CoQEx: Entity Counts Explained
Shrestha Ghosh, Simon Razniewski, Gerhard Weikum
Max Planck Institute for Informatics, Germany

Aggregating context and instance evidences over multiple texts to increase answer coverage and user comprehension

Motivation
➢ Infer a representative count from the numeric distribution over multiple text snippets
➢ Interpret count contexts w.r.t the main answer through semantic groups of synonyms and subgroups
➢ Exemplify through instances by ranking instances from the snippets that are counted in the main answer

System Overview
1. Input user query (live or over CoQuAD)
2. Extract query components using CoQEx
3. Top 50 text snippets from Bing API
4. Extraction and consolidation

Answer with Explanations
Query: how many languages are spoken in Indonesia

Answer inference
Count from weighted median of extracted counts
Context from representative extraction
estimated 700 languages

Extraction and Consolidation
User query + Query components + SE snippets

CoQuAD dataset
➢ how many novels did Agatha Christie write
➢ how many mayors in New York

Count Question Suggestions
Live questions
➢ number of songs by Bob Dylan
➢ how many astronauts have visited the ISS

Explanation by Provenance
Snippets annotated with context and instance candidates

<table>
<thead>
<tr>
<th>Explanation by Provenance (44 Snippets)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model confidence thresholds for candidate selection: Count contexts = 0.5, Instances = 0.4, Threshold not crossed</td>
</tr>
<tr>
<td>1. Indonesia 0.83 recognizes only a single national language, and indigenous languages are recognized at the regional level, although</td>
</tr>
<tr>
<td>2. There are estimated 700 languages 0.71 used in Indonesia, most of which are Austronesian 0.47 languages. The de jure official</td>
</tr>
</tbody>
</table>

Methodology

Demo
nlcounqer.mpi-inf.mpg.de

Contact
ghosh@mpi-inf.mpg.de