ELG Pilot Project:
European Clinical Case Corpus
E3C European Clinical Case Corpus

A project aiming at the creation of a corpus of clinical cases in 5 European languages

Start date: July, 6 2020
Duration: 12 months
Languages: Italian, Spanish, French, English, Basque
ELG coach: Khalid Choukri

https://e3c.fbk.eu/
Three types of annotations:

- **clinical entities**: pathologies, symptoms, etc.
- **temporal information**: events, time expressions and temporal relations;
- **factuality**: event factuality and assessment of the effect of negation, uncertainty and hedge expressions.

A 25-year-old man with a history of **Klippel-Trenaunay syndrome** presented to the hospital with **mucopurulent bloody stool** and **epigastric persistent colic pain** for **2 wk.**
E3C European Clinical Case Corpus

A project aiming at the creation of a corpus of clinical cases in 5 European languages

Annotations Guidelines

E3C Annotation Guidelines. Deliverable D1 (2020)

Data Collection

<table>
<thead>
<tr>
<th>Language</th>
<th>Clinical cases</th>
<th>Tokens</th>
<th>Tok./doc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Italian</td>
<td>1,323</td>
<td>73K</td>
<td>55.1</td>
</tr>
<tr>
<td>English</td>
<td>9,533</td>
<td>928K</td>
<td>97.2</td>
</tr>
<tr>
<td>French</td>
<td>1615</td>
<td>548K</td>
<td>339.1</td>
</tr>
<tr>
<td>Spanish</td>
<td>1,400</td>
<td>531K</td>
<td>379.27</td>
</tr>
<tr>
<td>Basque</td>
<td>122</td>
<td>26K</td>
<td>214.2</td>
</tr>
</tbody>
</table>

First release (CC-BY-NC)

The E3C Corpus: Data collection, planned January 2021

Publications

The E3C Project: Collection and Annotation of a Multilingual Corpus of Clinical Cases" to appear at CliC-it 2020, the Seventh Italian Conference on Computational Linguistics, Bologna, 1-3 March 2021