thomas Bartz-Beielstein

INNOVATOR OF THE SEQUENTIAL PARAMETER OPTIMIZATION TECHNOLOGY

Since 2006, when he became a professor for Applied Mathematics at CUAS, Thomas has built a research team of international status and visibility. He is head of the research center CIOP (Computational Intelligence, Optimization and Data Mining) at CUAS, and initiator of the SpotSeven team. His expertise lies in optimization, simulation, and statistical analysis of complex real-world problems. He is one of the leaders in the field of statistical analysis of optimization algorithms and the inventor and the driving force behind the sequential parameter optimization technology (SPOT).

Thomas has more than 100 publications on computational intelligence, optimization, simulation, and experimental research. His books on experimental research are considered as milestones in this emerging field. Thomas serves as program chair for all major conferences in computational intelligence as well as a reviewer for several scientific institutions, e.g., Dutch National Science Foundation, Natural Sciences and Engineering Research Council of Canada, Academy of Finland, and Federal Ministry of Education and Research Germany. In addition, Thomas serves as a reviewer for leading journals in the field of Computational Intelligence, Optimization, and Simulation.

Thomas studied Mathematics and Computer Science at the TU Dortmund University, and Mathematics and Educational Science at the Institute Waldorf Education Witten-Annen. He enjoys jogging and relaxing with his family at the seaside, when he is not busy at work.

Dr. Boris Naujoks

ONE OF THE LEADING SCIENTISTS ON MULTI-CRITERIA DECISION MAKING IN GERMANY

Boris received his PhD from the CI research group in Dortmund, where he focused on industrial applications of evolutionary algorithms at an early stage. Before joining the SpotSeven team in 2010, he was with Nutech Solution GmbH and log!n GmbH. There he managed different projects in applying evolutionary multi objective optimisation techniques in different real-world applications from airfoil design in aerospace industry to vehicle routing problems in logistics.

Dr. Katya Vladislavleva

EXPERT IN DATA-DRIVEN MODELING (PREDICTIVE ANALYTICS) FOR RESEARCH AND DEVELOPMENT

SpotSeven is very well connected to various groups and companies developing and practicing modern state-of-the art computational intelligence methods, and is driven to involve their expertise whenever necessary to make a project a success. A partnership with Evolved Analytics is one of such examples. Dr. Ekaterina (Katya) Vladislavleva, a chief data scientist and partner at Evolved Analytics, is an affiliated member at SpotSeven.

Katya's addition to the group's capability is Evolved Analytics' technology for the advanced function discovery and non-linear feature selection for process analytics and system design. Examples of projects where this technology was enabling successful solutions are the development of new modeling standards for melt-tension predictions of low-density polyethylenes, development and deployment of empirical models for rheological properties of polymers, non-linear feature selection for improved flavor design, customer segmentation based on quantitative product surveys, new standards in predicting mean opinion scores for video quality estimation.

Dipl.-Inform. Oliver Flasch

PHD STUDENT AT CUAS

Oliver's research interests include scalable methods for the automatic discovery of mathematical models from data (symbolic regression via genetic programming). Oliver studied Computer Science and Biology at TU Dortmund University and Ruhr-Universität Bochum. He is doing his PhD work in cooperation with TU Dortmund University. Before joining CUAS, he devised an innovative graphical stream processing language at DIP GmbH.

He is the CEO of the young startup sourcewerk, which provides flexible and scalable computing infrastructure for high performance and high throughput applications, including demanding simulation workloads. When not at work, he enjoys playing the piano in his office.

Dipl.-Inform. Martina Friese

PHD STUDENT AT CUAS

Martina is doing her PhD thesis in cooperation with Leiden University. Her research interests include real-world problems and modern optimization techniques such as sequential parameter optimization. Before Martina joined the team in 2009, she studied computer science at TU Dortmund University and worked as a software engineer at adesso AG.

Dipl.-Ing. Christian Jung

PHD STUDENT AT CUAS

Christian is doing his PhD studies in cooperation with TU Dortmund University. His research interests include process optimization and modeling of complex real-world problems in steel industry. Christian studied Electrical Engineering at TU Darmstadt and works at SMS Siemag AG. He joined the SpotSeven team in 2011.

M.Sc. Olaf Mersmann

PHD STUDENT AT TU DORTMUND UNIVERSITY

Olaf studied Physics, Computer Science, and Data Sciences at the TU Ilmenau, Philipps University of Marburg and the TU Dortmund. Before joining the SpotSeven Team in 2010, Olaf worked as an independent statistical consultant, was the CEO of the m3 Beratungsgesellschaft mbH. The research focuses on bridging the gap between classical statistics and black-box optimization. He is the recipient of a 3 year scholarship from the "Forschungsschule für Energieeffiziente Produktion und Logistik."

M.Sc. Beate Breiderhoff

MEMBER OF THE INSTITUTE OF COMPUTER SCIENCE AT CUAS

Beate is experienced in giving lectures in Applied Mathematics and computer programming and has a strong background in high level computer languages and computer algebra systems. In 2006, she received the M.Sc. degree from CUAS. Her research interests include optimization and numerical computation. Beate's work is motivated by the interaction between mathematicians, computational scientists, and engineers. She joined the SpotSeven team in 2011 and is currently developing optimization techniques for an environmental-technology and power-generating company.

Dipl.-Ing. (FH) Martin Zaefferer

MASTER STUDENT OF ENGINEERING IN AUTOMATION & IT
RESEARCH ASSISTANT IN THE SPOTSEVEN TEAM SINCE 2009

Martin studied electrical engineering with a focus on automation at CUAS and graduated in 2010. His research interests include computational intelligence, applications of knowledge discovery and sequential parameter optimization. Martin is an experienced programmer with a strong background in R and Matlab.

B. Eng. Jörg Stork

MASTER STUDENT OF ENGINEERING IN AUTOMATION & IT
RESEARCH ASSISTANT IN THE SPOTSEVEN TEAM SINCE 2010

Jörg studied electrical engineering with a focus on automation at CUAS and graduated in 2011. His research interests include genetic programming and parameter optimization. Jörg is an experienced programmer with a strong background in C and R.

About SpotSeven

The SpotSeven team consists of researchers working together under the guidance of Prof. Bartz-Beielstein. Three of our current projects are funded by the Federal Ministry of Education and Research (BMBF):

1. FIWA (Methoden der Computational Intelligence für Vorhersagemodelle in der Finanz- und Wasserwirtschaft, FKZ 17N2309)

2. MCIOP (Mehrkriterielle CI-basierte Optimierungsverfahren für den industriellen Einsatz, FKZ 17N0311)

3. CIMO (Computational Intelligence basierte Mehrzieloptimierungsverfahren, FKZ 17002X11)

SpotSeven was set up to build reliable relations between CUAS and partners from industry. Requests from industry initialized the team building. SpotSeven aims to intensify these cooperations and establish new industrial as well as research projects.

Publications / References (Books and WWW-Links)

SpotSeven team members have published over a hundred articles in peer-reviewed journals and conference proceedings. Most notably, Prof. Bartz-Beielstein has published two books:

BARTZ-BEIELSTEIN, T.: EXPERIMENTAL RESEARCH IN EVOLUTIONARY COMPUTATION - THE NEW EXPERIMENTALISM. SPRINGER, 2006.

BARTZ-BEIELSTEIN, T. (HRSG.); CHIARANDINI, M. (HRSG.); PAQUETE, L. (HRSG.); PREUSS, M. (HRSG.): EXPERIMENTAL METHODS FOR THE ANALYSIS OF OPTIMIZATION ALGORITHMS. SPRINGER, 2010.

[EISE12A] RETRIEVED FROM HTTP://EN.WIKIPEDIA.ORG/WIKI/THE_STAR_THROWER, FEB 14, 2012.
[TUKE77A] TUKEY, J. W.: EXPLORATORY DATA ANALYSIS. ADDISON-WESLEY, 1977.
[COX10A] COX, I.; GAUDARD, M. A.; RAMSEY, P. J.: VISUAL SIX SIGMA: MAKING DATA ANALYSIS LEAN. WILEY, 2010.
[MONT08A] MONTGOMERY, D. C.: STATISTICAL QUALITY CONTROL. WILEY, 2008.

Gefördert durch das:

Bundesministerium
für Bildung
und Forschung

Fachhochschule Köln
Cologne University of Applied Sciences

Campus Gummersbach

Address

Prof. Dr. Thomas Bartz-Beielstein
Fachhochschule Köln (CUAS)
Steinmüllerallee 1
51643 Gummersbach
02261 8196 6391
thomas.bartz-beielstein@fh-koeln.de
www.spotseven.de