

# **An Open System to Support Analyzing Scholarly Documents**

**Dr. Terry Lima Ruas**

(University of Göttingen)

## **Abstract:**

Natural language processing (NLP) has made remarkable progress recently. Large language models are influencing the lives of hundreds of millions of people and are expected to have substantial industrial and societal impact. The development of this technology and those who control it leads to shifts in power relations. It is important to reflect on where research is going, who has the power, what they value, and how we can democratize it. Answering these questions gives important cues on where we are right now, how today compares to our past, and where our field is headed. Our investigations also show that the industry presence among NLP authors has been steady before a steep increase over the past five years (180% growth from 2017 to 2022). More than 72% of US faculty in ACL 2023 have received industry funding. While these results already reveal marked shifts and give us agency to influence these trends, it is a snapshot of the situation now. To enable us and others to self-evaluate, we seek to develop a system at the University of Göttingen that combines different methodologies to scholarly. The web-based solution will make tens of millions of research papers easily accessible and provide various visualizations on the volume of papers, cross-field citations, temporal citation patterns, diversity of geolocations, publication language, gender, etc. Our system will help inform conference organizers, program chairs, committee members, editors, and mentors to analyze trends in their fields and subfields to help them make educated decisions.