NLP in Analytics & Cognitive
META-FORUM 2020
Deloitte’s Achievements in NLP

1. Text extraction in 55+ languages

2. Concept extraction in 16 languages
   (102 planned in pilot)

3. Text pre-processing in 16 languages
   eventually 100+)

4. Semantic vectors in 16 languages
   (currently working with multilingual models, with up to 104 supported languages)

4. Knowledge graphs
   (Identifying subject-predicate-object triplets, who did what to whom, Understanding relationships and events)

Fully model based approach which enables to use any language to define the requirements.
What we boast with

**Fully Automated NLP Techniques**

Full automated solution using state-of-the-art NLP techniques:
- Minimal manual input needed
- Fast implementation and fast onboarding of a new client
- Great scalability - not limited by amount of monitored companies

**Open-Source Technology**

Independent from expensive external data providers, running almost entirely on open-source technology and publicly available information.

**Multiple Languages within one Framework**

Possibility to work with multiple languages within one framework:
- No machine translation until the reporting stage
- Focus even on non-English sources

**End-to-end solution in the house**

Not only technical tool, but ability to offer subsequent services - consulting or business development.
Potential for ELG

Tap ELG for:

1. Experienced subject matter experts
2. Tools ready to test and implement
3. Data corpus

Looking for:

1. New real world cases
2. Potential partnerships / consortiums to respond to requests
Concept Linking
to real world entities

1. Organizations, people, places, products, and other “concepts” can be extracted and used to label and filter articles

2. Entities in English text can be linked to actual entities in Wikipedia (useful for developing a knowledge base)

“Paris is the capital of France”

wikipedia.org/wiki/Paris
wikipedia.org/wiki/France
Knowledge Graph

Extracting key events from text and representing them in a searchable graph:

1. **Identifying subject-predicate-object triplets**
   (who did what to whom)

2. **Linking concepts to entities in Wikipedia**

3. **Understanding relationships and events**

<table>
<thead>
<tr>
<th>subject</th>
<th>relation</th>
<th>object</th>
<th>subject_type</th>
<th>object_type</th>
</tr>
</thead>
<tbody>
<tr>
<td>launch Data</td>
<td>is in</td>
<td>cloud</td>
<td>NOUN_CHUNK</td>
<td>NOUN_CHUNK</td>
</tr>
<tr>
<td>database</td>
<td>hosted on</td>
<td>Microsoft Azure Microsoft Azure Microsoft Azure</td>
<td>NOUN_CHUNK</td>
<td>NOUN_CHUNK</td>
</tr>
<tr>
<td>Microsoft customer</td>
<td>know of</td>
<td>problem</td>
<td>NOUN_CHUNK</td>
<td>NOUN_CHUNK</td>
</tr>
<tr>
<td>Rotem</td>
<td>to</td>
<td>Discovery of unsecured database owner identified</td>
<td>NOUN_CHUNK</td>
<td>NOUN_CHUNK</td>
</tr>
<tr>
<td>database</td>
<td>hosted on</td>
<td>Microsoft Azure Microsoft Azure</td>
<td>NOUN_CHUNK</td>
<td>NOUN_CHUNK</td>
</tr>
<tr>
<td>data</td>
<td>hosting</td>
<td>data inside database hosted on Microsoft Azure in April</td>
<td>NOUN_CHUNK</td>
<td>NOUN_CHUNK</td>
</tr>
</tbody>
</table>
Thank You.

Julius Rusnak
CTO Analytics & Cognitive
Contact: jrusnak@deloitteCE.com