

INTRODUCTION

- UN ReliefWeb has 20-30 editors who manually process the PDFs produced by UN and NGOs.
- It takes about 10-15 minutes for an editor to process each PDF. An average of 33 reports per person in a 7-hour shift.
- Increase in humanitarian aid due to climate change, and increase in disasters, means the no. of reports for its programs will increase by an order of ~10 times.
- ReliefWeb wants their editors to continue to oversee the processing of PDFs and not completely replace them with AI system.

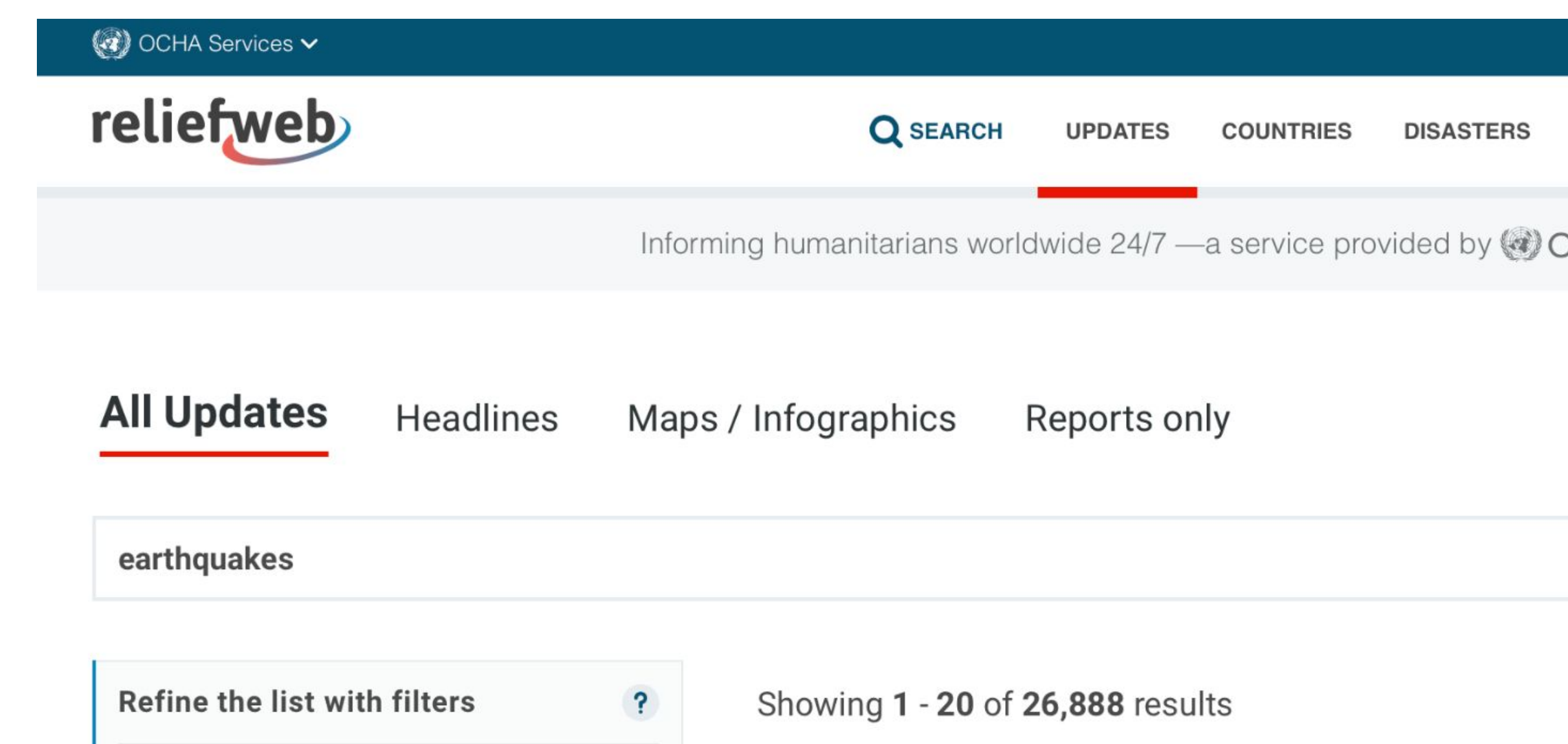
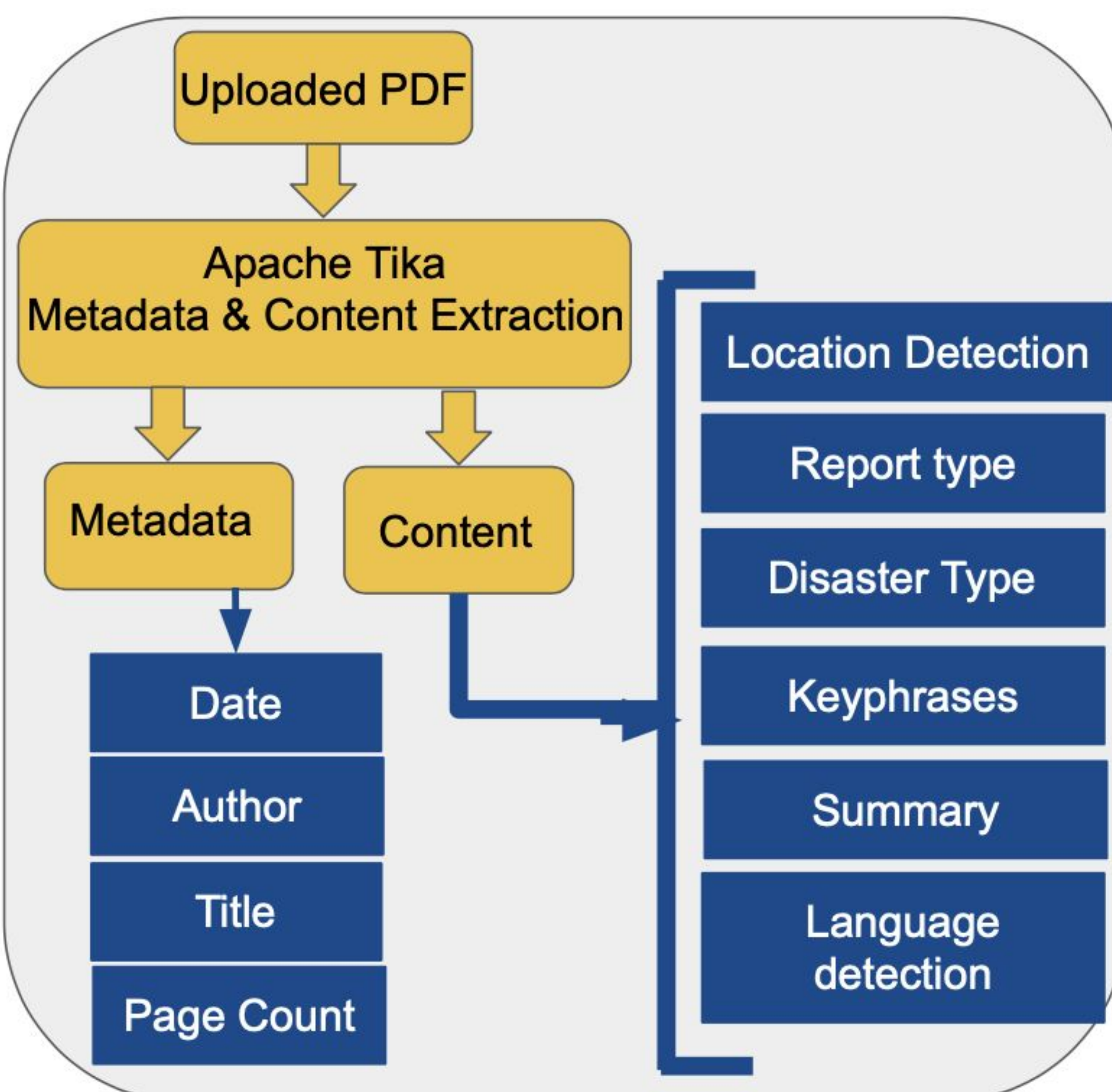
AIM

To create a document processing tool that would assist the editors at ReliefWeb (or any other NGO) to have a **faster and more efficient workflow**.

OBJECTIVE

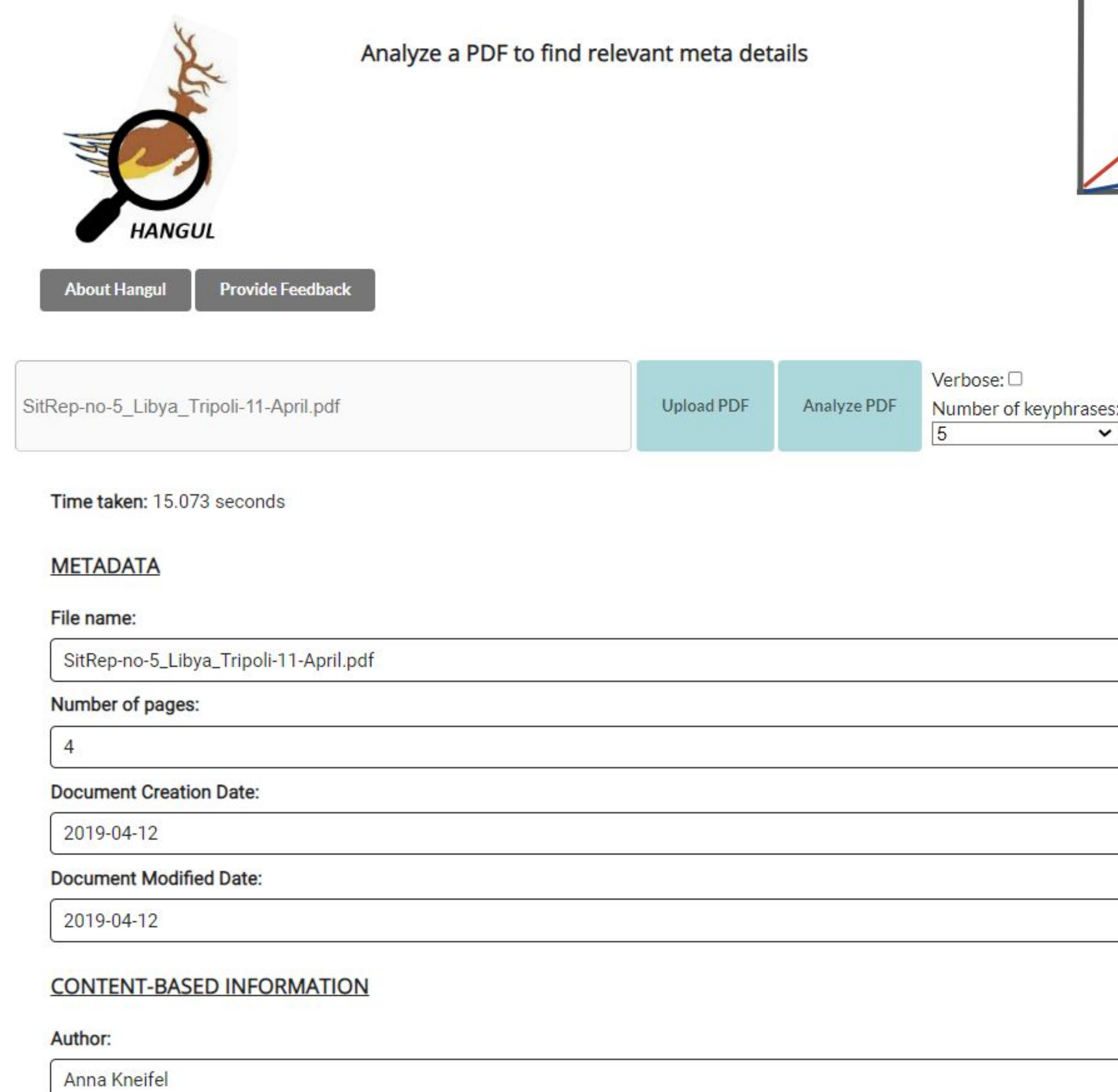
- Reduce time to process PDFs by employing NLP techniques to extract relevant information.
- Create a user friendly web-interface for easy access to the system.

METHOD



ReliefWeb website

Hangul Interface



Analyze a PDF to find relevant meta details

Time taken: 15.073 seconds

METADATA

File name: SitRep-no-5_Libya_Tripoli-11-April.pdf

Number of pages: 4

Document Creation Date: 2019-04-12

Document Modified Date: 2019-04-12

CONTENT-BASED INFORMATION

Author: Anna Kneifel

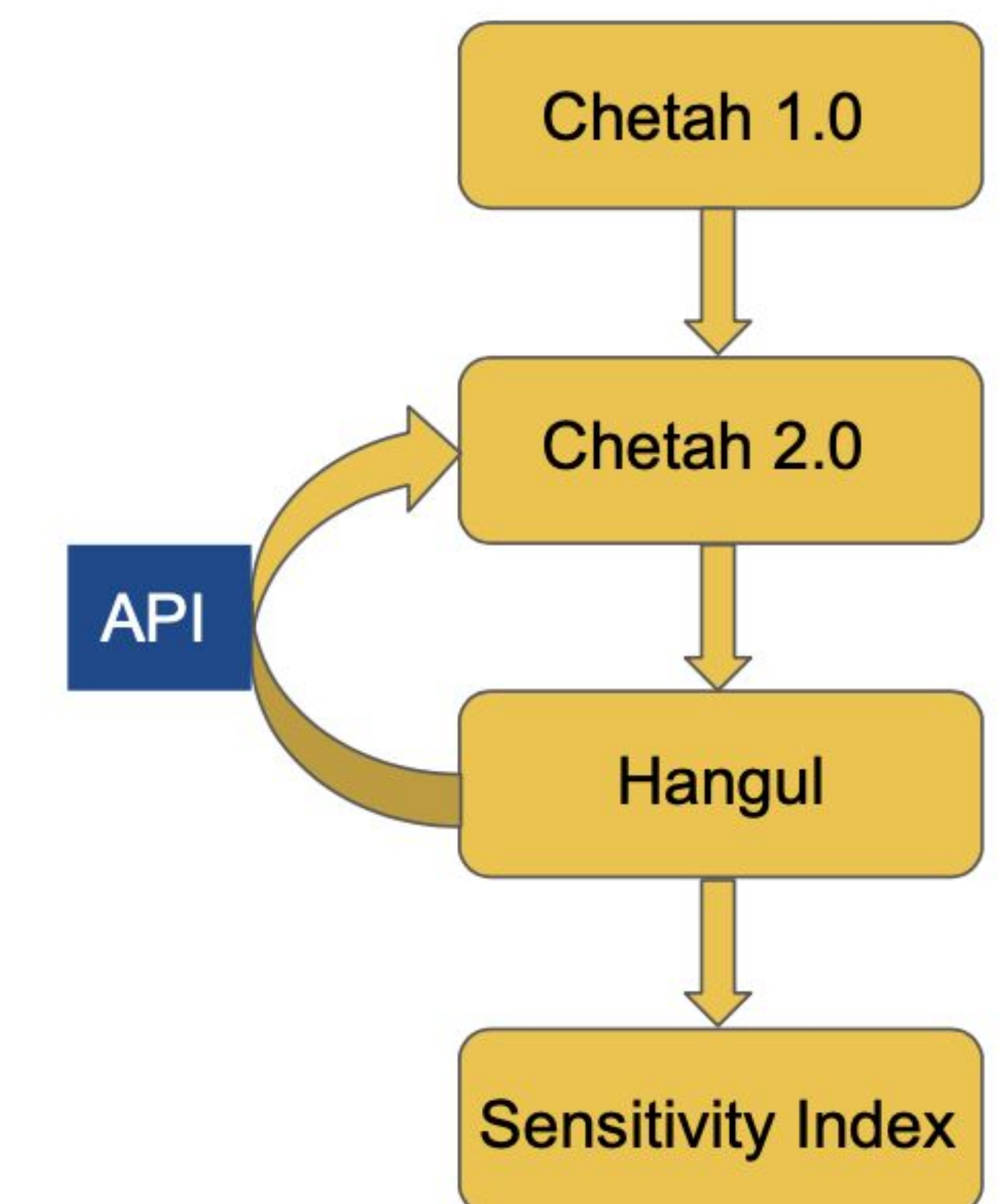
Try Hangul here:



We welcome your feedback and ideas.

Using NLP techniques such as: **Name Entity Recognition**, **Keyword Extraction**[1] and **Text filtering** to extract relevant metadata and content information from an uploaded PDF document.

HANGUL ECOSYSTEM



RESULT

PDF processing time is reduced from around 10-15 minutes to an average of one minute, increasing PDF processing capacity of editor by ~13 times in a 7-hour shift.

NEXT STEPS

- Markdown format for summaries.
- Automated theme detection.
- Sensitivity index.

ACKNOWLEDGEMENT

We would like to thank Edward G. Happ, the Data4Good Center and the NGO Search Engine MSI Capstone team (Fall '22).

REFERENCES

- [1] Campos, R., Mangaravite, V., Pasquali, A., Jatowt, A., Jorge, A., Nunes, C. and Jatowt, A. (2020). YAKE! Keyword Extraction from Single Documents using Multiple Local Features. In Information Sciences Journal. Elsevier, Vol 509, pp 257-289. Pdf
- [2] ReliefWeb (<https://reliefweb.int/>)