SwissText
SwissText
SwissText
SwissText
SwissText
SwissText SwissText SwissText
SwissText SwissText SwissText
SwissText SwissText SwissText
SwissText SwissText SwissText
SwissText SwissText SwissText
SwissText
SwissText
SwissText
SwissText
SwissText

SwissText 2018
3rd Swiss Text Analytics Conference

Program

12th and 13th of June 2018

Content

Preface

Welcome to the 3rd Swiss Text Analytics Conference!

Artificial Intelligence is booming! AI and Machine Learning classes are overcrowded, startups are founded everywhere, and not a single day goes by without breaking news from the field. The same applies to Text Analytics. Tremendous progress is made every year: for instance, machine translation has been shifted to a new level by DeepL, Microsoft has reached human level in Speech2Text, and Google just announced its autonomous calling agent Duplex.

These light-house projects also raise interest in Text Analytics in general, and both industrial and academic experts perceive increased need for NLP applications and solutions.

Fittingly with this development, the SwissText conference will grow this year in several directions: two days instead of one, more presentations, three interactive workshops, full scientific papers, an international program committee with 50 experts, published proceedings, etc.

But would people really appreciate the new format? Would they come to Winterthur for two days? As of now, I can say "yes": number of registrations is high, and more than 80% of the participants will come for both days!

On the other hand, each additional "feature" of the conference increased the workload for our team, and I would like to thank the local organizing committee here at ZHAW for their support and great work! In addition, I would like to thank all members of our program committee for their efforts, and of course our academic and industrial partners and sponsors for their support.

I am now looking forward to an exciting conference with fascinating presentations and lots of interesting discussions!

Mark Cieliebak

Conference Chair

Schedule
Tuesday - Day One, 12th June 2018

08:30	Registration + Coffee/Gipfeli
09:00	Welcome Message: Mark Cieliebak (ZHAW)
09:30	Keynote: **Disentangling the Thoughts: Latest News in Computational Argumentation.** Iryna Gurevych (TU Darmstadt)
10:15	Break
10:45	Parallel Sessions

Track 1 *Room: TN E0.46*	Track 2 *Room: TN E0.54*	Workshop *Room: TS 01.07*
Ruiqi Li, Guillem Collell and Marie-Francine Moens: **A** Genetic Algorithm for Combining Visual and Textual Embeddings Evaluated on Attribute Recognition	*Rémy Blättler, Mark Cieliebak and Jan Deriu:* **B**ilingual Term Extraction with Big Data	**D**ata Expedition into the Swiss Twitter Community Organizer: *SpinningBytes AG* Part 1 of 2
Georgios Balikas and Ioannis Partalas: **O**n the Effectiveness of Feature Set Augmentation using Clusters of Word Embeddings	*Massimo Lusetti, Tatyana Ruzsics, Anne Göhring, Tanja Samardzic and Elisabeth Stark:* **E**ncoder-Decoder Methods for Text Normalization	
Diego Saldana: **A**utomated Detection of Adverse Drug Reactions in the Biomedical Literature Using Convolutional Neural Networks and Biomedical Word Embeddings	*Ann-Sophie Gnehm:* **T**ext Zoning for Job Advertisements with Bidirectional LSTMs	

12:15	Lunch Break	
13:45	Poster Presentations	**D**ata Expedition into the Swiss Twitter Community Organizer: *SpinningBytes AG* Part 2 of 2
14:15	Poster Session and Exhibition Open Time to visit the Exhibition and Posters	
16:00	Keynote: **Cross-Linguality and Machine Translation without Bilingual Data.** Eneko Agirre (UPV/EHU)	
16:45	End of Day One	

Schedule

Wednesday - Day Two, 13th June 2018

08:30	Registration + Coffee/Gipfeli
09:00	Welcome Message: Mark Cieliebak (ZHAW)
09:30	Keynote: **Corpus Conversion: A Machine Learning Platform to Ingest Documents at Scale.** Costas Bekas (IBM)
10:15	Break
10:45	Parallel Sessions

Track 3 *Room: TN E0.46*	Track 4 *Room: TN E0.54*	Track 5 *Room: TN E0.58*
Noëmi Aepli and Simon Clematide: **P**arsing Approaches for Swiss German	*Ethan Brown and Saurabh Jain:* **R**ecommendations from Unstructured Data for Investment Banking	*Sibylle Peuker:* "**W**hat can I ask you?" - The User Experience of Conversational Interfaces
Donat Agosti: **T**ext and Data Mining Workflow to Make Scientific Publications Accessible	*Albert Weichselbraun, Philipp Kuntschik, Norman Süsstrunk, Fabian Odoni, Sandro Hörler and Adrian M.P. Brasoveanu:* **O**ptimizing Information Acquisition and Decision Making Processes with Natural Language Processing, Machine Learning and Visual Analytics	*Khalil Mrini, Marc Laperrouza and Pierre Dillenbourg:* **B**uilding a Question-Answering Chatbot using Forum Data in the Semantic Space
Markus S.T. Pilzecker: **A**utomatic Hyperlinking on a Juridicial Corpus via an RDFized Terminology	*Lorenz Bernauer and Patrick Bernauer:* **A**uto-generated Email Responses: Boost your Email Efficiency	*Jacky Casas, Nathan Quinteiro Quintas, Elena Mugellini and Omar Abou Khaled:* **R**upert the Nutritionist, the Efficient Conversational Agent

12:15	Lunch Break
13:30	Poster Presentations

13:45	Hate Speech Hackathon Organizer: *FHNW and ZHAW* Room: *TN 02.29*	Workshop: NLP Clinic Organizer: *Swiss Alliance for Data-intensive Services* Room: *TS 03.08*	Workshop: Future Actions for Swiss German Organizers: *University of Zurich and ZHAW* Room: *TS 01.07*

15:30	Poster Session and Exhibition Open Time to visit the Exhibition and Posters
16:00	Keynote: **Multilingual Event Detection in the Europe Media Monitor.** Hristo Tanev (ECJRC)
16:45	Closing Statement
17:15	Apéro

Iryna Gurevych

Disentangling the Thoughts: Latest News in Computational Argumentation

Abstract. Argumentation is omnipresent in our daily communication and is a crucial part of each decision-making process. The rapidly growing research field of Argumentation Mining aims at automatically recognizing argumentation structures in unstructured text in order to establish new intelligent systems for facilitating information access, writing skills acquisition and text summarization. In the first part of this talk, I will focus on large-scale argument search, classification and reasoning. In the second part, the focus will be on mitigating high annotation costs for argument annotation. Specifically, we tackle small-data scenarios for novel argument tasks, less-resourced languages or web-scale argument analysis tasks such as detecting fallacies. The talk presents the various solutions in Computational Argumentation developed at the Technical University of Darmstadt: Argumentation Analysis for the Web (*ArguAna*), Decision Support by Means of Automatically Extracting Natural Language Arguments from Big Data (*ArgumenText*).

Biography. Iryna Gurevych is Professor for Language Processing at Technical University of Darmstadt, and Director of the Ubiquitous Knowledge Processing Lab. Her research aims at lexical-semantics, resources and algorithms, and innovative applications of language processing to social sciences and humanities.

Organization.   Technical University of Darmstadt

Contact. gurevych@ukp.informatik.tu-darmstadt.de

Eneko Agirre

Cross-Linguality and Machine Translation without Bilingual Data

Abstract. Machine translation is one of the most successful text processing application. Current state-of-the-art systems leverage large amounts of translated text to learn how to translate, but is it possible to translate between two languages without having any bilingual data? In this presentation we will show that this is indeed the case. We will first map the word embedding spaces of two languages to each other, with and without seed bilingual dictionaries. This allows to produce accurate bilingual dictionaries based on monolingual corpora alone. Based on these mappings, it is then possible to train a neural machine translation system without accessing any bilingual data.

Biography. Eneko Agirre is Professor at the University of the Basque Country and member of the IXA Natural Language Processing group. His research focuses on lexical and computational semantics, with applications in information retrieval and machine translation. He has been president of SIGLEX, member of the editorial board of Computational Linguistics, and has received a Google research award.

Organization.   University of the Basque Country (UPV/EHU)

Contact. e.agirre@ehu.eus

Costas Bekas

Corpus Conversion: A Machine Learning Platform to Ingest Documents at Scale.

Abstract. Over the past few decades, the amount of scientific articles and technical literature has increased exponentially in size. Consequently, there is a great need for systems that can ingest these documents at scale and make the contained knowledge discoverable. Unfortunately, both the format of these documents (e.g. the PDF format or bitmap images) as well as the presentation of the data (e.g. complex tables) make the extraction of qualitative and quantitative data extremely challenging. In this paper, we present a modular, cloud-based platform to ingest documents at scale. This platform, called the Corpus Conversion Service (CCS), implements a pipeline which allows users to parse and annotate documents (i.e. collect ground-truth), train machine-learning classification algorithms and ultimately convert any type of PDF or bitmap-documents to a structured content representation format. We will show that each of the modules is scalable due to an asynchronous microservice architecture and can therefore handle massive amounts of documents. Furthermore, we will show that our capability to gather ground-truth is accelerated by machine-learning algorithms by at least one order of magnitude. This allows us to both gather large amounts of ground-truth in very little time and obtain very good precision/recall metrics in the range of 99% with regard to content conversion to structured output.

Biography. Costas Bekas manages the Foundations of Cognitive Computing group at IBM Research – Zurich. Costas' research covers high-performance computing, massive scale analytics and cognitive computing. He is a recipient of the PRACE Award (2012) and the ACM Gordon Bell Prize (2013 & 2015).

Organization.   IBM Research – Zurich

Contact. bek@zurich.ibm.com

Hristo Tanev

Multilingual Event Detection in the Europe Media Monitor

Abstract. Europe Media Monitor (EMM) performs large-scale media analysis in up to seventy languages and recognizes various types of trends, some of them combining information from news articles and social media posts. Information Extraction is an integral part of EMM, both as enabling technology, as well as a tool for document analysis. The core information extraction applications in EMM perform two tasks: named entity recognition and crisis-related event detection and metadata extraction. In this talk I will describe our multilingual event metadata detection and extraction system and its applications in the domains of security, disasters and health. Also, I will present a recently released corpus with event metadata information.

Our event extraction technology is used in three main applications: NEXUS, a system for detection of reports about security-related events, including crime, armed conflicts, man-made and natural disasters, a Medical Event Extraction system, whose primary functionality is detection of disease outbreak reports, and a system for detection of news about trans-border violations. In the talk we will explain the structure of the underlying grammars, the heuristics used in these systems, as well as the learning and the maintenance of the lexical resources.

Biography. Hristo Tanev, PhD, has been working in the domain of Natural Language Processing for more than 25 years already. His work has been concentrated mostly in the area of Question Answering, Terminology Extraction, Grammar Induction, and Event Extraction. Being researcher and developer of working applications, Hristo Tanev's work resulted in several software systems for Question Answering, Event Extraction and Terminology Learning.

Organization.   European Commission, Joint Research Centre

Contact. hristo.tanev@ec.europa.eu

Presentation

Ruiqi Li, Guillem Collell and Marie-Francine Moens

A Genetic Algorithm for Combining Visual and Textual Embeddings Evaluated on Attribute Recognition

Abstract. We propose a genetic-based algorithm for combining visual and textual embeddings in a compact representation that captures fine-grain semantic knowledge – or attributes – of concepts. The genetic algorithm is able to select the most relevant representation components from the individual visual and textual embeddings when learning the representations, simulating how complementary visual and linguistic knowledge is combined. We evaluate the proposed model in an attribute recognition task and compare the results with a model that concatenates the two embeddings and models that only use monomodal embeddings.

Biography. Ruiqi Li is a PhD researcher at the Language Intelligence and Information Retrieval group in KU Leuven. Her research topics are word representation, sentence embedding, and graph structural-context similarity. She is experienced in machine learning, especially deep learning techniques for natural language processing.

Organization.   Katholieke Universiteit Leuven

Contact. ruiqi.li1993@outlook.com, guillem.collelltalleda@cs.kuleuven.be, sien.moens@cs.kuleuven.be

Georgios Balikas and Ioannis Partalas

On the Effectiveness of Feature Set Augmentation using Clusters of Word Embeddings

Abstract. Word clusters have been empirically shown to offer important performance improvements on various Natural Language Processing (NLP) tasks. Despite their importance, their incorporation in the standard pipeline of feature engineering relies more on a trial-and-error procedure where one evaluates several hyper-parameters, like the number of clusters to be used. In order to better understand the role of such features in NLP tasks we perform a systematic empirical evaluation on three tasks, that of named entity recognition, fine grained sentiment classification and fine-grained sentiment quantification.

Biography. Georgios Balikas studied Electrical and Computer Engineering at the Aristotle University of Thessaloniki, Greece. He holds a PhD in Machine Learning from the University of Grenoble-Alps and works as a Data Scientist at Kelkoo Group. His research topics are in the intersection of Natural Language Processing (NLP), Information Retrieval (IR) and Machine Learning with a special focus on multilingual text mining applications. He is experienced in neural networks and machine learning for NLP and IR tasks.

Ioannis Partalas works as machine learning scientist in Expedia group in various projects on content and revenue management. He was previously working as data scientist in the Viseo group (France) on NLP projects and prior to that he was a post-doctoral researcher in AMA team at University of Grenoble working on large-scale classification.

Organization. Kelkoo, Expedia

Contact. georgios.balikas@kelkoogroup.com, ipartalas@expedia.com

Massimo Lusetti, Tatyana Ruzsics, Anne Göhring, Tanja Samardzic and Elisabeth Stark

Encoder-Decoder Methods for Text Normalization

Abstract. Text normalization is the task of mapping non-canonical language, typical of speech transcription and computer mediated communication (CMC), to a standardized writing. For example, Swiss German words viel, viil, vill and viu all map to the normalized form viel. It is an important up-stream task, necessary to enable the subsequent direct employment of standard natural language processing (NLP) tools for text mining applications and indispensable for languages such as Swiss German that have no written standard. Text normalization has been addressed with a variety of methods, most successfully with character-level statistical machine translation (CSMT). In the meantime, machine translation has changed and the new methods, known as neural encoder-decoder (ED) models, resulted in remarkable improvements. Text normalization, however, has not yet followed. A number of neural methods have been tried, but CSMT remains the state of the art. In this work, we normalize Swiss German WhatsApp messages using the ED framework. We exploit the flexibility of this framework, which allows us to learn from the same training data in different ways. In particular, we modify the decoding stage of a plain ED model to include target-side language models operating at different levels of granularity: characters and words. Our systematic comparison shows that our approach results in an improvement over the CSMT state-of-the-art.

Biography. Massimo Lusetti is a Master's degree student in Multilingual Text Analysis at the University of Zurich, where he currently works as a graduate assistant. His main interests are machine translation, machine learning and neural networks. He received his Bachelor's degree at the University of Bologna.

Tatyana Ruzsics is a doctoral student at Language and Space Lab (Text Group), University of Zurich. She received her MSc degree in Mathematics from Lomonosov Moscow State University and MSc degree in Quantitative Finance from UZH/ETH. Her research interest lies in deep learning and its application to language processing. She participated in a team that won CoNLL 2017 shared task on morphological reinflection.

Organization. University of Zurich

Contact. massimo.lusetti@uzh.ch, tatiana.ruzsics@uzh.ch

Diego Saldana

Automated Detection of Adverse Drug Reactions in the Biomedical Literature Using Convolutional Neural Networks and Biomedical Word Embeddings

Abstract. Monitoring the biomedical literature for cases of Adverse Drug Reactions (ADRs) is a critically important and time-consuming task in pharmacovigilance. The development of computer assisted approaches to aid this process in different forms has been the subject of many recent works.

One particular area that has shown promise is the use of Deep Neural Networks, in particular, Convolutional Neural Networks (CNNs), for the detection of ADR relevant sentences. Using token-level convolutions and general-purpose word embeddings, this architecture has shown good performance relative to more traditional models as well as Long Short-Term Memory (LSTM) models.

In this work, we evaluate and compare two different CNN architectures using the ADE corpus. In addition, we show that by de-duplicating the ADR relevant sentences, we can greatly reduce over-optimism in the classification results. Finally, we evaluate the use of word embeddings specifically developed for biomedical text and show that they lead to a better performance in this task.

Biography. Diego Saldana Miranda obtained his Ph.D. at Universite Pierre et Maria Curie (UPMC) Paris applying machine learning methods to predict properties of biofuel mixtures. After this, he joined Novartis to work on applying machine learning in various domains including clinical studies and real-world data analysis, as well as natural language processing.

Organization. Novartis Pharma A.G.

Contact. diego.saldana_miranda@novartis.com

Ann-Sophie Gnehm

Text Zoning for Job Advertisements with Bidirectional LSTMs

Abstract. We present an approach to text zoning for job advertisements with bidirectional LSTMs (BiLSTMs). Text zoning refers to segmenting job ads into eight classes that differ from each other regarding content. It aims at capturing text parts dedicated to particular subjects (e.g. the publishing company, qualifications of the person wanted, or the application procedure) and hence facilitates subsequent information extraction. As we have 38,000 job ads in German available, published in Switzerland from 1950 to 2014 (Swiss Job Market Monitor corpus), each labelled with text zones, we benefit from a large amount of training data for supervised machine learning. We use BiLSTMs, a class of recurrent neural networks particularly suited for sequence labelling. Our best model reaches a token-level accuracy of 89.8%, which is 2 percentage points above results from previous approaches with CRFs and implies an error rate reduction by 16%. Models with task-specific embeddings perform better than models with pretrained word embeddings, due to the large amount of labelled training data. When optimizing the model for future application on recently published job ads, the inclusion of older training data lowers performance, as some sort of out-of-domain effect counteracts the effect of more training data. Ensembling, i.e. to aggregate classification decisions of five models, brings the largest improvement of all optimization steps, raising accuracy by 0.5 percentage points. In conclusion, we succeeded in building a high performing solution to automatic text zoning for job ads based on neural networks.

Biography. Ann-Sophie Gnehm works as a research assistant at the University of Zurich, on the Project Swiss Job Market Monitor. She holds a M.A. in Sociology and Economics and is currently studying Computational Linguistics & Language Technology at the University of Zurich as second degree. Her research interests are information extraction and text mining for the Social Sciences.

Organization.   University of Zurich

Contact. gnehm@soziologie.uzh.ch

Presentation

Rémy Blättler, Jan Deriu and Mark Cieliebak
Bilingual Term Extraction with Big Data

Abstract. We are building an open-source system that automatically extracts a translation memory and a terminology list from a multilingual website.

This works by first scanning the whole website with a JavaScript capable spider and then links the related websites to each other. From those linked websites, a TM is created by running an alignment algorithm that uses Machine Translation (MT) to find related sentences (Example: Bleualign https://github.com/rsennrich/Bleualign). From the TM, special words and phrases are extracted into a terminology list. In this talk, we first present the overall architecture of the system showing the current challenges. Then, we present the terminology extraction algorithm based on corpus statistics.

Biography. Rémy Blättler holds a Master of Computer Engineering degree from Northwestern University in Chicago. He lived and worked in Manhattan for eight years. Since 2006, he has been developing the Supertext system as CTO.

Jan Deriu received his Master degree at ETH, and was part of the team that won the SemEval2016 task for Sentiment Analysis. His main interests lie in Natural Language Processing and Deep Learning. He is pursuing a PhD on dialogue systems at University of Zurich and ZHAW.

Organization.   Supertext, Zurich University of Applied Sciences, SpinningBytes

Contact. remy@supertext.ch, deri@zhaw.ch, ciel@zhaw.ch

Ethan Brown and Saurabh Jain

Recommendations from Unstructured Data for Investment Banking

Abstract. Senior bankers at investment banking firms spend much of their time manually scrolling news feeds for potential deals. Here we suggest a way to augment and automate this process leveraging natural language understanding tools. More specifically, we use a combination of entity extraction and classification to surface events that are relevant to a particular investment banker. For the entity extraction step, we have a domain-trained word embedding feed into a bi-directional recurrent neural network and end in a single-layer classifier which projects into ~10 phrase classes in the investment banking domain. Next, an additional extraction is performed on the classified phrase to pull out relevant details, e.g. company name, deal size, region, etc. Following this, we have two routes to provide recommendations. In the first, each extracted phrase is scored by a simple regressor, with some of the extracted details used as input features. These scores are then aggregated by one of the extracted details (not used in the input), e.g. company name. The second approach is to simply predict details missing from the phrase extraction, and use these predictions as a recommendation. To tune both underlying models, the banker is able to provide both explicit and implicit feedback. This approach has been successfully deployed at several banks around the world.

Biography. Dr. Ethan Brown received his PhD in computational physics from the University of Illinois before coming to ETH Zürich for a postdoc research position studying quantum optimization and machine learning. He joined Squirro after having his own UK-based company using machine learning to control data center infrastructure.

Saurabh Jain received his Master's degree in Computer Science from EPFL, Lausanne. Before that, he finished his Bachelors in Electronics & Communication Engineering from IIIT-Allahabad (India). He is currently working as a Machine Learning Engineer at Squirro.

Organization.   Squirro

Contact. ethan@squirro.com, saurabh@squirro.com

Presentation

Sibylle Peuker
"What can I ask you?" - The User Experience of Conversational Interfaces

Abstract. Conversational Interfaces are considered the new heroes of human-machine interaction. Smart assistants like Alexa from Amazon and Google Home are catching up in more and more households, and there's hardly any industry that does not experiment with chatbots. But can humans and machines really understand each other? We wanted to know if such conversational interfaces are as easy to use as is claimed. For this purpose, we have investigated many chatbots and other conversational interfaces within a year, we observed many people (individuals and groups) interacting with such interfaces, and we also built chatbots.

To substantiate the hypotheses built during this year, we tested the user experience and acceptance of language assistants and chatbots in an exploratory study with 16 users. In in-depth interviews, we saw that think big is okay, but start small is the way to success. In the talk, we discuss what we have learned from user research:

• What are the hurdles in the communication between man and machine?
• For which tasks is a conversational interface suitable?
• When is it useful?
• When should the chatbot rather hand over to a human employee?
• How do I design a conversation so that it feels natural and meaningful to the user?

We back up our user research with 17 years of experience with user behavior in usability tests. We will discuss several examples to show the potential and the pros and cons of conversational interfaces.

Biography. Dr. Sibylle Peuker, is partner and UX architect at Zeix, the agency for User-Centered Design and User Experience in Zurich. She is a mathematician and computer scientist by training. With curiosity and passion she creates user-friendly websites, web apps and chatbots for all industries. Sibylle always tries to infect others with her enthusiasm for user-centered design, as a speaker and as a lecturer at Berner Fachhochschule and Hochschule für Technik Rapperswil.

Organization. Zeix AG

Contact. sibylle.peuker@zeix.com

Presentation

Jacky Casas, Nathan Quinteiro Quintas, Elena Mugellini and Omar Abou Khaled

Rupert the Nutritionist, the Efficient Conversational Agent

Abstract. Rupert is a conversational agent, or chatbot, whose goal is to coach people who wish to improve their food lifestyle. Two choices are available to the users: help them reduce their consumption of meat or consume more fruits and vegetables. The data gathering method is easy and especially fast in order to simplify the life of the user. The follow-up is done daily with a more comprehensive summary at the end of each week.

The purpose of this chatbot is to demonstrate that an interaction via a chat interface can be fast and efficient, and has nothing to envy to existing coaching mobile applications.

The daily interaction duration is up to the user, from few seconds to several minutes. Users who want to know more about different topics related to food can start a discussion with the agent. A guided conversation is then launched and users will learn interesting facts like meat consumption in Switzerland, meat "budget" and other tips and statistics about food in general. Since the target audience is Switzerland, the information relates to this country. The chatbot currently speaks French, but will also speak German and Italian in the future. A coach must be friendly and speak the language of the user. Rupert is developed to run on Facebook Messenger.

Further tests, aiming at analyzing user's behaviour on this conversational agent, will be conducted.

Biography. Jacky Casas is currently a PhD student in computer science at the University of Fribourg while Nathan Quinteiro is doing his master thesis at EPFL. Both are interested in artificial intelligence, machine learning and natural language processing. This led them to the chatbots world. And here they are.

Organization. University of Applied Sciences of Western Switzerland, Ecole Polytechnique Fédérale de Lausanne

Contact. jacky.casas@hefr.ch, nathan.quinteiro@epfl.ch, elena.mugellini@hefr.ch, omar.aboukhaled@hes-so.ch

Presentation

Khalil Mrini, Marc Laperrouza and Pierre Dillenbourg
Building a Question-Answering Chatbot Using Forum Data in the Semantic Space

Abstract. We attempt to combine both conversational agents and conceptual semantic representations by creating a Chatbot based on an online autism forum, and aimed at answering questions of parents of autistic children.

First, we collected the titles and posts of all threads in two forums. We filter the threads based on their titles, such that only threads which titles are questions are kept. We remove threads without replies and obtain 36,000 threads, totalling 600,000 replies.

Then, to provide the best answers, we use Amazon Mechanical Turk to obtain usefulness labels on five levels for a part of the data set: about 10'000 replies. We train a variety of models to learn from these labels and apply them on the unlabelled replies. We use seven standardized continuous features, with three features on sent2vec cosine similarity. The Random Forest Classifier came on top with an F1-score of 0.66.

Afterwards, we compute the sentence vectors of questions and replies by averaging word2vec embeddings. When the Chatbot is asked a question, its sentence representation is computed and compared to all forum questions. The replies of the most cosine-similar question are first filtered to keep the ones with the highest usefulness label, and then the most cosine-similar reply is returned as answer.

An example of how the Chatbot works is its answer to "What is Autism?": "Autism has always been difficult for some people to explain, but I do know what it is not: Pretty colors and sparkly gems".

Biography. Khalil Mrini is a Master's student in Computer Science at EPFL. His research interests are in natural language processing, and more specifically representation learning and its uses. He has worked on research projects at the Idiap Research Institute, at Infosys (Bangalore, India) and at Nanyang Technological University (Singapore), and is now at Swisscom's research wing at EPFL. He is due to start a PhD in NLP at the University of California, San Diego.

Organization.   Ecole Polytechnique Fédérale de Lausanne

Contact. khalil.mrini@epfl.ch, marc.laperrouza@epfl.ch, pierre.dillenbourg@epfl.ch

Lorenz Bernauer and Patrick Bernauer

Auto-Generated Email Responses: Boost Your Email Efficiency

Abstract. Emails are a time-consuming and poorly automated task within today's business world. Repetitive emails can be under-challenging and frustrating for highly skilled workforce. This pain is addressed by Mailytica, a new service for Email Smart Responses based on Natural Language Processing and Artificial Intelligence.

Mailytica analyzes emails by applying Text Analysis and Machine Learning pipelines in order to derive topics without human input. All new incoming emails are constantly examined and referred to existing topics. Subsequently, Mailytica does generate emails responses automatically on basis of answers which were given in the past to the same topic. This especially automates emails of repetitive business transactions.

Mailytica significantly differs from comparable systems:

1. The training is done by itself. No manual training needed. This is achieved by classifying a dataset of old emails and by analysing how the user responded to these emails.
2. Deployment of most up-to-date NLP and AI algorithms. Mailytica does understand not only single words, but also the real meaning and context of all sentences and messages.
3. Self-learning algorithms. Ensuring that the solution is advancing constantly.
4. Incorporation of different functionalities like classification, routing, analytics of emails and auto-generated Smart Responses.

Based on a real-world case study, we demonstrate how Mailytica works, how it integrates and how users benefit from it.

Biography. Lorenz Bernauer graduated at the Yonsei University in Seoul. His research and publications aim the improvement of Natural Language Processing and Artificial Intelligence algorithms. He is now engaged in the development of production-ready applications which utilize both, NLP and AI. The most recent achievement is the co-foundation of Mailytica, an Email Chatbot which aims to automate repetitive email tasks in enterprises.

Organization.   Mailytica

Contact. lorenz.bernauer@mailytica.com

Presentation

Albert Weichselbraun, Philipp Kuntschik, Norman Süsstrunk, Fabian Odoni, Sandro Hörler and Adrian Brasoveanu

Optimizing Information Acquisition and Decision Making Processes with Natural Language Processing, Machine Learning and Visual Analytics

Abstract. The Web, social media and textual resources within organizations contain a growing wealth of data that is potentially useful to improve an organization's products, services and business processes. Key technologies such as natural language processing, machine learning and visual analytics provide means for acquiring data from external and internal sources, extracting information, and performing analytics that help in obtaining insights in regard to specific industries, companies, products and services. Applying these methods to real-world use cases, therefore, has the potential to play a pivotal role in improving an organization's competitiveness. Nevertheless, combining data sources, technology and domain knowledge to use cases that unfold the economic potential of these methods is a challenging task.

The following presentation, therefore, introduces three industry research projects from the domains of (i) recruiting, (ii) company valuation, and (iii) digital asset management to demonstrate how natural language processing, machine learning and visual analytics can be used to improve the efficiency of information gathering processes, enable data-driven decision making and improve customer value. We will discuss data sources (Web, Deep Web and internal data) relevant to each use cases, methods for processing these data, their adaptation to the domain and integration in the company's business processes.

Biography. Albert Weichselbraun is Professor of Information Science at the University of Applied Sciences Chur and co-founder and technical director of the webLyzard technology GmbH, a commercial spin-off focusing on Web Intelligence.

He guides the development of the webLyzard Web Intelligence platform and is involved in national and European projects. His research focuses on the fields of Natural Language Processing, Artificial Intelligence and commercial applications of the developed methods.

Organization.   University of Applied Sciences HTW Chur

Contact. albert.weichselbrauch@htwchur.ch, philipp.kuntschik@htwchur.ch, norman.suesstrunk@htwchur.ch, fabian.odoni@htwchur.ch, sandro.hoerler@htwchur.ch, adrian.brasoveanu@htwchur.ch

Presentation

Noëmi Aepli and Simon Clematide
Parsing Approaches for Swiss German

Abstract. This paper presents different approaches towards universal dependency parsing for Swiss German. Dealing with dialects is a challenging task in Natural Language Processing because of the huge linguistic variability, which is partly due to the lack of standard spelling rules. Building a statistical parser requires expensive resources which are only available for a few dozen high-resourced languages. In order to overcome the low-resource problem for dialects, approaches to cross-lingual learning are exploited. We apply different cross-lingual parsing strategies to Swiss German, making use of the Standard German resources. The methods applied are annotation projection and the model transfer. The results show around 60% Labelled Attachment Score for all approaches and provide a first substantial step towards Swiss German dependency parsing. The resources are available for further research on NLP applications for Swiss German dialects.

Biography. Noëmi Aepli holds a MA in Computational Linguistics (CL) and Language Technology from the University of Zurich (UZH) and currently works as a Data Scientist at Swisscom. She has been working on Natural Language Processing for Swiss German since her BA studies.

Simon Clematide studied German Literature and Linguistics, Informatics and Philosophy at UZH. He holds a PhD in CL and works as a Senior Researcher and Lecturer at the Institute of CL at UZH. His research topic is automatic text analysis.

Organization.   University of Zurich

Contact. noemi.aepli@uzh.ch, siclemat@cl.uzh.ch

Presentation

Markus S.T. Pilzecker

Automatic Hyperlinking on a Juridicial Corpus via an RDFized Terminology

Abstract. In collaboration with the terminology section of the Bundeskanzlei, the Bundesarchiv Bern, in its role as national coordination body for Open Data, commissioned a project, whose intention it was to demonstrate the usefulness of Linked Data technologies on open terminologic data. The two principal parts of this project were:
- "TERMDAT-LD", the RDFification of the juridical subdomain of TERMDAT, an official Swiss multilingual terminology
- Automatic Hyperlinking on the "Systematische Sammlung des Bundesrechts (SR)" and TERMDAT itself
Automatic Hyperlinking is a technology, where a machine enriches an otherwise unaltered web document with hyperlinks to other web addresses. In contrast to a human, a machine places such hyperlinks, following a strategy, incarnated in its implementation. Juridicial documents are rich and comparably precise in refering other juridicial documents. But even in the carefully, manually edited SR, only few of such references are made explicit as hyperlinks, placed by a human. And this is due to a simple fact: cost. We decided to combine Automatic Hyperlinking with Linked Data technologies and, instead of chosing as linking strategy a direct link to the authoritative web page, we went over the web surface page of the TERMDAT-LD entry of the title of the legal document as a pivot point. Since the pivot page exposes the whole RDF graph structure of the terminology, for the price of "some clicks more", it allows to study many more juridicial workflow scenaria than just direct links. 30000 terms and phrases (DE, EN, FR, IT, RM) triggered construction of a bit less than a million new links in about 120k documents. Precision was better than 99%. Due to limitations in budget, we could not determine recall. A first inspection showed, that it is considerably below 50%, since our primitive algorithm is not very robust against syntactical noise, e.g. HTML markup in longer phrases.

Biography. Markus Pilzecker has a degree in theoretical physics from the Technical University of Darmstadt. In his career as software engineer and architect, a cornerstone was Model Driven Architecture, which brought ontological and epistemological perspectives into IT. His interest in languages led to a stay at the DFKI, where he built multilingual NLP applications. Having been terminology manager at Roche, he is now as Knowledge Architect and Engineer founder of the Swiss SME "Die Wissens-Ingenieure".

Organization.   Die Wissens-Ingenieure

Contact. swisstext2018@die.wissens.in.gen.ieu.re

Presentation

Donat Agosti, Guido Sautter and Reto Gmür
Text and Data Mining Workflow to Make Scientific Publications Accessible

Abstract. Scientific publications ought to contribute to the dissemination of research results. Whilst the quintessence of current scientific publications lays in creating largely unstructured natural language publications from often highly structured primary data, the Swiss company Plazi does the opposite, i.e., creates structured, findable, citable, and reusable data from natural language scientific publications. This workflow is based on standalone open source applications written in plain Java, with modular Natural Language Processing tools. These tools can be chained into highly customizable pipelines, allowing TDM to be highly automated, including discovery of anything ranging from named entities to entire specific text blocks, or figures and their associated captions, and deposit them both in TreatmentBank (TB) and in the Biodiversity Literature Repository (BLR). Because data is not work in a legal sense, and thus is not copyrighted, and Swiss law allows to make temporary copies of publications for mining, the Plazi workflow can operate legally under Swiss law – a competitive advantage for Switzerland.

TreatmentBank includes text-based entities such as 220,000 taxonomic treatments extracted from 31,000 articles, including millions of extracted facts, close to 1M bibliographic references. BLR, a community within CERN's Zenodo repository, includes 172,000 extracted scientific illustrations and 32,000 article deposits, with a daily input. Each of the deposits includes meta data, links to related items and a digital object identifier (DataCite DOI). Upon upload, this data is immediately submitted to some of the world's largest science infrastructure, such as the NCBI and the Global Biodiversity Information Facility, and names fed into the Swiss Linked Open Government Portal LINDAS.

Biography. Donat Agosti is a biologist with a PhD degree from ETHZ (1989); research posts at the Natural History Museum London, American Museum of Natural History and the Jet Propulsion Laboratory, California Institute of Technology. Since 2008 founding president of Plazi, a Swiss company promoting open access to scientific results. Involved since 1996 in fostering open access to biodiversity data, data conversion workflows to build open data repositories, such as the Biodiversity Literature Repository at Zenodo. Member of Swiss University Council SUK P-5 expert panel.

Organization.   Plazi, Factmission

Contact. agosti@plazi.org, gsautter@googlemail.com, reto@factsmission.com

Posters

A Morphological Toolset for Rumantsch Grischun

Author	Simon Clematide
Organization	University of Zurich
Contact	siclemat@cl.uzh.ch

A Supervised Dataset for Stylometric Analysis of Swiss Text

Author	Alireza Ghasemi
Organization	ELCA Informatique SA
Contact	alireza.ghasemi@elca.ch

Automated Transformation of Documents into Data Sources

Author	Aaron Richiger, Martin Keller, Patrick Emmisberger
Organization	turicode AG
Contact	aaron.richiger@turicode.com, martin.keller@turicode.com, patrick.emmisberger@turicode.com

BioMedical Text Mining Activities of the BioMeXT group

Author	Fabio Rinaldi
Organization	University of Zurich
Contact	fabio@ontogene.org

Cutter - a Universal Multilingual Tokenizer

Author	Johannes Graën, Mara Bertamini, Martin Volk
Organization	University of Zurich
Contact	maratiziana.bertamini@uzh.ch, graen@cl.uzh.ch, volk@cl.uzh.ch

Differences between Swiss High German and German High German via data-driven methods

Author	Gerold Schneider
Organization	University of Zurich
Contact	gschneid@ifi.uzh.ch

Enhancing Search with Facets, Categories and Clusters Extracted From Free Text

Author	Alexandros Paramythis, Doris Paramythis
Organization	Contexity AG
Contact	alexandros.paramythis@contexity.ch, doris.paramythis@contexity.ch

Evaluating Neural Sequence Models for Splitting (Swiss) German Compounds

Author	Don Tuggener
Organization	Zurich University of Applied Sciences (ZHAW)
Contact	tuge@zhaw.ch

History of Patents in Switzerland

Author	Arnaud Miribel, Yu Yamashita, Soufiane El Badraoui
Organization	Ecole Polytechnique Fédérale de Lausanne
Contact	yu.yamashita@epfl.ch, arnaudmiribel@gmail.com, soufiane.elbadraoui@epfl.ch

Is Reading Mirrored in the Face? A Comparison of Linguistic Parameters and Emotional Facial Expressions

Author	Egon Werlen, Ivan Moser, Christof Imhof, Per Bergamin
Organization	Institute for Research in Open-, Distance- and eLearning (IFeL), Swiss Distance University of Applied Sciences (FFHS)
Contact	egon.werlen@ffhs.ch, ivan.moser@ffhs.ch, christof.imhof@ffhs.ch, per.bergamin@ffhs.ch

Merging Haystacks to Find Matching Needles: A Transliteration Approach

Author	Fernando Benites, Gilbert Duivesteijn, Pius von Däniken, Mark Cieliebak
Organization	Zurich University of Applied Sciences, Deep Impact, Spinningbytes AG
Contact	benf@zhaw.ch, ciel@zhaw.ch, pvd@spinningbytes.com, gilbert@deep-impact.ch

Mining Goodreads: A Text Similarity Based Method to Measure Reader Absorption

Author	Simone Rebora, Moniek Kuijpers, Piroska Lendvai
Organization	University of Göttingen, University of Basel, University of Verona
Contact	simone.rebora@univr.it, moniek.kuijpers@unibas.ch, piroska.lendvai@uni-goettingen.de

NOAH 3.0: Recent Improvements in a Part-of-Speech Tagged Corpus for Swiss German Dialects

Author	Noëmi Aepli, Nora Hollenstein, Simon Clematide
Organization	Swisscom, University of Zurich, IBM
Contact	noemi.aepli@uzh.ch, nora.hollenstein@uzh.ch, siclemat@cl.uzh.ch

Quantifying Collaboration in Synchronous Document Editing

Author	Adrian Pace, Louis Baligand, Stian Håklev, Jennifer Olsen, Nore de Grez, Bram De Wever
Organization	Ecole Polytechnique Fédérale de Lausanne, Ghent University
Contact	adrian.pace@epfl.ch, louis.baligand@epfl.ch, stian.haklev@epfl.ch, jennifer.olsen@epfl.ch, Nore.DeGrez@UGent.be, Bram.DeWever@UGent.be

SlowSoft Speech Technology Components for Swiss German and Romansh

Author	Christof Traber
Organization	SlowSoft GmbH
Contact	christof.traber@slowsoft.ch

Spitch - Technology Stack & Applications

Author	Igor Nozhov, Juerg Schleier
Organization	Spitch AG
Contact	igor@spitch.ch, juerg.schleier@spitch.ch

Swiss German Language Detection in Online Resources

Author	Pius von Däniken, Mark Cieliebak
Organization	SpinningBytes AG
Contact	pvd@spinningbytes.com, mc@spinningbytes.com

Swiss Twitter Corpus

Author	Christoforos Nalmpantis, Fernando Benites, Michaela Hnizda, Daniel Kriech, Pius von Däniken, Ralf Grubenmann, Mark Cieliebak
Organization	Zurich University of Applied Sciences, SpinningBytes AG
Contact	christofernal@gmail.com, benf@zhaw.ch, pvd@spinningbytes.com, rg@spinningbytes.com, krih@zhaw.ch, mhnizda@gmail.com, ciel@zhaw.ch

Text-based Indexation and Monitoring of Corporate Reputation Using the R Package "sentometrics"

Author	Samuel Borms, David Ardia, Keven Bluteau, Kris Boudt
Organization	Université de Neuchâtel, Vrije Universiteit Brussel
Contact	david.ardia@unine.ch, keven.bluteau@unine.ch, kris.boudt@vub.be, samuel.borms@unine.ch

The surprising Utility of NLP Algorithms on non Text Data

Author	Ruben Wolff
Organization	D-ONE
Contact	ruben.wolff@d1-solutions.com

Would you like to start an innovation project?

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Innosuisse – Swiss Innovation Agency

Innosuisse is the Swiss Innovation Agency. We promote science-based innovation via funding, networks and advice.

Most of our funding is put towards supporting innovation projects carried out jointly by companies and research institutions (CHF 151.4m in 2017). SMEs can investigate the feasibility of their ideas with the innovation cheque scheme. Innosuisse also funds high-risk projects with a high degree of innovation run by researchers who do not yet have a partner to implement the results of their project.

Visit www.innosuisse.ch for more information and follow us on social media to stay up to date.

www.linkedin.com/company/Innosuisse
www.twitter.com/_Innosuisse

Sponsor

IBM **Research**

IBM Research – Zurich is one IBM's 12 research centers around the globe.

This network of some 3000 scientists is one of the largest industrial IT research organizations in the world. The Zurich laboratory was established in 1956 and is home to world-class scientists representing more than 45 nationalities. Cutting-edge research and outstanding scientific achievements—most notably two Nobel Prizes—are associated with this lab. The spectrum of research activities range from exploratory research in nanoscience and –technology and quantum computing for future computing, to cloud and computing infrastructure, security and privacy, computational sciences, Blockchain, data analytics and artificial intelligence (AI). A focus is the application of AI to solve industry-specific innovation challenges. So called cognitive systems are able to learn at scale, reason with purpose and interact with humans naturally. Rather than being explicitly programmed, they learn and reason from their interactions with us and from their experiences with their environment. Cognitive systems can make sense of huge amounts of data, including the 80 percent of the world's data that computer scientists call "unstructured", and help us discover new insights to make better decisions.

www.zurich.ibm.com

Contact:

Claudia Longchamp
HR Business Partner
Phone: +41 724 87 47
E-Mail: lao@zurich.ibm.com

DATA DRIVES DIGITIZATION

Memory Prices
1 Giga • 250 000 €

IOT
Machine Learning
Big Data / Smart Data

* Moore's law
* MS Windows
* Google
* Facebook
* iPhone
* Self Driving

Data growth

• 15 €   • 0,1 €

| 50s | 60s | 70s | 80s | 90s | 00 | 10 | 20 | 30 |

COMPUTER:

per company   per desk   pocket   per anything
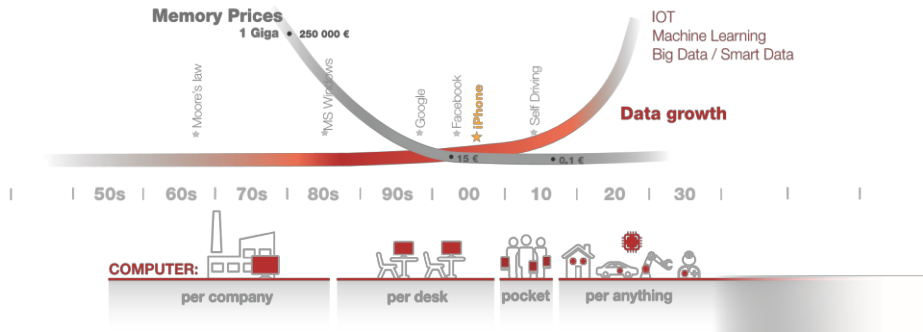
D|ONE
WE MAKE SENSE.

Artificial intelligence. And human excellence.

Accelerating paradigm shifts in the financial industry by leveraging new technologies.

The Swiss Start up Parashift is specialized in Software Development and applied Machine Learning. We are providing new solutions and custom implementations to complex business problems in the financial industry and accounting.

Our current services revolve around automated information extraction and document processing.

Further Information

www.parashift.io

Contact

alain.veuve@parashift.io

die **Mobiliar**

Die Mobiliar is Switzerland's oldest private insurer, and the market leader in several major insurance segments. Every third household and every fourth company in Switzerland is insured with us.

Our presence in over 160 locations throughout the country cements our reputation as Switzerland's most personal insurance company.

The Data Science & Analytics team at Die Mobiliar works on solving problems centered around insurance data. We combine our expertise on machine learning, text mining, natural language processing, image analysis, and recommender systems with Die Mobiliar's nearly 200-year-old expertise on the insurance business.

The result: analytics solutions that optimise processes and support decision-making for Die Mobiliar and the wider insurance industry.
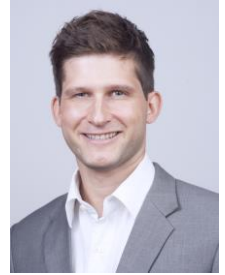
Dr. Mirco Rossi
Head of Cognitive Computing & Disruptive Analytics
mirco.rossi@mobi.ch

Schweizerische Mobiliar Versicherungsgesellschaft AG

Bundesgasse 35 · 3001 Bern · Switzerland · +41 (0)31 389 61 11

www.mobiliar.ch · diemobiliar@mobiliar.ch

Partner

The Swiss Alliance for Data-Intensive Services is a technology network for innovative companies, academic institutes and individuals with a focus on data-driven value creation: services, products and business models based on digital data. It is a community that helps companies to move forward with digitalization and brings key innovators together.

In doing so, we rely on three pillars for our success:

- R&D projects by our members for pushing forward innovation, and cooperation within an interdisciplinary network of experts from innovative companies and universities to combine knowledge from different fields into marketable products and services.

- Top employees and best-in-class education.

- Inspiration and exchange via connecting domain experts and joint workshops, conferences and Expert Groups such as Machine Learning Clinic, Natural Language Processing and Predictive Maintenance.

To boost innovation several initiatives are launched in 2018, e.g., the initiative "From ideas to projects" and a start-up grant 2018/2019. The Swiss Alliance for Data-Intensive Services makes a significant contribution in creating data-driven added value in Switzerland.

Get in touch:

More information: data-service-alliance.ch

Contact: Gundula Heinatz-Bürki (gundula.heinatz@data-service-alliance.ch)

zhaw datalab
www.zhaw.ch/datalab

What to do with 70+ data scientists?

The ZHAW Data Science Laboratory, Datalab, is a virtual research group spanning three departments and five institutes of one of Switzerland's biggest universities of applied sciences. Being very strong in research, it brings together all researchers engaged in the particular aspects of data science under one roof for collaborative research and industry projects.

Founded in early 2013 as one of Europe's first groups dedicated to data science, ZHAW Datalab currently comprises more than 70 researchers from as diverse areas as law, analytics, computer science and entrepreneurship. It is one of the leading data science research centers in Switzerland and beyond.

Our offer: R&D projects

ZHAW Datalab offers the perfect team composition for any individual data science project. From our pool of domain experts, we form interdisciplinary project teams with the right mix of methodical and industry expertise. We are used to working together and crossing departmental borders.

Possible project settings are third party- or directly funded R&D projects as well as student's thesis projects, feasibility studies and consulting. Typical project durations vary from a few weeks to several years. We adopt our processes, tools and technologies to our partner's needs.

**Our service: education & community building**

ZHAW Datalab designed one of the first dedicated data science curricula in Europe: The "Master of Advanced Studies in Data Science" for professional education. Our associates also commit considerable amounts of time to various Bachelor, Master and Ph.D. degree programmes, thus sustaining the idea interchange between projects, students and industry.

We are committed to the Swiss data science community by means of several events and conferences we (co-) organize. One particular example is the "Swiss Conference on Data Science" series since 2014. We highly value this chance for idea exchange, trend spotting and passing on of lessons learned.

Get in touch

Find more information at www.zhaw.ch/datalab (e.g., about individual team members and projects), contact us directly via datalab@zhaw.ch, or follow us on Twitter:

@DataScienceCH
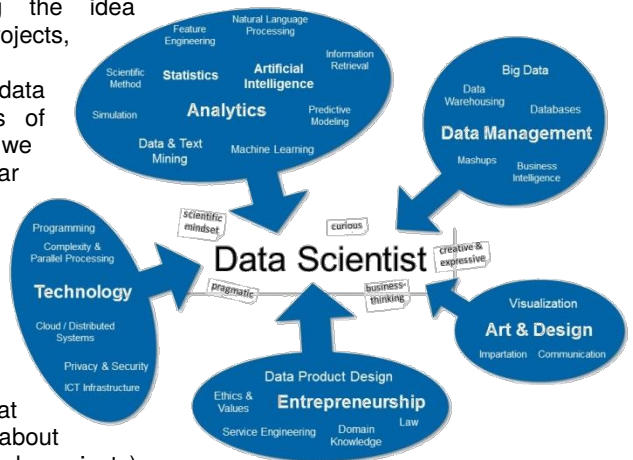
Bern University of Applied Sciences

DSE Research Group

The Data Science and Engineering (DSE) Research Group is part of the Institute for ICT-based Management at the Bern University of Applied Sciences. Our interdisciplinary team consists of scientists and technical staff with backgrounds in computer science, mathematics, and economics.

Competencies

We design and develop data science applications using modern methods (e.g., Design Thinking, Scrum) and technology stacks (such as Apache Spark, TensorFlow, and R). We have extensive experience with decision support systems and cockpits for the public sector. This includes the design of suitable management processes, supporting the development of organizational strategies as well as developing information systems that integrate data from different sources (e.g., accounting systems, citizen registers, and spatial information systems). In addition to structured data stored in relational databases, insights may also be derived from textual data from documents or social media. Data is analyzed via machine learning, mathematical models, physical laws or state-of-the-art organizational processes. Finally, data visualization and integration in work and decision contexts deliver value to partners in the industry and government.

Offer

We offer opportunities for joint research projects, including consulting, concept elaboration, and rapid prototyping of data science applications. We have a track record of 30+ projects that were completed successfully in the last 6 years, ranging from focused consulting projects for the industry to multi-year research projects with international partners. Further, we offer customized on-site training and workshops.

Contact

Prof Dr. Jürgen Vogel
juergen.vogel@bfh.ch
http://ictm.bfh.ch

Institute for Information and Communication Technologies (IICT)

The School of Management and Engineering Vaud (in French, Haute École d'Ingénierie et de Gestion du Canton de Vaud, HEIG-VD) is one of several schools within the University of Applied Sciences and Arts of Western Switzerland (HES-SO). As all the eight Swiss universities of applied sciences, the main missions of HEIG-VD include basic and continuing education, applied research and development, and services to other institutions. Created in 2006 by the merger of the cantonal schools for engineering and management, HEIG-VD has today more than 2000 students enrolled in ten Bachelor programs. HEIG-VD professors also contribute to three HES-SO Masters programs. HEIG-VD is located on three sites in the town of Yverdon-les-Bains, one of them being Y-Parc, the first and largest technology park in Switzerland.

The Institute for Information and Communication Technologies (IICT) is the largest R&D institute of HEIG-VD. Its activities pertain to software engineering, telecommunications and IT, with particular emphasis on intelligent data analysis, security, and advanced communication systems and networks. The IICT carries out about 50 research projects every year, often in collaboration with industrial partners. Over the past decade, the IICT has created four startup companies. The IICT has currently about 60 members, including 18 professors and several PhD students in co-supervision with other universities. One of its most promising research axes is machine learning and big data, with applications to sensor data analysis, bio-inspired methods and biological data, human computer interaction, information retrieval, natural language processing, and machine translation.

HEIG-VD | Route de Cheseaux 1 | CP 521 | 1401 Yverdon-les-Bains

+41 (0) 24 557 63 30 | https://heig-vd.ch | http://iict.heig-vd.ch/

Partner

HE-Arc is part of the University of Applied Sciences and Arts of Western Switzerland (HES-SO), the largest University of Applied Sciences in Switzerland and the second largest higher education institution of Switzerland. HE-Arc undertakes research projects with a wide range of partners, including research centres and universities in Switzerland or abroad, as well as public or private companies and institutions. Strongly anchored in the regional economy of the "Arc jurassien", territories of Neuchâtel/Bern/Jura Cantons, HE-Arc collaborates closely with SMEs, and its R&D also extends to certain aspects of industrial-scale production.

Dedicated to computer sciences applied to different areas varying from industrial-driven applications to research and academic-driven issues, Data Analytics Group - DAG at HE-Arc is a research group active in the fields of Data Mining, Big Data Predictive Analytics, High Performance Computing and Complex Numeric Algorithms. DAG has a solid theoretical and practical background in cutting edge technologies and latest advancements in the state of the art of Natural Language Processing such as textual data classification, clustering and regression algorithms deployed in CPU and GPU architectures. DAG has recently realised several solutions to solve concrete industrial problems related to big data analytics and genetic optimization on distributed ecosystems such as Hadoop/Yarn and Spark.

School of Computer and Communication Sciences
Machine Learning and Optimization Laboratory

Research:

Our main research directions are centered around

- Machine Learning
- Optimization
- Text Understanding

For example, we study unsupervised learning of representations of text (words and documents), such as sent2vec, or CNN based models. Furthermore, we develop and analyze new optimization algorithms for training machine learning models. Last but not least, our interest is in scalable distributed training algorithms and systems for classic ML models as well as deep learning, and open-source implementations and benchmarking of such systems.

Further Information:

Detailed information about our research, open positions and course materials can be found here: https://mlo.epfl.ch.

SUPSI

Information Systems and Networking Institute

The Information Systems and Networking Institute (ISIN) is part of the Department of Innovative Technologies at SUPSI. ISIN carries out teaching (Bachelor, Master and Continuing education) and applied research activities in the information and communication technology domain with a special focus of Internet of Things applications.

The robust research portfolio of the Institute spans many disciplines, including:

- Data and semantic analysis
- Multimedia processing
- Natural User Interfaces
- Cyber-security
- Pervasive communication
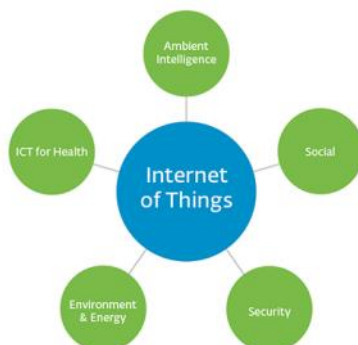
The wide range of competences is mostly exploited in Internet of Things projects involving local companies and other organizations.

Since it was established, the institute has considerably grown in terms of people, projects and applied research activity reaching about 40 employees and 3 Mio CHF yearly in applied leading-edge research projects. In the last few years dozens of KTI, EU and SNF projects have been acquired and completed.

SUPSI - Dipartimento tecnologie innovative

Istituto sistemi informativi e networking

Galleria 2, Via Cantonale 2c

CH-6928 Manno

T +41 (0)58 666 65 83

F +41 (0)58 666 65 71

isin@supsi.ch

Lucerne University of
Applied Sciences and Arts

HOCHSCHULE
LUZERN

Information Technology

FH Zentralschweiz

Lucerne School of Information Technology

The Lucerne University of Applied Sciences and Arts is the first Swiss university of applied sciences to have its own School of Information Technology. More than 600 students are enrolled in the bachelor's degree programs in Digital Ideation, Computer Science, Information & Cyber Security, International IT Management and Business Information Technology and master's degree programs in Engineering, Business Information Technology and Specialized Media and IT Teaching and Learning Methods at the Zug-Rotkreuz campus. The School of Information Technology is focusing its research on new areas of interest, ranging from artificial intelligence, machine learning and mobile and smart systems all the way to blockchain, smart contracts and visual computing. Numerous partners from the public and private sector benefit from the project skills and professional expertise of the researchers.

Continuing and executive education programs include the Master of Advanced Studies and Certificate of Advanced Studies in four attractive specialist areas. Courses, seminars and conferences lasting one or more days are also available. The Lucerne School of Information Technology also has outstanding links to domestic and foreign universities as well as to partners from all areas of economic, administrative, cultural and social life. Thanks to its range of programs, the School of Information Technology offers added value for business, society and culture.

For more information about this institution, please contact Prof. Dr. René Hüsler, Dean of the Lucerne School of Information Technology (rene.huesler@hslu.ch).

Institut für Wirtschaftsinformatik

Universität St.Gallen

Competence Center Crowdsourcing

University of St. Gallen

Institute of Information Management, Competence Center Crowdsourcing

The Institute of Information Management at the University of St. Gallen (IWI-HSG) is dedicated to applied and design-oriented research at the interface between business and information technology. The main research activities of the institute take place within its competence centers. The Competence Center Crowdsourcing (CCC) focuses on research related to the fields of crowdsourcing, data science, text analytics, collective and artificial intelligence, open innovation, and user-centricity in information systems.

The competence center's areas of interest include:

- Understanding the business value of text analytics and machine learning

- Developing methodologies for the generation and implementation of data-driven innovation, smart products, and smart services

- Investigating the potential of data and text mining approaches for processing user-generated content on large-scale online platforms

- Leveraging text mining and machine learning in software development and testing

- Combining collective intelligence embedded in crowdsourced data with text mining and machine learning algorithms for improved decision-making in organizations

- Designing intelligent decision support and business analytic systems based on the principles of text mining and machine learning

- Exploring the adoption and acceptance of automated decisions by decision-makers

The Competence Center Crowdsourcing works in close collaboration with several multinational partner organizations from different industries. The consortium aims at fostering the interaction and the exchange of knowledge between research and practice.

Contact

Prof. Dr. Ivo Blohm
Assistant Professor for Data Science and Management
Head of Competence Center Crowdsourcing

Institute of Information Management | University of St. Gallen
Müller-Friedberg-Str. 8 | CH-9000 St. Gallen
Phone: +41 71 224 33 21 | Fax: +41 71 224 27 77
Mail: ivo.blohm@unisg.ch
Internet: http://crowdsourcing.iwi.unisg.ch

University of Zurich UZH

Institute of Computational Linguistics

Computational Linguistics

Computational Linguistics investigates how human language is used as a means of transmitting, storing and processing information, and how these processes can be modelled on a computer and made available to specific applications. Searching information in the World Wide Web, analyzing texts in blogs and forums to gain insight in people's opinions, automatic text summarization or machine translation – Computational Linguistics attempts to make information available for our knowledge-based society.

Study

Computational Linguistics combines linguistics and information science and is the right choice for everyone that is interested in both areas.
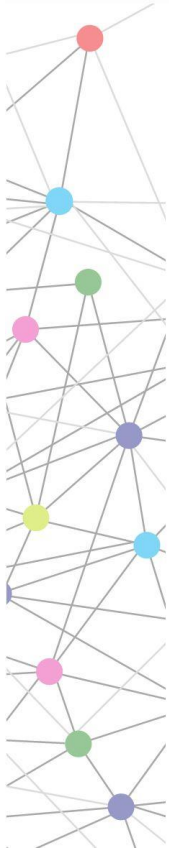
Research

Research topics of our institute include:

– Multilingual Text Analysis
– Machine Translation
– Sentiment Analysis and Opinion Mining
– Automated Media Analysis
– Biomedical Text Mining

Further information

Detailed information about our research and studying Computational Linguistics can be found at:
www.cl.uzh.ch

Swiss Distance University of Applied Sciences

FFHS
Fernfachhochschule Schweiz
Zürich | Basel | Bern | Brig

Mitglied der SUPSI

Created in 1998, the Swiss Distance University of Applied Sciences (Swiss Distance UAS) offers nationally recognised Bachelor and Master's degrees as well as a range of further education courses in economics, information technology, law and health.

In 2004, the Swiss Distance UAS became an affiliated school of the Scuola Universitaria Professionale della Svizzera Italiana (SUPSI). The Swiss Distance UAS offers degree courses in applied science based on blended learning. This means that about 80% of the training takes place through supervised distance study, while the remaining 20% is done at the regional centres in Zurich, Berne, Basel and Brig. By offering innovative courses that combine different types of e-learning with traditional education methods, and by training professionals without removing them from their jobs, the Swiss Distance UAS has closed a gap in the Swiss educational system. Thanks to this strategy it achieves a high degree of specialisation in each field of study.

In its three research institutes, the Swiss Distance UAS engages in applied research covering web science, management, innovation and e-learning that complies with all national Swiss education standards:

Institute for Research in Open-, Distance- and eLearning (IFeL)

The Institute for Research in Open-, Distance- and eLearning (IFeL) investigates the use of modern technology in learning. The interdisciplinary team of the IFeL consists of social and computer scientists and collaborates with both national and international partners. They apply a design-based research methodology and focus on emotions in reading and learning. Since autumn 2016, the institute has been home to the UNESCO Chair for Personalised and Adaptive Distance Education. In this context, it investigates the framework conditions and implementation concepts for personalised and adaptive learning.

The Institute for Management and Innovation (IMI)

The IMI research in the area of management and innovation. The institute cooperates with enterprises, organizations and groups as well as partner institutes from other colleges.

The core competencies of the IMI is in the thematic areas of innovation management, Organizational Intelligence: How to foster interdisciplinary collaboration to increase innovation organizational intelligence, interdisciplinary collaboration, organizational capabilities, and absorptive capacity. Research in these areas has been conducted, published in scientific journals and practitioner magazines, and presented at various conferences. The research have a strong background in qualitative research methods and is experienced in collaborating with management practitioners.

Laboratory for Web Science (LWS)

The Laboratory for Web Science (LWS) is a joint research unit of the FFHS and the Information Systems and Networking Institute (ISIN) of the SUPSI. The LWS focuses on the field of data science and it closely collaborates with industry partners on research projects to facilitate innovation and technology transfer.

ffhs.ch/en

Special Interest Group of the Swiss Informatics Society for Artificial Intelligence and Cognitive Science

The Special Interest Group for Artificial Intelligence and Cognitive Science (SGAICO) brings together researchers, practitioners and other parties interested in the subjects of artificial intelligence and cognitive science (AI/CO).

SGAICO pursues the goal of promoting intelligent technologies for innovation in our society. It provides a platform for exchange between industry and universities. You are cordially invited to join in our various activities and become a member of our Swiss network of experts as a platform for your ideas and initiatives.

- Discussion and dissemination of AI/CO-related knowledge
- Exchange on application problems occurring in Swiss enterprises and establishment of contacts between users and experts in Switzerland
- AI/CO methods and technologies in interdisciplinary contexts such as for example engineering, medicine, psychology and law
- Successful applications of AI/CO around the world and their relevance for Switzerland
- Topics and needs of AI/CO education in Switzerland

SGAICO supports and participates in many different events on AI /CO-related topics. It also supports Swiss researchers in a wide range of activities such as community exchange, establishing contacts, or the organization of events.

SGAICO is a member of the European Coordinating Committee for Artificial Intelligence (ECCAI), which is coordinating the European national AI/CO societies.

More Information: https://sgaico.swissinformatics.org/

Contact: sgaico@swissinformatics.org

President: Jana Koehler, Hochschule Luzern

Deputy Presidents: Jean-Daniel Dessimoz, HESSO-HEIG-VD, Marc-Oliver Gewaltig, EPFL, Thilo Stadelmann, ZHAW

schweizer informatik gesellschaft
société suisse d'informatique
società svizzera per l'informatica
swiss informatics society

informiert sein
dabei sein vernetzt sein
weiterkommen

Partner

Swiss Association for Analytics

The Swiss Association for Analytics (SAA) has been created in 2012. It is the very first Swiss group entirely dedicated to predictive and descriptive analytics. Our main objective is to raise awareness of Swiss companies to the benefits of analytics. By analytics, we mean the use of data mining and machine learning algorithms for data-driven decision making.

Whether the domain of your company is banking, finance, pharma, e-commerce or telecom doesn't matter. To benefit from analytics, companies need to have data, tools and know-how. The SAA can provide support in these three areas.

The SAA is a non-lucrative organization with the following objectives:

• Promote analytics within Switzerland
• Show the added value of analytics to Swiss companies
• Provide networking facilities for practitioners
• Exchange with other associations having related objectives

We have several means to achieve the above-mentioned objectives. We manage a LinkedIn group (www.linkedin.com/groups/4586163) in which we discuss topics such as trends, challenges, case studies, events and job offers. We also organize 3-4 analytics events each year (www.meetup.com/swiss-analytics). Our events regularly gather around 100 people in analytics. We also publish a printed magazine dedicated to Analytics, twice a year (www.swiss-analytics.com/magazine). If you are interested to be a sponsor, speaker or author, feel free to contact us at info@swiss-analytics.com.

Feedback

https://www.swisstext.org/2018/feedback

Impressum

Conference Chair
Dr. Mark Cieliebak

Program Committee
See next page.

Organization Committee
- Fernando Benites
- Jan Deriu
- Amrita Prasad
- Don Tuggener
- Dirk von Grünigen
- Ada Wan

The conference is co-organized by Zurich University of Applied Sciences and the Swiss Alliance for Data-Intensive Services.

Contact
Dr. Mark Cieliebak
+41 (0) 58 934 72 39
https://www.swisstext.org
info@swisstext.org

Program Committee

Swiss Track

- Ivo Blohm, University of St. Gallen
- Martin Jaggi, Ecole Polytechnique Fédérale de Lausanne
- Aurelien Lucchi, ETH Zurich
- Roberto Mastropietro, Scuola universitaria professionale della Svizzera italiana (SUPSI)
- Andrei Popescu-Belis, School of Business and Engineering Vaud (HEIG-VD)
- Fabio Rinaldi, University of Zurich
- Jürgen Spielberger, Zurich University of Applied Sciences (ZHAW)
- Jürgen Vogel, Bern University of Applied Sciences (BFH)
- Manfred Vogel, Fachhochschule Nordwestschweiz (FHNW)
- Tim vor der Brück, Lucerne University of Applied Sciences and Arts (HSLU)
- Egon Werlen, Swiss Distance University of Applied Sciences (FFHS)

Scientific Track

- Giuseppe Attardi, Università di Pisa
- Adrien Barbaresi, Österreichische Akademie der Wissenschaften (ÖAW)
- Roberto Basili, University of Roma Tor Vergata
- Eduardo Blanco, University of North Texas
- Ivo Blohm, University of St. Gallen
- Noah Bubenhofer, University of Zurich
- Ernst Buchberger, Medical University of Vienna
- Manuel De Buenaga, Universidad Europea de Madrid
- Stephan Busemann, German Research Center for Artificial Intelligence (DFKI)
- Miriam Butt, Universität Konstanz
- Pascual Cantos-Gómez, University of Murcia
- Sylviane Cardey, Universite de Franche-Comte
- Ernesto William De Luca, Georg-Eckert-Institute for international Textbook Research
- Gaël Dias, University of Caen Normandie
- Nicola Ferro, University of Padova
- Alexander Gelbukh, Instituto Politécnico Nacional, Mexico
- Hatem Ghorbel, University of Applied Sciences HES-SO, HE-Arc Ingénierie
- Xavier Gómez Guinovart, Universidade de Vigo
- Martin Jaggi, Ecole Polytechnique Fédérale de Lausanne
- Manfred Klenner, University of Zurich
- Valia Kordoni, Humboldt University Berlin
- Udo Kruschwitz, University of Essex
- Ekaterina Lapshinova-Koltunski, Saarland University
- Aurelien Lucchi, ETH Zurich
- Roberto Mastropietro, Scuola universitaria professionale della Svizzera italiana (SUPSI)
- Alexander Mehler, Goethe University Frankfurt
- Wolfgang Menzel, University of Hamburg
- Margot Mieskes, University of Applied Sciences, Darmstadt
- Marie-Francine Moens, Katholieke Universiteit Leuven
- Alessandro Moschitti, Qatar Computing Research Institute (QCRI)
- Guenter Neumann, German Research Center for Artificial Intelligence (DFKI)
- Johann Petrak, The University of Sheffield
- Hannes Pirker, Austrian Academy of Sciences (ÖAW)
- Andrei Popescu-Belis, School of Business and Engineering Vaud (HEIG-VD)
- Uwe Quasthoff, University Leipzig
- Fabio Rinaldi, University of Zurich
- Paolo Rosso, Universitat Politècnica de València
- Tanja Samardzic, University of Zurich
- Yves Scherrer, University of Helsinki
- Gerold Schneider, University of Zurich
- Rico Sennrich, The University of Edinburgh
- Marcin Skowron, Austrian Research Institute For Artificial Intelligence (OFAI)
- Jürgen Spielberger, Zurich University of Applied Sciences (ZHAW)
- Angelika Storrer, Universität Mannheim
- Manfred Vogel, Fachhochschule Nordwestschweiz (FHNW)
- Juergen Vogel, Bern University of Applied Sciences (BFH)
- Tim vor der Brück, Lucerne University of Applied Sciences and Arts (HSLU)
- Egon Werlen, Swiss Distance University of Applied Sciences (FFHS)
- Fabio Massimo Zanzotto, University of Rome Tor Vergata
- Torsten Zesch, University of Duisburg-Essen
- Heike Zinsmeister, University of Hamburg

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Innosuisse – Swiss Innovation Agency

Google

data +service
Swiss Alliance for
Data-Intensive Services

IBM **Research**

SPITCH

D|ONE
WE MAKE SENSE.

SPINNINGBYTES

SUPERTEXT

zh aw

Zurich University of
Applied Sciences

die **Mobiliar**

ARGUS DATA INSIGHTS®
WISSEN ZUM ERFOLG

parashift

Institut für Wirtschaftsinformatik

datalab
www.zhaw.ch/datalab

Universität St.Gallen

F
H
**Bern University
of Applied Sciences**

EPFL
ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Machine Learning and
Optimization Laboratory

Lucerne University of
Applied Sciences and Arts

HOCHSCHULE
LUZERN

WIN LINK
INFORMATION AND
COMMUNICATION TECHNOLOGY

swiss group for artificial intelligence
and cognitive science

SGAICO

Swiss
Association
for Analytics

SUPSI

University of
Zurich UZH

ETH zürich
Data Analytics Lab

FFHS

heig-vd
HAUTE ÉCOLE
D'INGÉNIERIE ET DE GESTION
DU CANTON DE VAUD
www.heig-vd.ch

n|w
University of Applied Sciences and Arts
Northwestern Switzerland