JOHANNES KEPLER UNIVERSITY LINZ

Institute of Computational Perception

DATA FRIENDLY SPACE

Student Research Assistant Position

January 2021

About Data Friendly Space

Data Friendly Space (DFS) is a 501(c)(3) non-profit organization based in the United States with a presence in Switzerland and Nepal. DFS' guiding principle is to improve information management and analysis capacity, tools and processes in the humanitarian and development community to enable better informed and more targeted assistance.

DFS staff is composed of experts from the humanitarian information management and analysis field who specialize in real time secondary data review and build humanitarian applications that support fast extraction of information from large volumes of unstructured data.

DFS also focuses on creation of data centric web applications, websites and mobile applications to support humanitarian organizations in extracting the most insights from their data and execute their missions. When building software, DFS focuses on the intersection between data automation processes powered by Artificial Intelligences and human knowledge and skills, in particular when one can help the other to execute analysis.

DFS is closely linked to the DEEP project, with several of its staff members having contributed to its foundation and development. All DFS contracts that involve the use of DEEP provision budget for DEEP maintenance and further development.

Data Friendly Spaces (DFS) core principles include:

- Technology for humans, by humans: Understanding user design process and developing interfaces that facilitate interaction with your data, at the speed of thought.
- Technological responsibility. While we are passionate about new technologies, we are also realistic about how much can be expected and achieved from them. DFS takes a forward-thinking approach to review and select tools and approaches that are the most appropriate to solve given problems. We firmly believe that technology is best used when it enhances brain capabilities, not when it tries to do our thinking for us.
- Creating sustainable solutions. The ultimate success of DFS come when its services are no longer needed and humanitarian organizations are able to make the best of their data without external support. DFS aims to empower organizations with self-sufficient methodologies, technologies and workstreams through appropriately tailored solutions and capacity building. There is nothing more rewarding for us than seeing projects we supported still producing value years after we have stopped supporting them.

DEEP Platform

DFS is the technical host of the Data Entry and Exploration Platform (thedeep.io), a tool used by humanitarians all over the world to monitor and assess crises. DEEP is a tool used by analysts to manually tag information extracted from multiple documents and sources according to the classes and taxonomies of a corresponding analytical framework. Tagging this data leads to the structuring of large volumes of information that enables effective analysis of the humanitarian conditions of the populations of interest and empowers humanitarians to identify information gaps and to provide sound recommendations in terms of needs assessment strategies. The functionalities of DEEP are enhanced with **DEEP Language** – an under-development module which exploits Natural Language Processing and Machine/Deep Learning approaches to address the following topics:

- Automating time-consuming and laborious processes manually performed by analysts
- Maximising the impact generated by the use of DEEP in the management of information during crises
- Integrating with other humanitarian information management platforms to provide custom NLP services

The innovation of DEEP relies upon leveraging recent advances in Natural Language Processing and Deep Learning to automate the process of secondary data review. In order to achieve this, DFS establishes partnerships with academia and conducts research projects with leading universities such as EPFL and JKU. These projects are conducted by Master students or PhD candidates, supervised by the DFS staff as well as professors and, when possible, result in the publication of a paper and actual deployment within the platform.

More information on the DEEP and how it is being used can be found here:

- DEEP's Website
  https://www.thedeep.io/
- DEEP's YouTube channel (tutorials and demos)
  https://www.youtube.com/channel/UCO3naDryeQIFny6BsEJwCaA
- IFRC's tutorial for conducting Needs Assessments with DEEP:
  https://deephelp.zendesk.com/hc/en-us/articles/360041904812-4-DEEP-Using-the-DEEP-Platform-

Position

The **student research assistant** will explore the latest advances in Natural Language Processing and Deep Learning and apply them to the various aspects and challenges concerning the DEEP Language platform. The position offers an employment of **10-20 hours per week** or a period of up to **6 months** at the Computational Perception Institute under the supervision of Dr. Navid Rekabsaz. Prolongation of contract could be possible in the case of the availability of additional funding. The scope, time, and complexity of the defined project for the position fits to the work required for a **Master thesis**. It is therefore ideal, if the candidate conducts and concludes his/her Master thesis throughout the project.

Skills & Knowledge:

- Good knowledge of Python Data Science and Machine Learning libraries
    - The required expertise is provided by the *NLP* course at JKU
- Good knowledge of Deep Learning with an NLP focus
    - The required expertise is provided by the *NLP with Deep Learning* course
- Good research writing / presentation skills (report writing, PowerPoint presentations)
- Interest in the humanitarian and the applications of technology for the humanitarian sector
- Familiarity with humanitarian settings and the United Nations/NGO systems is desirable

Overall area of some potential topics (non-exhaustive list):

- **Summarisation:** DEEP users often perform bulk uploads of documents to add them as sources for their projects, but not all documents might be relevant for their analysis and reading through all of them can be very time-consuming. Abstractive and extractive summarisation algorithms could significantly reduce the time spent on reading these documents and could help analysts target the most relevant information therefore improving the speed and quality of their analysis.
- **Text classification:** analysts that use DEEP for secondary data review manually tag excerpts from the sources they have uploaded in the platform according to the analytical framework they have chosen. The tagging process is laborious, and taggers need to go through extensive training to understand how to correctly tag the excerpts. Supervised and unsupervised classification could help in the identification of the most relevant excerpts in large documents and could provide recommendations to the taggers on how to appropriately tag those excerpts.
- **Question answering:** DEEP users often need to find very specific information in large volumes of unstructured data (i.e. the number of people displaced in a crisis or the number of COVID-19 related deaths in a country). A question & answering algorithm that automatically outputs the excerpts from documents associated to a

pre-defined set of ‘questions’ could significantly reduce the time analysts spend looking for a specific type of information.

Language: Fluent in written and spoken English.

Advantages and learning experience (outside the context of CP institute and JKU):

- **Research publication:** Based on the results of the research, there will be the possibility of publishing a paper in a major Journal (previous work has been submitted to the European Chapter of the Association for Computational Linguistics). It also provides opportunity to present at conferences around the world.
- **Networking and exposure:** DFS partners include major humanitarian organisations such as the IFRC, UNHCR, UNICEF, UN OCHA and many others. Students will get a chance at networking with all kinds of professionals from the sector through participating in meetings, webinars, presentations and discussions. DEEPL has also been involved in hackathons such as CERN’s Port and AMLD.
- **Impact:** DFS has positioned themselves as the go-to partner for humanitarian organisations when it comes to developing and implementing innovative technology solutions for the humanitarian sector. Students will have the opportunity to apply their technical knowledge to build solutions that contribute to improving humanitarian outcomes during crises.

In order to apply, please send your updated CV and final university script (list of passed courses and their notes) to navid.rekabsaz@jku.at. In the email, mention your stage of study, and a potential start date.