ELG Pilot Project: Streaming Language Processing in Manufacturing
SLAPMAN: Streaming Language Processing in Manufacturing

Dominik Riemer, Steffen Thoma
01.12.2020
Motivation

Exploiting text data in the manufacturing domain

Idea #1
Machine log data

Idea #2
Service technicians

How to utilize machine log data?

How can I conduct flexible data analysis without deep IT skills?

How to utilize knowledge of service technicians?
What is Apache StreamPipes?

“An open source IIoT toolbox to enable non-technical users to connect, explore and analyze IoT data streams.”

Developed @FZI since 2015, transition to Apache Incubator in 2019

> 30 industrial data connectors to connect industrial assets within minutes

> 100 included algorithms for self-service data analytics

live dashboard & historical data explorer

streampipes.apache.org/ | @streampipes | github.com/apache/incubator-streampipes

> 100 Pipeline Elements

[Diagram showing various features of StreamPipes]

1. Connect IoT sources
2. Leverage reusable data processors & sinks
3. Deploy & execute the way you like
4. Realize use cases

- Continuous Asset Monitoring
- Calculate live KPI
- Wrap your ML models and use them on sensor/image data
Project objectives

Investigating the role of industrial textual (e.g., log) messages in the manufacturing domain to improve product and process quality

Main Objectives

- Develop/Integrate various text analytics features into Apache StreamPipes
- Develop methods to easily train new text analytics models for machine assets
- Provide a customized StreamPipes version to the ELG platform

ELG NLP Components for StreamPipes

- NER (BERT, Gate, Cogito), Sentiment, Translation, …
- Integration with StreamPipes using the provided SDK and runtime wrappers

NLP Learning Component for StreamPipes

- Starting with an initial model, i.e. pre-trained general model or learned on a specified text corpus
- Continuous adaptation to the use case at hand via adaptation in regards to the incoming textual data stream
Let's connect!

- streampipes.apache.org
- streampipes.apache.org/docs
- apache/incubator-streampipes
- @streampipes

... and leave a star on Github if you like it 😊

{riemer|thoma}@fzi.de
**1 December 2020 – META-FORUM 2020 Day 1**

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
<th>Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>12:00 – 13:30 (CET)</td>
<td>E3G (European Clinical Case Corpus)</td>
<td>Link to the Virtual meeting room</td>
</tr>
<tr>
<td></td>
<td>LSDISCO (Lingsoft Solutions as Distributable Containers)</td>
<td>Link to the Virtual meeting room</td>
</tr>
<tr>
<td></td>
<td>MKS-LLOD (MKS as Linguistic Linked Open Data)</td>
<td>Link to the Virtual meeting room</td>
</tr>
<tr>
<td></td>
<td>Smart euSpeaker (Basque-speaking smart speaker based on Mycroft AI)</td>
<td>Link to the Virtual meeting room</td>
</tr>
<tr>
<td></td>
<td>EVALITA-ELG (Italian EVALITA Benchmark Linguistic Resources for the ELG platform)</td>
<td>Link to the Virtual meeting room</td>
</tr>
<tr>
<td></td>
<td>OPUS-MT (Open Translation Models, Tools and Services)</td>
<td>Link to the Virtual meeting room</td>
</tr>
<tr>
<td></td>
<td>Text2TCS (Extracting Terminological Concept Systems from Natural Language Text)</td>
<td>Link to the Virtual meeting room</td>
</tr>
<tr>
<td></td>
<td>PARA4DLM (Textual paraphrase dataset for deep language modeling)</td>
<td>Link to the Virtual meeting room</td>
</tr>
<tr>
<td></td>
<td>YouTwinDI (Virtual Personal Assistant Prototype)</td>
<td>Link to the Virtual meeting room</td>
</tr>
<tr>
<td></td>
<td>SLAPMAN (Streaming Language Processing in Manufacturing)</td>
<td>Link to the Virtual meeting room</td>
</tr>
<tr>
<td></td>
<td>ELG (European Language Grid)</td>
<td>Link to the Virtual meeting room</td>
</tr>
</tbody>
</table>

**Virtual Project Expo**

Thank you!