LANG.TEC
SEMANTIC TEXT PROCESSING
LANG.TEC

SEMANTIC TEXT PROCESSING
Presentation Outline

1. Who we are
2. Business areas
3. Clients
4. Introducing QUEST – the MT Quality Estimator
5. Comparing translation modes (HT, PEMT, QUEST)
6. Demo
7. Model architecture
8. Parting words
Who we are

LangTec:

• 10 years in business, 15 employees
• We’ve been working on MTQE since 2018
• QUEST is our second fully functioning QE model and a significant advance over its predecessor
Clients

1. Semantic Analytics

Text Mining

- abbvie
- BSS
  Beiersdorf Shared Services
- PRIME
- RESEARCH
- billiger.de
- vodafone
- Universität Hamburg
  DER FORSCHUNG | DER LEHRE | DER BILDUNG
- trendne
- mgm
  technology partners
- shareifyoulike
- tritra
- pwc

Technology Stack

- DB
- BOSCH
- Continental
- starfinanz
- otto group
- Beiersdorf
- Siemens
- Otto
- notebooksbilliger.de
- ING
- DiBa
- getabstract
  compressed knowledge
- tesa

Text Analytics

LANG.TEC

SEMANTIC TEXT PROCESSING

Confidential
## Clients

### 2. Automated Text Generation

<table>
<thead>
<tr>
<th>TextWriter Client</th>
<th>NLG Consulting</th>
<th>Bespoke Software</th>
</tr>
</thead>
<tbody>
<tr>
<td>Campaign Trail</td>
<td>starfinanz</td>
<td>LANG.TEC</td>
</tr>
<tr>
<td>WELT N24</td>
<td>MeteoGroup</td>
<td>LANG.TEC</td>
</tr>
<tr>
<td>Anonymer Kunde</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regional-Nachrichten.net</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bild am Sonntag</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lagerhaus G</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tradegate exchange</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Confidential

Proceedings of the 14th Conference of the Association for Machine Translation in the Americas
October 6 – 9, 2020, 1st Workshop on Post-Editing in Modern-Day Translation
Clients

3. Computational Linguistics

Localisation

Resources

Linguistic Analysis

Microsoft

Lionbridge

novomind
defined crowd
CARMEQ.
gracenote.

Cognotekt

OnMaCon

voicebox

resmio

ELSEVIER
Introducing QUEST –
MT Quality Estimator
Overview

1. Our in-domain QUEST model is capable of predicting post-editing effort with unprecedented accuracy: only 2% error.

2. QUEST can be used with any language pair, any machine translation engine and for any domain.

3. QUEST can also be used out-of-domain still providing superior prediction accuracy compared with experience-based estimations and other machine learning models.

4. QUEST’s prediction accuracy enables translation service providers to achieve much more competitive pricing and deadline-setting.
In-Domain QUEST Model

- When historical domain data are available we can also train in-domain models.
- The in-domain baseline beats the out-of-domain baseline.
- The in-domain baseline also beats the out-of-domain TCA.
- The out-of-domain QUEST model beats all baselines and the out-of-domain TCA.
- The in-domain QUEST model beats all baselines and all TCA models, very closely approximating the actual post-editing effort.
## Comparing approaches

<table>
<thead>
<tr>
<th></th>
<th>Manual Translation (HT)</th>
<th>MT + Post-Editing (In-domain baseline)</th>
<th>MT + Post-Editing + QUEST (out-of-domain)</th>
<th>MT + post-editing + QUEST (in-domain)</th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word Count</td>
<td>22,635</td>
<td>22,635</td>
<td>22,635</td>
<td>22,635</td>
<td>22,635</td>
</tr>
<tr>
<td>Predicted Post-Editing Operations</td>
<td>0</td>
<td>10,474</td>
<td>9,538</td>
<td>6,443</td>
<td>6,561</td>
</tr>
<tr>
<td>Prediction Error</td>
<td>0%</td>
<td>60%</td>
<td>45%</td>
<td>2%</td>
<td>0%</td>
</tr>
<tr>
<td>Total Turnover Effort</td>
<td>56.59 hrs. 7.07 person days</td>
<td>48.00 hrs. 6.00 person days</td>
<td>43.72 person hrs. 5.46 person days</td>
<td>29.53 hrs. 3.69 person days</td>
<td>30.07 hrs 3.76 person days</td>
</tr>
<tr>
<td>Fraction of words to be processed</td>
<td>100%</td>
<td>46%</td>
<td>42%</td>
<td>28%</td>
<td>29%</td>
</tr>
<tr>
<td>Per-Word Rate</td>
<td>EUR 0.09</td>
<td>EUR 0.05</td>
<td>EUR 0.04</td>
<td>EUR 0.03</td>
<td>EUR 0.03</td>
</tr>
<tr>
<td>Total Cost</td>
<td>EUR 2,037</td>
<td>EUR 1,132</td>
<td>EUR 905</td>
<td>EUR 679</td>
<td>EUR 679</td>
</tr>
</tbody>
</table>
Demo

Different PE effort estimations
Model architecture
QUEST’s usage contexts

1. Post-Edit Effort Prediction
2. MT-Recommendation Tool
3. MT-Profiling Tool
4. MT Quality Document Sorting
5. MT Quality Threshold Plug-In
Parting words

• QUEST can be used on-premise or in the cloud and custom models can be built within hours.

• Our machine learning model also scales to any language pair

• QUEST works as solution for leveraging MT for lesser-resourced languages

• LangTec can provide a free customized trial QUEST model for interested parties
Thank you!

LangTec

Dr. Christopher Reid
Associate Product Owner QUEST

Bundesstraße 28 a
20146 Hamburg
Germany

Phone: +49 . 40 . 180 610 98
E-mail: christopher.reid@langtec.de
Web: www.langtec.de
LANG.TEC

SEMANTIC TEXT PROCESSING