JURISINNOV'

Groupe P13-G04

Team Members :

Adam Lahrioui [1], Aimane El Mrabet [1], Chaymae Dahhassi [1], Nada Elhouari [1], Ouanzougui Abdelhak [1]

Supervised by:

M.Mohammed El Rhabi [2], M.Jean-Pierre LLORED [3], Mme.kawtar zerhouni [4]

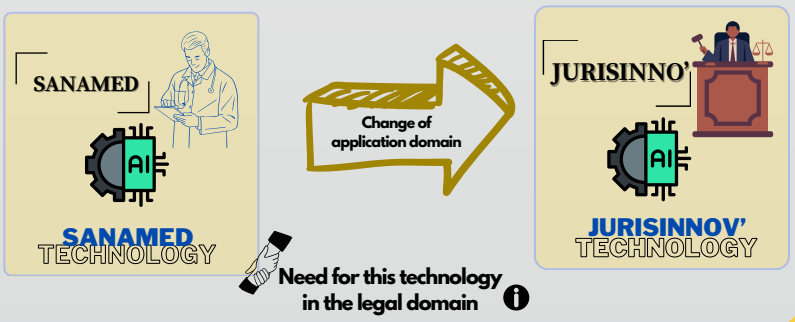
[1] 2nd year of engineering studies at Ecole Centrale Casablanca, [2] Prof. Dr., Applied Mathematics, Data Processing, [3] Associate Professor in Humanities and Social Sciences, [4] Ph.D, Assistant Professor at ECC

aimane.elmrabet@centrale-casablanca.ma

+212 636-388931

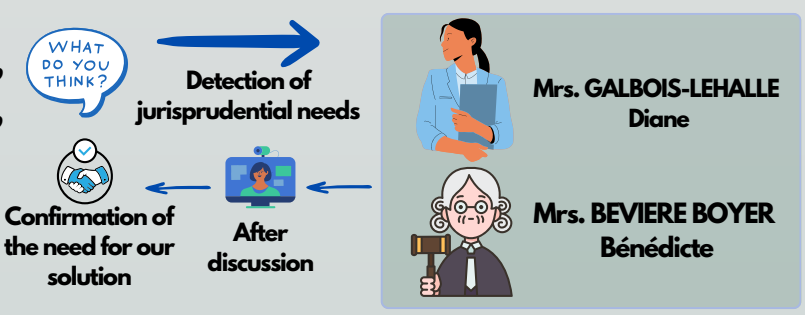
1 Introduction & Motivation

In our intrapreneurial project, we aim to apply a technology initially developed for medical records to the domain of jurisprudence. This technology, created by the Sanamed team, extracts patient metadata and generates medical document summaries. Our objective is to explore and customize this solution to enhance the efficiency of processing legal documents, ensuring increased accessibility for legal professionals, particularly those pertaining to jurisprudence.

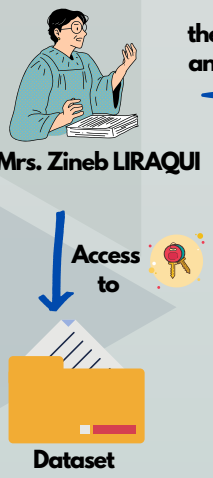


2 Analysis & description of the group work

To implement our solution successfully, we began by meeting with legal professionals, including jurists *Mrs. GALBOIS-LEHALLE Diane* [1] and *Mrs. BEVIERE BOYER Bénédicte* [2], to grasp the needs of the jurisprudential realm and assess the adaptability of technology.

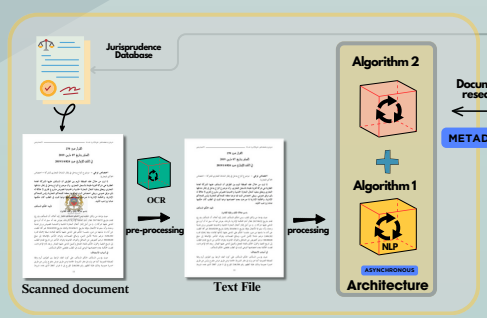


Following that, we contacted *Mrs. Zineb LIRAQUI* [3], a lawyer managing an online firm [4]. After negotiations, she provided us access to a database containing a collection of scanned files on which we conducted our work.

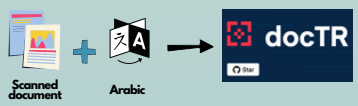


To maximize the use of the database, collaboration with third-year students was essential for pre-processing the dataset. They conducted a study to identify the optimal technology for converting scanned documents into editable text files, determining that Optical Character Recognition (OCR) technology, particularly docTR , excelled in speed and efficiency. Working together, we developed two algorithms for the asynchronous architecture, allowing simultaneous and independent operation. The first algorithm serves as a tool to extract metadata from our text files, while the second allows users to search for documents and retrieve metadata.

3 Methodology



1-Pre-processing

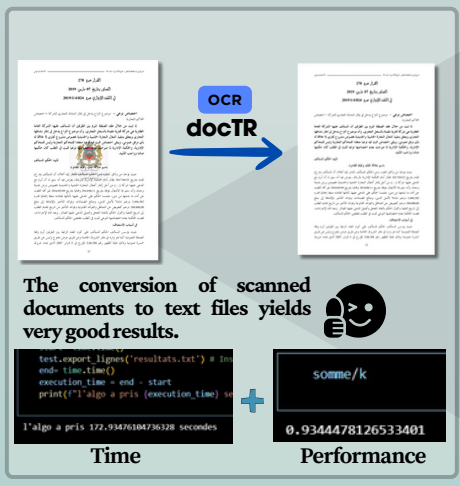


2-Algorithm 1

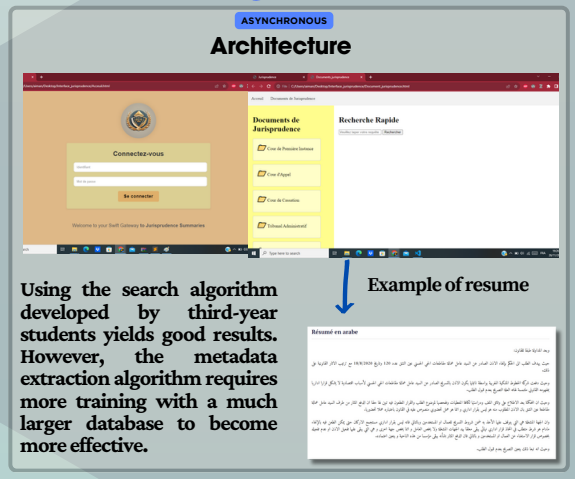
Algorithm using NLP method [5]

3-Algorithm 2

4 Results



The conversion of scanned documents to text files yields very good results.

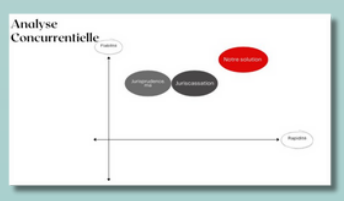


Using the search algorithm developed by third-year students yields good results. However, the metadata extraction algorithm requires more training with a much larger database to become more effective.

5 Innovation & Business targets

Innovation

- Providing a complete solution for digitizing jurisprudence data to improve the quality of decisions made by jurists.
- A comprehensive database of jurisprudence from the past 23 years.
- An automatic summary of each jurisprudence.
- Optimal search speed.



Business targets

- Monthly or annual subscription for platform usage.
- Fees for additional services such as personalized training or specific integrations and generating extra summaries of Law documents.
- Strategic partnerships with institutions for large-scale projects

6 Bibliography

1. Diane Galbois-Lehalle | LinkedIn (no date). Available at: https://www.linkedin.com/in/diane-galbois-lehalle-07b226185/?originalSubdomain=fr.
2. Bénédicte BOYER BEVIERE | LinkedIn (no date). Available at: https://www.linkedin.com/in/b%C3%A9n%C3%A9dicte-boyer-beviere-8868b519/?originalSubdomain=fr.
3. Zineb Laraqui | LinkedIn (no date). Available at: https://www.linkedin.com/in/zineb-laraqui-avocat/?locale=en_US.
4. 'Cabinet Zineb Laraqui – Avocat au barreau de Marrakech' (no date). Available at: https://www.zineblaraqui.com/ .
5. 'AraBERTv2 / AraGPT2 / AraELECTRA' (2024). AUB MIND. Available at: https://github.com/aub-mind/arabert.