

Position: Research Assistant for Topic Modelling Analysis

We are five researchers across the University of Cape Town and the Swedish University of Agricultural Sciences working on a project on maize in South Africa, looking into aspects of agrarian change, relational theory, local and Indigenous knowledge, and historical-future trajectories. We are exploring the possibility to analyse journal articles dating back to the early 20th century using topic modelling. As we have the thematic expertise, but lack the practical coding and computational skills, we are looking for a research assistant to help us.

For a first step of this project, we are looking for someone who could build a database for us, doing the following:

- Scrape PDFs from an open access repository containing journal articles from the 1900's up until the 1980's. Some articles have poor visual quality, so the candidate will need skills to deal with this. An initial indication is that it could be up to 10 000 PDFs, of varying length.
- Convert text to machine readable format, organise it to allow for topic modelling at a later stage.
- Perform basic analysis of word frequencies etc, based on input from our team on themes and search strings.
- Provide input and recommendations to the process, regarding for example data formatting and preparation to allow for topic modelling at a later stage.

We are looking for someone who is skilled and reliable, and is able to take lead on these tasks as we ourselves in the team don't have the technical skills. That is, the person must be able to work independently on the individual tasks - however of course with guidance from us.

Since we also want to be able to do a more in-depth topic modelling if we have budget for it, we ideally want a person who could also do that.

We would like to start as soon as possible, but considering the holidays coming up, perhaps realistically at the beginning of next year. Ideally the candidate will be able to work full-time during this first step, if that is however not possible, we can adjust the timeline.

It will be a paid research assistant position, with a South African salary with the exact tier depending on the qualifications of the candidate. There will of course be acknowledgement in any publications, and could also entail co-authorship depending on how the collaboration unfolds.

The work can be done remotely, or with us at the Department of Environmental and Geographical Sciences at the University of Cape Town (for anyone already living in the city).

In case of interest, please contact Dr. Kajsa Resare Sahlin at kajsa.sahlin@uct.ac.za
Please describe your interest and previous relevant experience, as well as attach a CV if relevant. We will then begin by scheduling a brief meeting, and depending on fit, candidates will then be asked to provide a suggestion for a plan to execute the tasks listed above, including an estimate of the time required.