Orgpedia - Information Extraction for Public Good

Orgpedia Foundation is a Section 8 (non-profit) company based in Pune, India. Orgpedia intends to make it easy for people to understand the workings of government organisations; this is for both - people outside the organization as well as those inside the organization.

To achieve this broad goal, Orgpedia is focusing on solving two technology problems

1. Build information extraction technology for extracting information from government orders.
2. Present the information in a way that is easy to consume for even a common man.

Orgpedia Projects:

1. **Chat friendly extraction of Maharashtra Government Orders**: Orgpedia has converted the last five years of Maharashtra Government orders (GRs) that are in Marathi PDFs, to English and Marathi text files. They are in a format that is friendly for chat LLMs to consume. The pipeline does table structure extraction, translation and paragraph extraction. There are about 50,000+ orders in the repository and the data is available at https://github.com/orgpedia/mahGRs. The repository is updated weekly.

On this data there is a simple chatbot that answers specific questions video, the chatbot uses OpenAI's apis. There is also a website where weekly summaries are published. The summaries are generated based on district and department and annotated with additional information like, funds allocated, category of the order. The goal is to provide rich annotations so that users can sign up for information that is relevant for them. https://mahsummary.orgpedia.in/

2. **Chat friendly extraction of Gazette Data (WIP)**: We are currently processing Gazette data from Maharashtra and Karnataka, both these datasets are predominantly in Marathi and Kannada, respectively. The goal is to process the gazette data and split it in smaller relevant documents so that they can be easily consumed by a specialised chatbot.

3. **Deep Extraction on Government Orders**: The idea is to convert a government order into a JSON object so that it can be added to a database and queries can be run on the database. We have been focussing on Government Orders that relate to posting and transfers of government officers. The goal here is to extract the information and build an org chart for that department.

Currently we have only released Org-chart for Indian Cabinet Ministers. The site contains the cabinet information and the hierarchy for the last 75 years, starting with the country's first order to the date, the site is available in 22 Indian languages. Code: https://github.com/orgpedia/cabsec/.

Help needed for Orgpedia:

Orgpedia in near term wants to focus on two problems 1) Deep extraction on the orders and gazette data and 2) Chat interface in native Indian language for a specific department/domain.

To achieve both these goals Orgpedia wants to build a specialised Indian language model specifically for government data and for these two use cases. To achieve this goal, it will require converting large amounts of government data in Indian languages to text files and further computing resources for building the model. We need help on both these fronts

1. Cloud credits for running the data pipeline.
2. Computing Credits for building the specialised language model.

All of Orgpedia's data & code is in open source and available at https://github.com/orgpedia. For more information please email Mukund Deshpande (mukund@orgpedia.in)