



# LWDA 2016

Hasso Plattner Institute

September 12 - 14, 2016 | Potsdam

# Welcome

2 - LWDA 16

## **Welcome to LWDA Conference 2016 in Potsdam**

The LWDA 2016 conference provides a joint forum for experienced and young researchers, to bring insights to recent trends, technologies and applications and to promote interaction in the research field of Big Data and beyond. Recent research in the field is presented and discussed from the viewpoint of machine learning, data mining, knowledge extraction, knowledge management, information retrieval, personalization, database management, information systems, big data management and big data analytics to name a few.

The Hasso Plattner Institute (HPI) at the University of Potsdam is proud to host the LWDA 2016 conference. For the technical program the organizers would like to thank the workshop chairs and their programme committees for their hard work as well as the keynote speakers for their inspiring talks. Finally we acknowledge the generous support of our thematic partners Amazon and SAP SE as well as our sponsors Idealo and M2 consulting. We hope the participants will keep the venue as an inspiring event with fruitful discussions in mind and the readers will enjoy studying the scientific contributions in the LWDA 2016 Proceedings.

Prof. Dr. Emmanuel Müller  
Chair of Knowledge Discovery and Data Mining  
Hasso Plattner Institute (HPI) & German Research Centre for  
Geosciences (GFZ)

Dr. Ralf Krestel  
Junior Research Group Leader Web Science  
Hasso Plattner Institute (HPI)

# Content

3 - LWDA 16

Welcome	2
Progam	4 - 6
Partners	7 - 8
Keynote Speaker	9 - 11
Locations	12 - 13

# Program

4 - LWDA 16

## MONDAY, September 12, 2016

12:30 PM Welcome (HS 1)

01:00 PM Keynote (HS 1)

Argument Mining: Manual and automatic annotation of short user-generated texts  
by Prof. Dr. Stede

02:00 PM Joint Research Session (HS 1)

Harmony Assumptions: Extending Probability Theory for Information Retrieval  
Plackett-Luce Networks for Dyad Ranking  
Case Representation and Similarity Assessment in the selfBACK Decision Support System  
The Evolution of HyPer

03:30 PM Coffee Break

04:00 PM Keynote (HS 1)

Amazon: A Playground for Machine Learning  
by Dr. Archambeau

05:00 PM Parallel Sessions 1

KDML (HS 1)	DB (HS 2)	WM (H-E.51)	IR (H-E.52)
Synthesizing Invariants via Iterative Learning of Decision Treese	Robust Query Processing in Co-Processor-accelerated Databases	Understanding online financial communities: What constitutes a valuable information echange for users?	Classification of German Newspaper Comments
The Partial Weighted Set Cover Problem with Applications to Outlier Detection and Clustering	Visualization-Driven Data Aggregation	Knowledge Analytics For Workplace Learning	Modern Tools for Old Content – in Search of Named Entities in a Finnish OCR'd Historical Newspaper Collection 1771–1910
k-Means Clustering via the Frank-Wolfe Algorithm	SABER: Window-Based Hybrid Stream Processing for Heterogeneous Architectures	A Reference Model for Anti-Money Laundering in the Financial Sector	
	Challenges with Continuous Deployment of NoSQL-backed Database Applications	Experience - the neglected success factor in enterprises?	

06:30 PM Poster Sessions (Building H, Foyer) & BBQ

08:30 PM Closing Poster Session

## TUESDAY, September 13, 2016

09:00 AM Keynote (HS 1)

Algorithm Engineering for Graph Clustering  
by Prof. Dr. Wagner

10:00 AM Coffee Break

10:30 AM Parallel Sessions 2

KDML (HS 1)	DB (HS 2)	WM (H-E.51)	IR (H-E.52)
Probabilistic Inference with Stochastic Discrete Clenshaw-Curtis Quadrature	Discovering Data Transformations in Web Resources	CAPLAN: An Accessible, Flexible and Scalable Semantification Architecture	An Interactive e-Government Question Answering System
Variable Attention and Variable Noise: Forecasting User Activity	Exploring the Application Potential of Relational Web Tables	Building an integrated CBR-Big Data Architecture based on SEASALT	Graph n-grams for Scientific Workflow Similarity Search
Mining Subgroups with Exceptional Transition Behavior	Massively distributed BIG DATA management for Enterprises	BISHOP – Big Data Driven Self-Learning Support for High-performance Ontology Population	A System for Creating and Searching Semantic Documentation for Conservators
	A Clustering Approach for Holistic Link Discovery	Ontology-based Communication Architecture Within a Distributed Case-Based Retrieval System for Architectural Designs	

12:00 PM Lunch (Mensa Griebnitzsee)

01:00 PM Keynote (HS 1)

Intelligent Enterprise  
by Dr. Sebastian Wiczorek

02:00 PM Joint Applied Session (HS 1)

The digital transformation is just the beginning  
Case-Based Decision Support on Diagnosis and Maintenance in the Aircraft Domain  
Scalable Detection of Emerging Topics and Geo-spatial Events in Large Textual Streams  
Comparing contextual and non-contextual features in ANNs for movie rating prediction

03:30 PM Coffee Break

04:00 PM Community Meetings (KDML HS 1, DBHS 2, WM H.E.51, IR H.E.52)

05:15 PM Social Event (Boat departure at 5:30PM!) + Dinner on your own

# Program

6 - LWDA 16

## WEDNESDAY, September 14, 2016

09:00 AM Keynote (HS 1)

Web-Scale Domain-Specific Information Extraction by Prof. Dr. Leser
--

10:00 AM Coffee Break

10:30 AM Parallel Sessions 3

KDML (HS 1)	DB (HS 2)	WM (H-E.51)	IR (H-E.52)
Is Web Content a Good Proxy for Real-Life Interaction? A Case Study Considering Online and Off-line Interactions of Computer Scientists	Model-Driven Integration of Compression Algorithms in Column-Store Database Systems	Towards a case-based reasoning approach for cloud provisioning	Using key phrases as new queries in building
Min-Hashing for Probabilistic Frequent SubtreeFeature Spaces	Podiumsdiskussion: Data Sciences - Ein neuer Studiengang oder nur eine (Informatik-)Vorlesung?	Case Completion of Workflows for Process-Oriented Case-Based Reasoning	Vereinheitlichung internationaler Bibliothekskataloge
Sampling Methods for Random Subspace Domain Adaptation		SEMAFLEX - Semantic Integration of Flexible Workflow and Document Management	
		Ähnlichkeitsbasiertes Retrieval von BPMN-2.0-Modellen	

12:00 PM Lunch (Mensa Griebnitzsee)

01:00 PM Parallel Sessions 4

On the Evaluation of Outlier Detection: Measures, Datasets, and an Empirical Study Continued	From Cloud to Fog and Sunny Sensors	Understanding Mathematical Expressions: An Eye-Tracking Study (Resubmission)	
Scalable Inference in Dynamic Admixture Models	MapReduce Frameworks: Comparing Hadoop and HPCC	Linked Data City - Visualization of Linked Enterprise Data	
	Query-driven Data Integration	Approaches for Annotating Medical Documents	
	Finding Trees in Mountains -- Outlier Detection on Polygonal Chains	The Reevaluation of Subject Analysis: A Knowledge-based Approach facilitating Semantic Search	
	Towards rapidly developing database-supported machine learning applications	Topical Video-On-Demand Recommendations based on Event Detection	



## Hasso Plattner Institute

The Hasso Plattner Institute (HPI) in Potsdam holds an unique position on the landscape of German universities. The highly ranked Bachelor, Master and PhD degree programs in IT-Systems Engineering have a high practical and innovation orientation. HPI has no tuition fees and is financed through private funds donated by its founder, Professor Hasso Plattner, who co-founded the software giant SAP. Plattner is the most important private supporter of science in Germany and will contribute to further promote this unique university-level training in IT Systems Engineering.

One of HPI's major goals is to become an elite, world-class educational facility. The institute already offers an optimal study and work environment and cooperates very closely with the business community. HPI also maintains strong international contacts, for example with Stanford University in Palo Alto, California.

At HPI, approximately 500 highly gifted young students are currently pursuing their Bachelor's and Master's degrees in IT Systems Engineering with the help of twelve professors and over 200 senior researchers, lecturers and research assistants. Study programs are structured according to the needs of the IT industry, and students become acquainted not only with individual approaches to software technologies, but also with overall approaches to entire IT systems.

For more information visit [www.hpi.de](http://www.hpi.de)



## Gesellschaft für Informatik (GI)

German Informatics Society (Gesellschaft für Informatik e.V. (GI)) is a non-profit organization with about 20,000 members living across the world.

Most of them develop the discipline of informatics by teaching, researching or working in the field of informatics, while others are involved in related business and political areas. The main purposes of this network of professionals are motivate for informatics, develop the scientific discipline and promote the impact informatics has economy, business and society.

For more information visit [www.gi.de](http://www.gi.de)

# Partners

8 - LWDA 16

## Thematic Partners



## Sponsors







## **Prof. Dr. Manfred Stede**

Universität Potsdam

Manfred Stede studied Computer Science and Linguistics at TU Berlin and Edinburgh University, and received an M.Sc. in Computer Science from Purdue University (USA). In 1996, he earned his Ph.D. at the University of Toronto with a thesis on multilingual text generation. From 1995 to 2000, he worked at TU Berlin in the large national „Verbmobil“ project, which built a system for translating spoken language between German, English, and Japanese. After a short interlude at a company in Berlin, he became a professor in Applied Computational Linguistics at Potsdam University in 2001. His research mainly revolves around issues of text structure, ranging from theoretical models to its automatic analysis, with applications in, e.g., text mining and summarization. Recently, a focus of his research is on different dimensions of subjectivity in language, where speakers convey their attitudes, opinions, and arguments. The well-known computational application is Sentiment Analysis, where Stede contributed to a successful system implementing a lexicon-based approach for English. As a follow-up step, he is now interested in Argument Mining, i.e., the automatic discovery of author’s claims, reasons supporting them, and possible objections.



## **Cédric Archambeau**

Amazon

Cedric Archambeau is a Senior Machine Learning Scientist with Amazon, Berlin. He manages the algorithms team and served as a technical advisor to Sebastian Gunningham, Amazon Senior Vice President Seller Services. Recently, his team delivered the learning algorithms offered in Amazon Machine Learning ([aws.amazon.com/machine-learning](https://aws.amazon.com/machine-learning)). He is interested in large scale probabilistic inference and Bayesian optimization. He holds a visiting position in the Centre for Computational Statistics and Machine Learning at University College London. Prior to joining Amazon, he was leading the Machine Learning and Mechanism Design area at Xerox Research Centre Europe, Grenoble.

# Keynote Speaker

10 - LWDA 16



## **Prof. Dr. Dorothea Wagner**

Karlsruher Institut für Technologie

Dorothea Wagner is a full professor for Informatics at the Karlsruhe Institute of Technology (KIT). Her research interests include design and analysis of algorithms and algorithm engineering, graph algorithms, computational geometry and discrete optimization, particularly applied to transportation systems, energy systems, network analysis, data mining and visualization. Among other activities she is member of the German Council of Science and Humanities (Wissenschaftsrat). From 2007 to 2014 she was vice president of the DFG (Deutsche Forschungsgemeinschaft - German Research Foundation) and 2004 to 2013 speaker of the scientific advisory board of Dagstuhl - Leibniz Center for Informatics. In 2012 she received a Google Focused Research Award, she is member of Academia Europaea and Fellow of the GI (Gesellschaft für Informatik).



## **Dr. Sebastian Wieczorek**

SAP SE

Dr. Sebastian Wieczorek is Director at the SAP Innovation Center Network. In this role he is responsible for the development of SAP's Machine Learning Platform. He is also serving as an academic expert and reviewer for the European Commission and the German Ministry of Education and Research. In previous positions at SAP, Sebastian was coordinating all startup engagement activities of the SAP Innovation Center Network, managed EU-funded research projects, was product owner for SAP Web Analytics, team lead for the „Application Engineering Group“ at SAP Research and futurist for the Development Experience team, working from Germany, Israel and the US. Besides working for SAP, Sebastian held lectures on Web Engineering at the Technical University of Darmstadt and obtained a PhD from the Technical University of Berlin (Germany) under supervision of Prof. Ina Schieferdecker. Before joining SAP, Sebastian worked as a Software Developer in Moscow (Russia) and studied computer sciences at the Technical University of Dresden (Germany) and the Northumbria University at Newcastle (United Kingdom).



## **Prof. Dr. Ulf Leser**

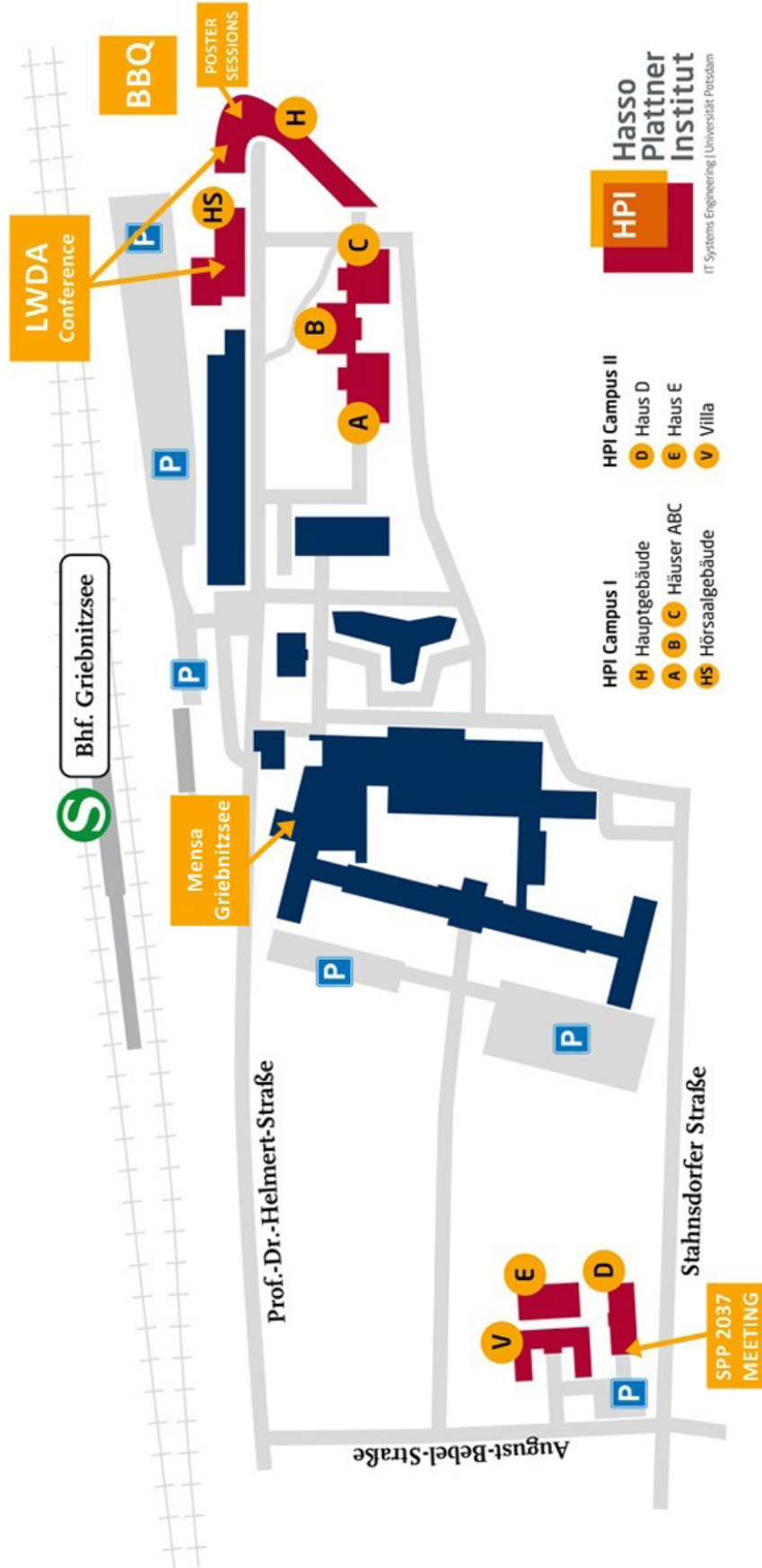
Humboldt-Universität zu Berlin

Ulf Leser studied computer science at the Technische Universität München and obtained his PhD in Data Integration and Query Planning from Technische Universität Berlin. After positions in research institutes and in the private sector, in 2002 he became a professor for Knowledge Management in Bioinformatics at Humboldt-Universität zu Berlin. His research focuses on scientific data management, statistical Bioinformatics, biomedical text mining and infrastructures for large-scale biomedical analysis and is typically carried out in interdisciplinary projects with domain scientists, especially from Medicine and Biology. He is speaker of the DFG-funded graduate school „SOAMED - Service-oriented architectures for medical applications“, chairman of the coordinated BMBF project „PREDICT - Comprehensive Data Integration for Personalized Ontology“, PI of the DFG research unit Stratosphere, and a board member of the DFG-excellence graduate school „BSIO - Berlin School for Integrative Oncology“.

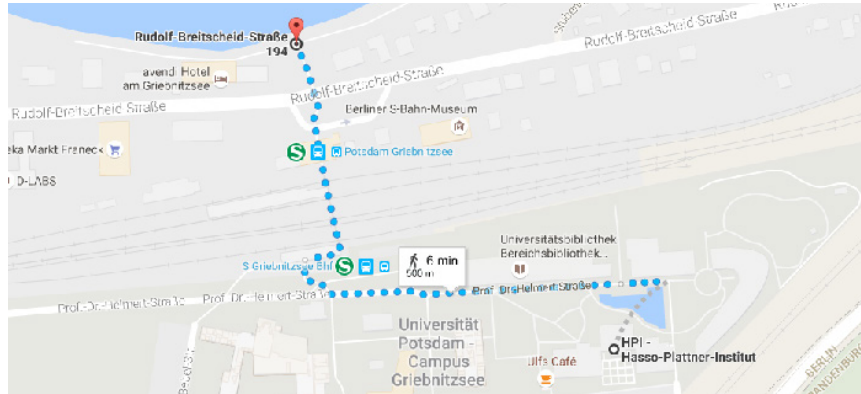
# Locations

12 - LWDA 16

## Campus Map

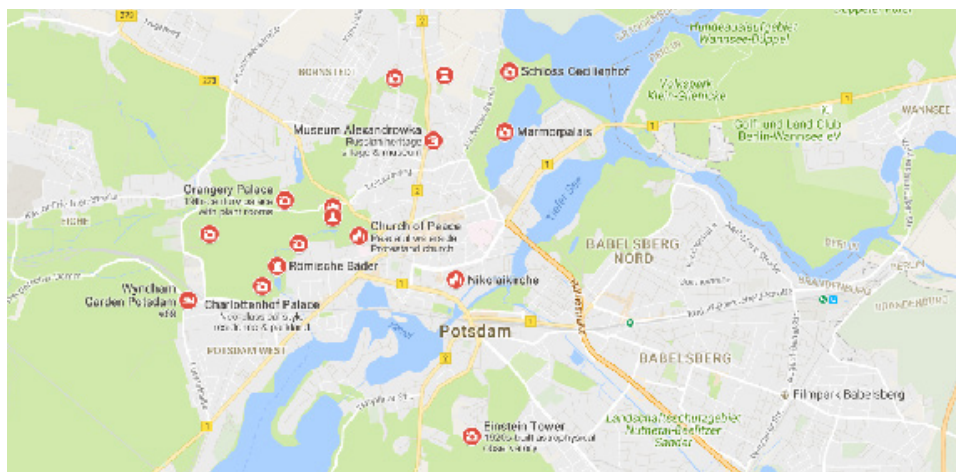


## Social Event



<b>Boat Trip</b> included in conference fee	<b>Conference Dinner</b> at your own expences
<b>Tuesday, September 13th</b>	
<b>05:15 PM</b>	<b>07:30 PM</b>
Boat stop Griebnitzsee, Rudolf-Breitscheid-Straße 194, 14482 Potsdam	Meierei - Brauhaus, Im Neuen Garten 10, 14469 Potsdam

## Potsdam - Points of Interest



**LWDA 2016**

Prof. Dr. Emmanuel Müller, Hasso Plattner Institute (General Co-Chair)  
Dr. Ralf Krestel, Hasso Plattner Institute (General Co-Chair)  
Dr. Davide Mottin, Hasso Plattner Institute (Proceeding Chair)  
Diana Stephan, Hasso Plattner Institute (Local Organization)

**FG-DB 2016 (Program Chairs)**

Prof. Dr. Felix Naumann, Hasso Plattner Institute  
Dr. Matthias Uflacker, Hasso Plattner Institute

**FG-IR 2016 (Program Chairs)**

Dr. Claus-Peter Klas, GESIS - Leibniz Institute for Social Sciences, Germany  
Dr. Ingo Frommholz, Luton University UK

**FG-KDML 2016 (Program Chairs)**

Dr. Stephan Günemann, Technische Universität München  
Prof. Dr. habil Ansgar Scherp, ZBW für Wirtschaftswissenschaften  
& Christian-Albrechts-Universität zu Kiel

**FG-WM 2016 (Program Chairs)**

Prof. Dr. Klaus-Dieter Althoff, Universität Hildesheim  
Sebastian Furth, Denkbare GmbH