Welcome Message

Mark Cieliebak
Conference Chair
Where do you come from?
Where do you come from?
Research Institutions

1. Åbo Akademi University
2. Ca' Foscari University of Venice
3. EPFL Lausanne
4. ETH Zurich
5. Helmut-Schmidt-University Hamburg
6. Lucerne University of Applied Sciences and Arts HSLU
7. Università della Svizzera italiana USI
8. Universitat Politècnica de València
9. University of Applied Sciences and Arts of Southern Switzerland SUPSI
10. University of Applied Sciences Northwestern Switzerland FHNW
11. University of Applied Sciences St. Gallen FHSG
12. University of Applied Sciences Western Switzerland HES-SO
13. University of Basel
14. University of Edinburgh
15. University of Neuchatel
16. University of Zurich
17. Zurich University of Applied Sciences ZHAW
>170 Participants from Research and…..

- 42matters AG
- Aebi, Völker, Und, AG für Kommunikation
- ARGUS der Presse AG
- AXA Winterthur
- Carpe Media
- CHUV
- Cognism
- Credit Suisse AG
- CSS Versicherung AG
- Deloitte Consulting
- Die Mobiliar
- ebay
- emineo AG
- Expert System
- finnova AG Bankware
- ForeKnowledge GmbH
- GateB
- Google
- HE-Arc Ingénierie
- Helsana Versicherungen
- Holidaycheck AG
- IBM
- Iprova
- Iterativ GmbH
- KPMG
- KPT Krankenkasse AG
- LUSTAT Statistik Luzern
- Microsoft
- Migros-Genossenschafts-Bund
- Modulo Language Automation LLC
- Namics AG
- Newscron SA
- OrganizationView
- Open Systems AG
- Plazi
- plus-IT AG
- PricewaterhouseCoopers
- Procter & Gamble
- Schreibwerkstatt GmbH
- Schweizer Radio und Fernsehen
- Semfinder AG
- SpinningBytes AG
- Spitch AG
- SRF Data
- Supertext AG
- Swiss Federal Archives
- Swiss Life
- Swiss Mobiliar
- Swiss National Science Foundation
- Swisscom
- Syntax Übersetzungen
- Tirsus GmbH
- Twygg
- UBS AG
- UPC
- Vidatics GmbH
- Wabion AG
- Yourposition AG
- Zentralbibliothek Zürich
- Zurich Insurance
Flipped Classroom (also known as “Inverted Teaching”) is a teaching method where lecture and homework are “flipped”: first, students prepare the topic of the next lecture at home, e.g. by reading part of a text book, watching e-lectures, or working with e-learning-modules. Then, in the lecture, students work with the teacher to clarify open questions, discuss the topic and solve exercises. Knowledge-transfer mainly happens in the first, preparatory phase, leaving time and space for more communicative and collaborative activities during the lecture. Under the teacher’s guidance, students establish cognitive connections to their prior knowledge - in traditional lectures, this task is left to be achieved by the students at home, after the lecture. Flipped Classroom lectures are highly interactive and students assume more responsibility for their own learning than in traditionally taught classes. For a successful implementation of the Flipped Classroom method, completion of the preparational tasks is crucial [4], and can be supported by online check-up questions, to be answered before the lecture. Flipped Classrooms have become very popular in recent years. They go back to the 1990’s, when Eric Mazur introduced Peer Instructions in his physics lectures at Harvard University [9]. Since then, the concept has evolved into an established teaching method that is now used successfully at elementary schools, high schools and universities worldwide. Hundreds of guidelines, reports, books, conference proceedings and research papers have been published on Flipped Classroom. For an extensive literature and research survey, see [1] and [6].
Text is Easy!

• Small alphabet

• Very structured: grammar, dictionaries, rules

• Easy to store, share and access
Text is not so Easy!

- Artificially constructed
- Different Languages
- Typos and Errors
Text is Ambiguos

Peter hit the man with an umbrella.

Each of us saw her duck.

I walked all the way across campus to hear the bacteria talk.

The Women were decapitated in an accident before attending the lecture.
Text is Fragile

I have a new cat
There are so many other topics...
…why are we so interested in Text?

TEXT IS IMPORTANT

Contracts  Diaries  Laws

Patents  Project Plans

Manuals  News  Research Results

Libraries  Love Letters
In 10 Seconds…

Type 5 Words on Computer

Publish 0.8 Research Papers

Send 35 Million Emails

Read 33 Words

Scan 13 Text Pages

Compute 85'000'000'000 Operations
Text Analytics in a Nutshell

Text → Algorithms → Information
Text Analytics in a Nutshell

Sources
• Medical research papers
• Legal texts - laws, court rulings
• Patents
• Social Media
• News
• Customer Feedback
• Websites
• Project Proposals
• Technical Documentation
• Speech Transcriptions

Goals
Classification
• Sentiment Detection
• Author Profiling (Age, Gender)

Information Extraction
• Named Entity Recognition
• Keyphrase extraction

Topic Analysis
• Document Clustering
• Hashtag prediction
• Topic Categorization

Text Generation
• Machine Translation
• Question and Answering
• Text Summarization
Text Preprocessing

Preprocessing

Text Analysis
Preprocessing

- Language Detection
- Sentence Splitting
- Tokenization
- Stemming/Lemmatization
- Stopword Elimination
- POS tagging
- Syntactic Parsing

The ball is blue. Mr. O'Neill thinks these aren’t U.S. cities.

Issues: C++, C#, B-52, B777, M*A*S*H, info@swisstext.org, +41 58 934 72 39
different, differently -> differ
cars -> car

a, about, above, across, ...
Issues: The president of the United States...

The ball is blue.

The dog bit the man.
Text Preprocessing

Preprocessing

Text Analysis

- Rule-Based
- Machine Learning
- Deep Learning
Machine Learning with Features

Train
- Training Data
  - Text documents
- Labels

Features

Machine Learning Algorithm

Application
- New Document
- Predictive Model
- Predicted Label
Machine Learning Algorithms

- AODE
- Artificial neural network
- Backpropagation
- Autoencoders
- Hopfield networks
- Boltzmann machines
- Restricted Boltzmann Machines
- Spiking neural networks
- Bayesian statistics
- Bayesian network
- Bayesian knowledge base
- Case-based reasoning
- Gaussian process regression
- Gene expression programming
- Group method of data handling (GMDH)
- Inductive logic programming
- Instance-based learning
- Lazy learning
- Learning Automata
- Learning Vector Quantization
- Logistic Model Tree
- Minimum message length (decision trees, decision graphs, etc.)

- Nearest Neighbor Algorithm
- Analogical modeling
- Probably approximately correct learning (PAC) learning
- Ripple down rules, a knowledge acquisition methodology
- Symbolic machine learning algorithms
- Support vector machines
- Random Forests
- Ensembles of classifiers
- Bootstrap aggregating (bagging)
- Boosting (meta-algorithm)
- Ordinal classification
- Information fuzzy networks (IFN)
- Conditional Random Field
- ANOVA
- Linear classifiers
- Fisher's linear discriminant
- Logistic regression
- Multinomial logistic regression
- Naive Bayes classifier
- Perceptron
- Support vector machines
- Quadratic classifiers
- k-nearest neighbor
- Boosting
- Decision trees
- C4.5
- Random forests
- ID3
- CART
- SLIQ
- SPRINT
- Bayesian networks
- Naive Bayes
- Hidden Markov models
- Unsupervised learning
- Expectation-maximization algorithm
- Vector Quantization
- Generative topographic map
- Information bottleneck method
- Artificial neural network
- Self-organizing map
- Association rule learning
- Apriori algorithm
- Eclat algorithm
- FP-growth algorithm
- Hierarchical clustering

- Single-linkage clustering
- Conceptual clustering
- Cluster analysis[edit]
- K-means algorithm
- Fuzzy clustering
- DBSCAN
- OPTICS algorithm
- Outlier Detection
- Local Outlier Factor
- Semi-supervised learning
- Reinforcement learning
- Temporal difference learning
- Q-learning
- Learning Automata
- SARSA
- Deep learning
- Deep belief networks
- Deep Boltzmann machines
- Deep Convolutional neural networks
- Deep Recurrent neural networks
- Hierarchical temporal memory
- Data Pre-processing
- List of artificial intelligence projects
Deep Learning

Unlabelled Data
Text documents

DNN

Neural Language Model

Unlabelled Data
Text documents

DNN

Neural Language Model

Training Data
Text documents

Features

Machine Learning Algorithm

Labels

Features

Machine Learning Algorithm

New Document

Features

Predictive Model

Predicted Label

Features

Predictive Model

Predicted Label
Challenges!

- Availability of Training Data
- Low Annotation Quality
- Disambiguation

“I found my wallet near the bank.”

- Co-reference Resolution

“The cat is in the house. It is white.”

- Irony

“What a great car – it stopped working after two days!”

- Slang

“#YouCantDateMe if u still sag ur pants super hard...dat shit is played the fuck out!!!”
What can you do with Text Analytics?

- Email Spam Detection
- Siri & co
- Machine Translation
- Social Media Monitoring
- Internet Search
- Newspaper Segmentation
- Company Profiling
- ADR Detection in Twitter

Applications
Thank You!

Mark Cieliebak

Zurich University of Applied Sciences (ZHAW)
Email: ciel@zhaw.ch, Website: www.zhaw.ch/~ciel
# Today’s Program

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:30</td>
<td>Welcome Message: Mark Cieliebak</td>
</tr>
<tr>
<td>10:00</td>
<td>Keynote: Paolo Rosso</td>
</tr>
<tr>
<td>10:45</td>
<td>Survey Session 1</td>
</tr>
<tr>
<td>11:40</td>
<td>Keynote: Jürg Attinger</td>
</tr>
<tr>
<td>12:10</td>
<td>Survey Session 2</td>
</tr>
<tr>
<td>13:00</td>
<td>Lunch Break</td>
</tr>
<tr>
<td>14:00</td>
<td>Presentations: 2 Parallel Tracks</td>
</tr>
<tr>
<td>15:35</td>
<td>Poster Session</td>
</tr>
<tr>
<td>16:05</td>
<td>Keynote: Katja Filippova</td>
</tr>
<tr>
<td>16:50</td>
<td>Closing + Apero</td>
</tr>
</tbody>
</table>
Best Presentation Award

- Selected by the audience
- Keynote speakers not eligible

www.swisstext.org/2016/feedback
Good to Know
Organizing Committee

Mark Cieliebak
Dominic Egger
Simon Müller
Daniel Schutzbach
Bettina Bhend