



EUROPEAN COMMISSION

DIGIT D1

Study on the use of Artificial Intelligence techniques for the electronic access to Commission documents (for EASE and new RegDoc systems)

Proposals for projects from D1

Date: 29/05/2019
Authors: D1

1. CONTEXT AND OBJECTIVES

This study should examine the different possibilities to start using Artificial Intelligence Techniques in information systems such as EASE and RegDoc. EASE being the new information system that should handle the requests for access to European Commission documents. This new system will provide an electronic workflow for this handling. It will improve corporate capabilities to identify similar requests submitted in the past and streamline the communication with third parties. In this area of identification and streamlining, the artificial intelligence techniques could come in use.

The objective of the EASE (Electronic Access to European Commission Documents) project is to ensure the Commission will be equipped with modern, electronic and integrated IT tools allowing the submission and handling of the requests for public access to documents. The solution will cover the public interface for communicating with applicants, the internal workflows within the European Commission, and the consultations of other EU Institutions, Member States and third parties, from the first request of the applicant to the final decision of the Commission. The ultimate goal is to bring the EU decision-making process closer to its citizens. The main objective of the project is to provide an information system that enables streamlining of the access to European Commission documents processes across the different stakeholders. The future system will improve the workflows linked to the submission, processing and preparation of replies to requests for access to European Commission documents. It also aims to rationalise internal workflows and enhance consistency between replies.

Regdoc is the Commission's Register of Documents; it is a public application that allows citizens to look for Commission documents. Together with the document, it displays metadata of the document. In case the document is not publicly available, a web-based form can be filled by the citizen. The current application will be rewritten in the coming years but this new project is at its early beginning so no specific documentation exists yet.

The introduction of AI techniques is meant to enhance the EASE project in the first place, but we would like to implement these tools also for the new RegDoc project as well as for the other registers.

An example of an AI technique is Doris, a text-mining tool, used in the information system Better Regulation Portal.

██████████ has sent us a first document for review in the light of potential AI applications (*Regulation 1049/2001 Excerpts from relevant case-law and other interpretative tools*). Key elements of our understanding are discussed in point 3.

2. Meeting inputs from D1:

1. A first list of potential AI (Text Mining) techniques that could be applied (tentative):

- a. Predictive model that will score each new EASE application from a citizen in terms of its probability to be

- accepted/ **accepted under conditions** / refused/ suspended/... **or any other categories for prediction that could be created if deemed relevant by the business stakeholders.** The predictive model would be based on the analysis of past decisions (data on past applications linked to the associated decision would be required).
- b. Search engine tool that would allow to collect/retrieve all documents related to a specific requested topic (chosen / defined by the applicant) across the different registries. Related to need 1 N1 from project charter.
 - c. Request assignment: a predictive model could be created that will automatically define the relevant service for the newly registered request. Machine learning model could automate the assignment of new requests. Related to Need description N21 from the Project charter
 - d. Automatic detection/identification of entities (locations, persons, Administrations) within documents using POS tagging. Related to Feature 10 F10 from the project charter
 - e. Topic modelling to define relevant categories of applications, also relevant categories of commission replies. F10
 - f. Creating a similarity metric analysis between cases and add a visual interface to it in order to see common cases as a connected graph.
2. D1 will engage in gathering information about relevant data inputs (texts, application forms, Commission decisions), verifying availability and quality of these data for the purpose of developing AI models.
 3. In concertation with all project stakeholders a relevant POC case will be selected and on this basis a MOU will be written and agreed by all parties.

3. Document first review “Regulation 1049/2001 Excerpts from relevant case-law and other interpretative tools:”

We will need some clarifications upon this document, but to our current understanding, we could summarize our comprehension as such:

We can see, in the left column, conclusions/ interpretations or outputs coming from court/tribunal on different requests (cases). The left column gives us details on the outcome and give a link to the case.

We could potentially use these cases as inputs for a model and use the Tribunal decision /outcome as target variable for each case.

Of course, for each case, the target variable would have to be redefined or relabelled to definite classes like (accepted/refused/ unclear). This could take some time to manually label each outcome/decision from the court.

Moreover some specific cases are linked to more than one outcome/decision.
See

Case “T-189/14” for example. A certain amount of manual “interpretation”work to define or label the categories of outcome has to be envisaged.

Does a document exist containing a clearer link between case number and outcome type exist ?

How can we reliably label the final decision/outcome of each request case?

4. Next steps:

1. Better understanding of available data/text documents (1 month of interactions (meetings, emails) with SEC Gen. representatives)
2. Acquiring data on our Data Platform (2-3 weeks)
3. AI analysis (3 - 5 weeks).
4. Delivering results of the POC in a shiny application (2 weeks)