

# NLP03 Identification of UN-Sustainable Development Goals (SDGs)

## Background

In 2015, the United Nations Member States (UN) agreed on 17 sustainable development goals (SDG) „for peace and prosperity for people and the planet, now and into the future“.

We want to support this endeavour by cluster research publications based on their relevance to these SDGs. The goal is to make relevant research findable and to strengthen access and visibility to such research.

## Goal

- Evaluation of the existing approach and enhancement of implementation

## Tasks

1. Implementation
  - Refactoring of existing R code
  - Build a Python implementation
2. Data science
  - Evaluation of used methods
  - Enlarge data set and compare results



Daniel Kurzawe

kurzawe@sub.uni-goettingen.de



Mathias Göbel

goebel@sub.uni-goettingen.de



# NLP04 Information Extraction from Research Papers for DIGIS

## Background

The objective is to devise approaches for the automated extraction of geochemical data and metadata from research papers and implement them prototypically pipeline for the geochemical data infrastructure DIGIS.

We extract specific mentions of methods from papers. This information can be part of the paper or included in tables or figures. The structure depends on the journal.

## Goal

- A prototype for extraction specific information from research papers

## Tasks

- Compare existing approaches for information extraction for a given set of papers
- Implement a prototype
- Draft a data pipeline



Digital Geochemistry Infrastructure  
for GEOROC 2.0

Daniel Kurzawe

kurzawe@sub.uni-goettingen.de



Mathias Göbel

goebel@sub.uni-goettingen.de



# MR06 Non-Statement View: A Set-theoretic Description of Theories

## Background

The non-statement view (or structuralistic theory concept) uses set theory to describe a scientific theory through its internal structure and in conjunction with larger theory networks. This philosophical framework allows a generic theory description.

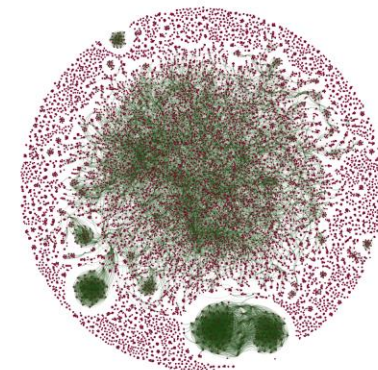
There are several publications about structural reconstructions of scientific theories, e.g. Newton particle mechanics. Due to its set-theoretical nature, a (semi-)automatic approach for such a reconstruction might be possible. This project explores this approach.

## Goal

- Extraction theory components theory and transformation into a structural theory description

## Tasks

- Explore concept for a semi-automatic reconstruction process
- Mapping semantic and concepts
- Build a theory network
- Implement a parser and transformer for a specific domain



Daniel Kurzawe  
kurzawe@sub.uni-goettingen.de



# PD07 Textual Criticism and Plagiarism

## Background

“The identification of textual variants, or different versions, of either manuscripts or of printed books” ([Wikipedia](#)) is a major task in philology entitled “textual criticism”. The analysis of a single text in different variants starting from the very first sketch up to the latest authorized version is provided with a historical-critical edition. Before the digital age, these editions used an obnoxious amount of signs marking and categorizing these differences, like the Leiden convention has standardized. However, newer visualization technologies provide more and more interactive views to these editions.

What are the shared approaches of plagiarism detection and textual criticism? Can they benefit from each other?

## Goal

- Investigate if/how a software for plagiarism detection can be utilized to deal with a set of documents that represent the same text. Comparison of both methods.

## Tasks

- Input material selection (assisted)
- Data conversion
- Usage of PD software/visualization, publish result using web technology
- Workflow to automatize main steps and to scale up (deal with as many editions as possible)

Transkription

Varianten

Factor Honoratissime, esse, vel ideo quod amissam tui videndi spem utuncque erro animorum ac methodorum consensus, video enim t ego incidere. Superiore anno ad D<sup>o</sup> Marchionem : mihi rationem omnes aequationes differentiales primi differentialis, in quibus abest constanti implens leges adituras: Id nunc tibi quoque innotuisse animadverto. an quaerendi naturam et tangentes curvarum ubi scilicet in aequatione curvae ipsa indeterminata in a multo tempore sum usus, et specimina etiam Itagenio li genus videbatur Ego sic procedi: Si verhi gratia seu  $x \int dx : x^{(3)} \int dy : y$ . Datur ergo  $\log. y$  ex data  $x$  et  $y$ . Porro differentiendo ex aeq. 3 fit  $dy : dx = y, 1 + \int dx : x$ . Ergo habetur et ductio : de Quadratura Figurae res est altioris indaginis. Nam nisi ordinarias quantitates habent ingredientes, immiscet line transcendentaltas seu infinitum. Aliquoties in Actis sodum calculo tractandi operi meo reservaveram, iametsi existem cum Logarithmis, Elegantisima videtur series iudicavi (!) perfectissimas; et quando rem hac reduxi, nihil aae aream quandam dictae figurae exhibet. Quomodo inde

deu... differentialibus) fhh L, erg. LH

quaerendi (!) tangentes curvarum, (a) quarum (b) ubi in a ingreditur (c) naturam ... ingreditur L.

LH

logarithmis, item aliter ex aeq. 3 dx : x = -adx : xx  $\int dxxy = -adx \int dy : y + axdy$  gestic | De quadratura L

L



Daniel Kurzawe  
kurzawe@sub.uni-goettingen.de



Mathias Göbel  
goebel@sub.uni-goettingen.de

