

# Follow-Up Project: Automated Text Mining on Job Offers Using SAP HANA

## Analyzing Skill and Competency Requirements for Industry 4.0

### Situation

**Industrie 4.0/Industry 4.0/the Industrial Internet:** combining known technologies in a new way: disruptive solutions

- ▶ The Internet of Things
- ▶ Cyber Physical Systems (CPS)
- ▶ Smart Factories
- ▶ Embedded Systems

(Kagermann et al. 2013, Industrie 4.0 2016)

### Complication

The way we live and work will change significantly

- ▶ New ways of business value creation
- ▶ Changing business models and strategies
- ▶ Adjusted business processes

(Kagermann et al. 2013)

### Resolution

Apply text analysis and text mining offered by SAP HANA and the SAP Predictive Analysis Library (PAL) on job offers collected from German online job portals to extract skill and competency requirements for Industry 4.0:

- ▶ Build on technical experience from former projects
- ▶ Apply technical knowledge to a new area
- ▶ Try to discover competency profiles from analyzing job offers

### System

- ▶ SAP HANA 2 SPS01, 1 TB RAM, 32 Cores (CPU)
- ▶ SAP HANA Studio, version 2.3.10
- ▶ PAL library
- ▶ Python, version 3.5.2
- ▶ Node.js 6.9.5 LTS

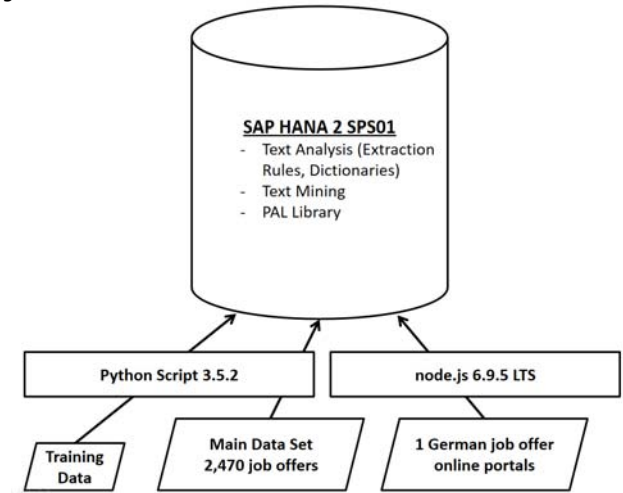
### Data Sets

- ▶ Manually collected German job offers, most tests done on collection of Nov – Apr. (2,470 job offers)
- ▶ Test data sets: T1: 15 job offers, T2: 50 job offers, manually classified

	Job Portal A	Job Portal B	
Nov. 2016	178	100	
Dec. 2016	271	250	
Jan. 2017	210	175	
Feb. 2017	234	174	
Mar. 2017	231	176	
Apr. 2017	273	198	
Mai 2017	293	176	
Jun. 2017	237	175	Total
<b>Total</b>	<b>1,927</b>	<b>1,424</b>	<b>3,351</b>

Source: Own illustration.

### Project Architecture



Source: Own illustration.

### Results

- ▶ **Web Crawler**
  - Implemented as stand alone application using node.js
  - Working with one German online job offer portal so far
- ▶ **Skill and competency requirement extraction**
  - Improvement of custom dictionaries and custom extraction rules
  - Improvement of extraction result in comparison to former project:
    - 68% of skills and competency requirements extracted
    - Precision: 92,8%, sensitivity: 77,5%, F1-score: 84,5%
  - Recommendations for further improvement
    - Convert parts of custom dictionary into custom extraction rules
    - Maybe apply the Grammatical Role Analysis to reduce false positives (only available in English so far)
  - Frequency analysis on discovered skill and competency requirements
    - Top 5: languages (64%), flexibility (40%), capacity for teamwork (36%), communication skills (29%), and self-responsibility (25%)
  - Association analysis: Weak results, difficulties with the attempt to discover competency profiles
  - Clustering: Good approach for deriving competency profiles from clusters

### Next Steps

- ▶ Integrate web crawler into SAP HANA application
- ▶ Extend functionality of web crawler
- ▶ Improve job profile discovery
- ▶ Analysis of skill and competency requirements with regards to metadata (e.g., job title, company, region)
- ▶ Implementation of a regular analysis to discover changes over time