# Text Mining for Literature Review

Filippo Chiarello, PhD

Lectures possibly to be held from 30 November to 23 December 2020.

An hands-on course on the use of automatic text analysis for analyzing scientific literature

The volume of scientific literature is growing in all disciplines. It is becoming increasingly hard for researchers  to monitor the evolution of an area and to define its boundaries (Chiarello et al., 2018). This issue is having a strong impact on highly innovative sectors, rapidly modified by the advent of digitalization and Industry 4.0 (Fareri et al., 2020). During the course the students will learn how to solve these kind of problems, using tools for paper retrieval, automatic text analysis (Silge & Robinson, 2017) and graphs analysis. The aim is to define the technical lexicon of a domain and reviewing a innovative-intense field of research.
To properly follow the lessons, students will need their personal computer with the following softwares:
RStudio: rstudio.com/products/rstudio/download
Tidyverse packages: tidyverse.org (to install using RStudio)
Gephi: https://gephi.org/

*Syllabus*

| Module | Lecture | Duration |
|---|---|---|
| Research Question | Practical approaches for research question definition, problem setting and framing | 1 h |
| Data Retrieval | Query Design | 1h |
| | Scopus API | 2h |
| Data Analysis | Introduction to R-Studio and the Tidyverse | 1h |
| | Metadata Analysis | 2h |
| | Text Mining for domain lexicon definition | 2h |
| | Graph analysis and visualization with Gephi | 1h |
| Case Study | IN depth analysis of a scientific article that uses the presented approaches (Chiarello et al., 2020) | 2h |

*References*

Chiarello, F., Trivelli, L., Bonaccorsi, A., & Fantoni, G. (2018). Extracting and mapping industry 4.0 technologies using wikipedia. Computers in Industry, 118.

*Info: filippo.chiarello @unipi.it*

Chiarello, F., Belingheri, P., Fantoni, G. (2020, Under Review). Data Science for Engineering Design: State of art and future direction. Submitted to: *Computers in Industry*

Fareri, S., Fantoni, G., Chiarello, F., Coli, E., & Binda, A. (2020). Estimating Industry 4.0 impact on job profiles and skills using text mining. *Computers in Industry*, 100.

Silge, J., & Robinson, D. (2017). Text mining with R: A tidy approach. " O'Reilly Media, Inc.".

**Short Bio:**

PHD in Management Engineering at the University of Pisa ("summa cum laude")  in 2018, Filippo Chiarello is a Researcher at the University of Pisa. His research is mainly focused on data science and text mining tools to build systems able to automatically extract knowledgeable information from technical documents. The ability to mix ontologies and technical dictionaries with natural language processing techniques brought his methods to be applied in both research and business. In the last year his focus shifted from technologies mapping to skills mapping, also mixing information coming from Human Resources related taxonomies such as ESCO and O*Net.

*Info: filippo.chiarello@unipi.it*