Taishi Nammoto

QBA 465

Description of "Hawaiian Texts Analyzer" Website

I. The Use

    A. By clicking on an upload button, an image selection display pops up.

    B. By clicking on a submit button, the selected image is uploaded to the database of the website.

    C. By uploading an image of a scanned Hawaiian written document, this website shows Hawaiian words on the image and the English translated texts.

    D. By clicking on a title of the uploaded image, the image is automatically downloaded from the website.

    E. By clicking on a delete button, the uploaded datasets including the image, texts, the added date, are permanently deleted from the database of this website.

II. The Structure

    A. Two main systems are implemented on this website. One is an Optical Character Recognition System (OCR) system. The OCR system detects Hawaiian texts from an image of a scanned Hawaiian document (Fig. 1). Another is a Machine Translation system. The translation system plays a role to translate Hawaiian to English.
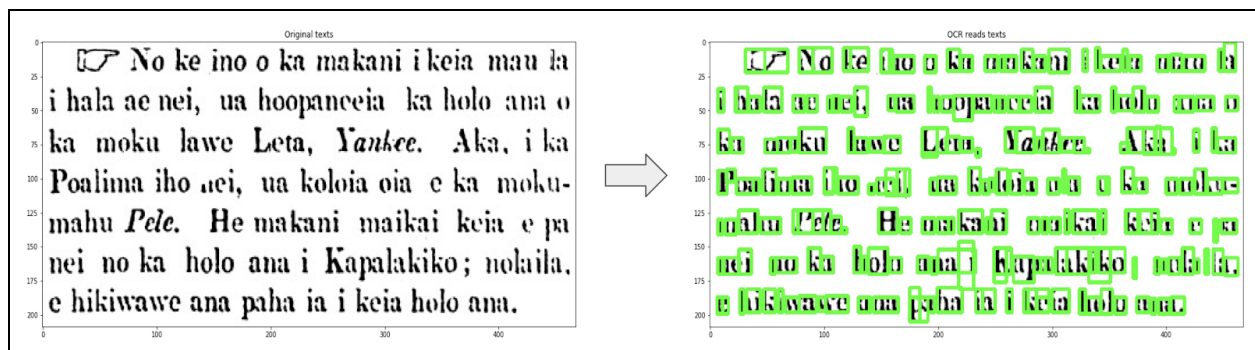

Fig. 1: Visualization of the OCR system

III. The system information

    A. Terratrac OCR is a trained OCR model which is a decentralized software development model designed to encourage open collaboration. By turning parameters of the recognition system, the system is specifically adjusted to read Hawaiian letters from images. The custom OCR is implemented on this website for converting image inputs into text outputs.

    B. UNdreaMT is an open source implementation of our unsupervised neural machine translation system. I plan to train this model by providing a sufficient English-Hawaiian paired corpus. The trained model will be implemented on this website.