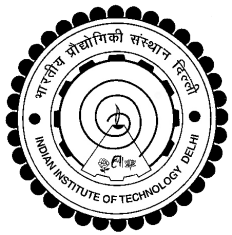
Web application for crowdsourcing of Image description and Preliminary integration of previously made RAVI tools

Prashant Ranwa 2018CS50414, K Laxman 2018CS50408

COD892 Major Presentation

Advisors: Prof.M.Balakrishnan, Prof.Volker Sorge, Abhishek Baghel
Supervisor: Prof. Vireshwar Kumar

January 04, 2023



Previously done work

Automated tool for converting inaccessible STEM documents into a fully accessible format such as EPUB

1. PDF is parsed to first convert digitally generated PDF into raw HTML.
2. Different analysis are done on this HTML example Math Detection, Text analysis, Table detection, Object Detection, Radical Sign Detection etc .
3. The resultant file is passed through a classifier to classify the content into text, equations, and text inside equations. This enables us to treat text and equations separately and adapt the document structure analysis algorithms accordingly
4. Further, the document is processed to extract various structural elements like a paragraph, list, table, caption, header, footer, page number, heading, and diagram.
5. Finally, the extracted information is encoded and saved as an accessible HTML document.

Relevant Text Extraction

Automated Generation of Image Descriptions for Visually Impaired by Gopi Veerendra. Providing alternate text description for images greatly helps understanding of context for visually impaired, and an effective template to write such descriptions was designed in Template based approach for augmenting image descriptions by Chahal. Algorithm to fill this template:-

1. Caption Extraction.
2. Corresponding image extraction.
3. Extracting labels and their position in image using Google Cloud Vision API a Computer Vision tool.
4. Extracting relevant text from nearby text using attention based models "Sentence Transformers" an NLP tool.
5. Finally used summarization(NLP tool) tool to summarize the extracted information i.e caption + labels in the figure + extracted relevant sentences and caption + extracted relevant sentences.

Functioning of web application made for conversion of PDF's to HTML

1. Login using username and password.
2. Now the webpage shows all files that were previously uploaded. A new file in pdf format can be uploaded here.
3. Now in the backend the pdf is converted to Html and saved.
4. If the conversion from pdf to Html is successful User will be able to **view, edit, save, download** the HTML source code and view **live preview** of HTML.

Problem statement

RAVI

1. **No image descriptions**This gives a refined html as output on which screen reader can be used by visually impaired people to hear all text present and understand the pdf but one of the biggest impediments for the visually impaired to understand STEM material is their inability to see figures which describe experiments or provide illustrations.

2. **Several different components** Different people had contributed code for a different type of contents like background analysis, mathematics analysis, Image analysis. Previously these code were tested individually and are now being combined together.

Problems faced in Relevant text Extraction

1. This code helps in providing illustration But the short summary generated from the NLP technique is still not as good as a human interpretation. The description of an image may not be good enough so we need to manually check it and improve the description of the image if needed.
2. The final description is not being saved anywhere and also the output is not in the form of HTML.
3. The code only works if an image has a caption, labels are present in image and some related text to both of these is present in the content close to image.
4. The code does not do any image analysis other than label detection.

Solutions

Made application for crowdsourcing of Image description

1. Initially input is taken as an HTML document in this application.
2. Access and view all the images present in HTML by scrolling.
3. For every image it shows an option to input and save the understanding of an image as Alt text.
4. The Predefined Alt-text will be shown in the input box either it can be updated or left same.
5. After saving all descriptions of images, either new Html can be downloaded or saved it in original location. This new Html contains the added description in the form of **alt text** in their respective img tag.

Changes made in Relevant text Extraction

- The final code takes the location of Html file as input and saves a new html in the original location. This Html is containing automatically generated image descriptions as alt tag property of that image respectively.
- Removal of all unnecessary parts in code which were present for testing heuristics and edited the code to make it more generalized.
- Added **Azure Computer Vision API** to do image analysis and get a single line understanding of images.

Integration

- Integrated all modules like Caption Detection, Column Detection, Mathematical Formula Detection etc, in RAVI code that were individually tested.
- Integrated **Image Description Adder UI** with RAVI code.
- Integrated RAVI code with the new updated **Relevant Text Extraction**.

Working of final code

The final code works in 3 steps:

1. First the pdf is converted to Html in the backend.
2. In the backend itself relevant text extraction code runs on this html and adds the image description in alt-text and outputs the returned Html to the database connected to frontend.
3. Using add-alt button in the front-end we can edit and make changes in the descriptions of images.

Future work

Alt text limit The Alt text helps screen-reading tools describe images but Screen-reading tools typically stop reading alt text at **125 characters**, cutting off long-winded alt text at awkward moments.

Image Descriptions A detailed explanation of an image that provides textual access to visual content; most often used for digital graphics online and in digital files. There is no upper limit for characters in this as this description can be inserted below the image in Html which the screen reader views in the same way as normal text.

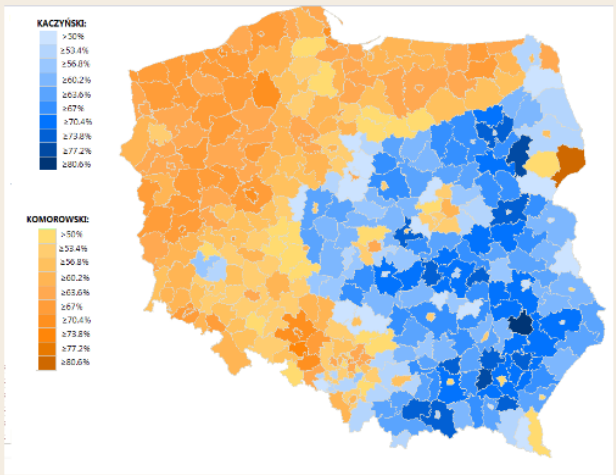


Image description: A map of Poland with voting districts outlined. The map outlines the results of the 2010 Presidential elections. The majority of the left side of the map is various shades of orange and the majority of the right side is various shades of blue. A key to the left outlines the shades of blue that represent the percentage of votes won by Komorowski and the shades of orange that represent the percentage of votes won by Kaczyński.

Alt text: A map of Poland with voting districts filled in with blue or orange.

Improving Automated Generation of Image Descriptions and Robustness of code

- One can classify the images into various categories like experimental setup, geometry, histogram, pie-chart, ray diagram, biological, etc. Further, we can develop category-specific codes to generate descriptions by using Digital Image Analysis and Computer Vision.
- Improving the identification and cropping of the images within the content.
- Exploring the **Poppler's pdftoHtml** for improved HTML output.
- Improve segmentation of pdf's to identify headings, sub-heading, maths, images, references, captions and tables.

Conclusion

The new additions we have made in the previously done work and the integration of the components has made us closer to our final goal which is to make it possible **to automatically convert all form of documents in a new standard compliant EPUB format document which can be understood by visually impaired students just by hearing in the same way as a normal student understands it by reading.**