

Black Swan

Discovering events that matter.

Visit us as www.BlackSwanEvents.org

The Black Swan Theory (Taleb 2007)

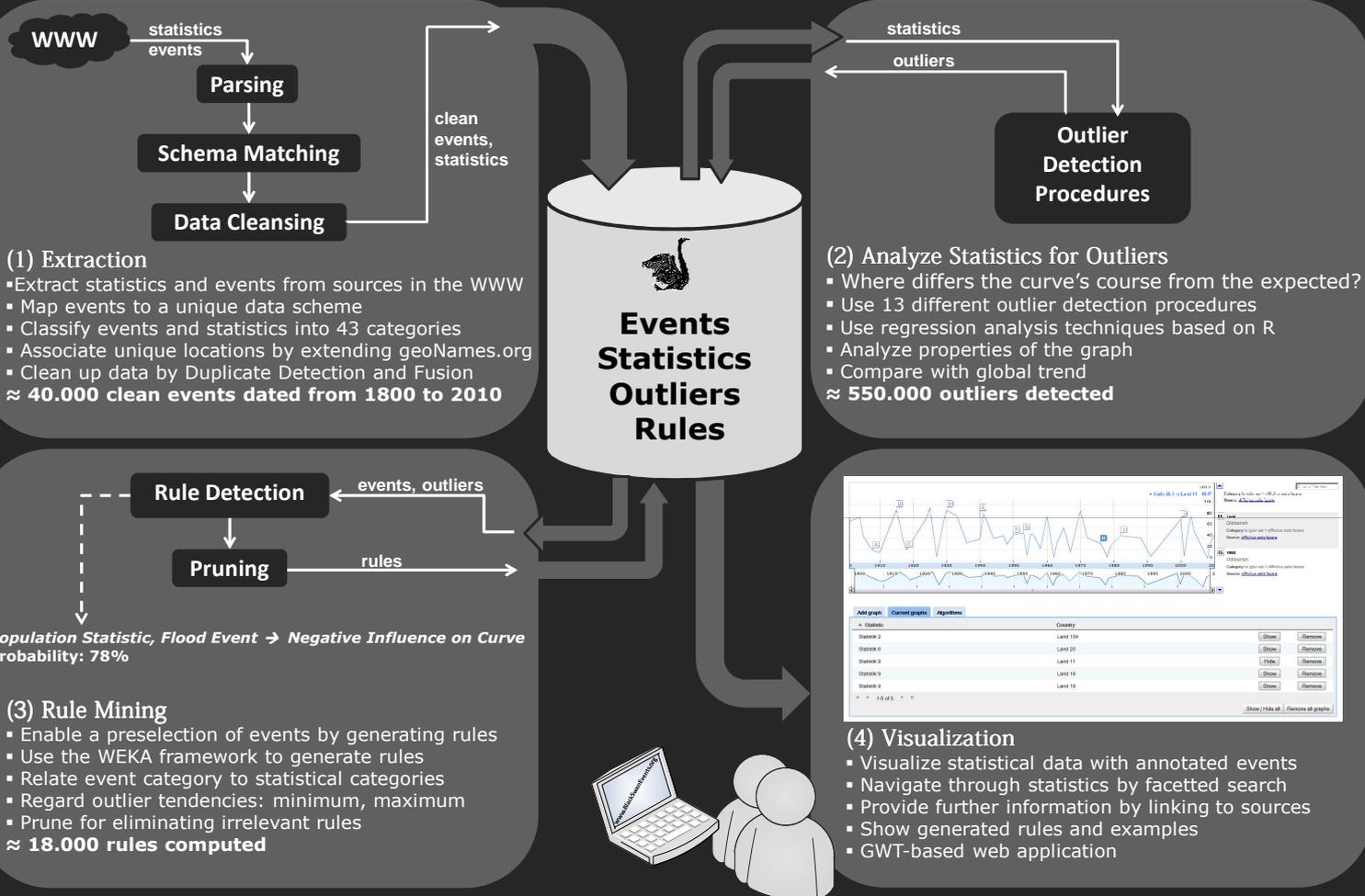
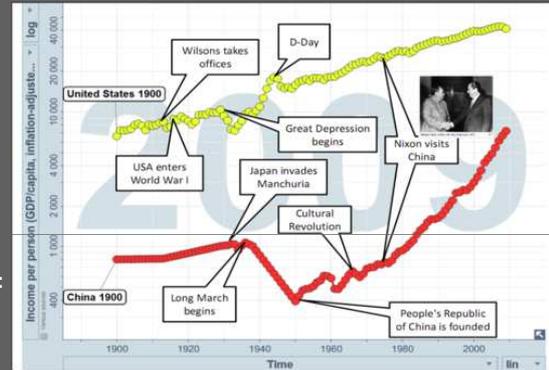
- Black Swans: unpredicted historic events with a major impact, rationalized by hindsight
- Understanding the past by Black Swans: *Why recur certain developments in history?*

Automated Annotation of Global Statistics

- Help historians to identify Black Swans
What historical events had a major impact?
- Detect anomalies in statistics
Where are inexplicable statistical trends?
- Associate causative events with statistics
What events triggered specific developments?
- Generate rules for automated detection
Which event types influence what statistics?

The Challenge

- Analyze statistics for outliers:
What defines an outlier?
- Perceive events and statistical data:
Where to search?
- Associate events to statistical outliers:
40.000 events, 450 statistics
- Visualize:
Tool-supported analysis of statistics



(1) Extraction

- Extract statistics and events from sources in the WWW
- Map events to a unique data scheme
- Classify events and statistics into 43 categories
- Associate unique locations by extending geoNames.org
- Clean up data by Duplicate Detection and Fusion
- ≈ **40.000 clean events dated from 1800 to 2010**

(2) Analyze Statistics for Outliers

- Where differs the curve's course from the expected?
- Use 13 different outlier detection procedures
- Use regression analysis techniques based on R
- Analyze properties of the graph
- Compare with global trend
- ≈ **550.000 outliers detected**

(3) Rule Mining

- Enable a preselection of events by generating rules
- Use the WEKA framework to generate rules
- Relate event category to statistical categories
- Regard outlier tendencies: minimum, maximum
- Prune for eliminating irrelevant rules
- ≈ **18.000 rules computed**

(4) Visualization

- Visualize statistical data with annotated events
- Navigate through statistics by faceted search
- Provide further information by linking to sources
- Show generated rules and examples
- GWT-based web application



Sören Discher², Cindy Fähnrich², Benedikt Forchhammer², Stefan Lemme², Andrina Mascher², Thorsten Papanbrock², Christoph Peschel², Peter Retzlaff², Stephan Richter², Thomas Stening², Sven Viehmeier², Armin Zamani Farahani², and Johannes Lorey¹, Felix Naumann¹

Information Systems Group at Hasso Plattner Institute, Potsdam, Germany

¹: firstname.lastname@hpi.uni-potsdam.de

²: firstname.lastname@student.hpi.uni-potsdam.de



IT Systems Engineering | Universität Potsdam