Flagship Projects

X.X Question and Answering (Q&A@OP) system – Study and Prototype

Description

Search became ubiquitous via websites and applications for any digital user (e.g. in shopping, entertainment, travel, and even in our cars). Virtual assistants are currently redefining the user experience and instead of just asking for simple search results, people began typing full questions and expecting actual answers from their search bar: they expect it to “work like Google.” Enterprise search must become better and smarter. Users expect applications and websites to provide them direct answers to complex questions on about any topic. This is not possible unless the answers are independently deduced and summarized instead of just curated from existing text index or content.

The Publications Office (OP) is exploring a Question and Answering (Q&A@OP) system that leverages “Artificial Intelligence” (AI), “Knowledge Graphs” (KG) and “Extractive Answers” technologies in order to complement the existing OP Portal advanced search system as well as OP’s chatbot. The system will allow users to ask complex questions in any of the 24 official languages and receive a deduced and summarized answer in return along with the rest of “traditional” search results.

In a first phase, the Q&A@OP proof of concept (PoC) will be limited to English and French interactions support in topics related to EU Law in Force and EU Whoiswho dataset on OP Portal and stored in Cellar, our content and metadata semantic repository. Answers may be derived from metadata and their relations or from content, a paragraph or a sentence within the content.

This PoC should allow the OP to evaluate and validate technical assumptions and the cost-benefit ratio before moving to industrialization of the solution.

Objectives

- Demonstrate that semantically accurate answers:
  - may be extracted directly from metadata held in the Cellar database and their relationships thus leveraging the knowledge graph
  - may be extracted from a paragraph or sentence within the content held in Cellar
  - Combine both above
- Demonstrate that the Q&A@OP System can be called via a RESTful API in order to ask Questions and integrate Answers in another application/system (e.g. OP Portal's chatbotOPIA)
- Understand prerequisites and the cost-benefit model in case of industrialisation on OP Portal
- Understand which AI algorithms, modules or cognitive services could be shared for common benefits between Q&A@OP system and OP’s chatbot, in case of industrialization on OP Portal.

Expected Benefits

<table>
<thead>
<tr>
<th>BENEFITS</th>
<th>RATIONALE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality &amp; Efficiency</td>
<td>✔ Providing highly accurate and contextualized answers to OP Portal users</td>
</tr>
<tr>
<td>BENEFITS</td>
<td>RATIONALE</td>
</tr>
<tr>
<td>----------</td>
<td>------------</td>
</tr>
<tr>
<td>• Provide OP Portal users with knowledge (contextualized pertinent answer to a question) instead of information or data (multiple search results)</td>
<td></td>
</tr>
<tr>
<td>User experience</td>
<td>✓</td>
</tr>
<tr>
<td>Discoverability</td>
<td>✓</td>
</tr>
<tr>
<td>Service improvement</td>
<td>✓</td>
</tr>
</tbody>
</table>

**Challenges**

- Having Google BERT pre-trained for EU domain ([EU Law in Force](https://eu-law.eu) and [EU Whoiswho](https://eu-whoiswho.eu)) specific terms. The use of the models of JRC’s SeTA toolkit, which are already pre-trained (also) on Cellar data, will be explored as an alternative

Apply the technology to our specific content and metadata ([EU Law in Force](https://eu-law.eu) and [EU Whoiswho](https://eu-whoiswho.eu) contacts) - Orchestrate the use of NLP models, Semantic model (BERT or SeTA) and machine learning in order to formulate accurate answers based on the give dataset from Cellar

**Deliverables**

- PoC environment available on OP Portal for testing - Operate in English and French language on [EU Law in Force](https://eu-law.eu) and [EU Whoiswho](https://eu-whoiswho.eu) content.  
- Presentations/Demonstrations to various audiences  
- Report detailing approach and cost-benefit value to move the application from PoC into Production