

# Bachelor of Informatics Honours

## Information & Knowledge in Organisations: Data Analytics

### *Assignment 3: What do Information Systems researchers write about?*

The rapid growth of Information Systems (IS) as a field of study has led to the publication of a large number of articles in this area. To gain insight into the research trends and patterns in IS, you are asked to analyze a bibliometric dataset that contains all IS articles published in the key journals and conferences in the field between 2011 and 2020.

The data contains information on the articles, including title, abstract, authors, publication year, and keywords. Your task is to analyze this dataset using text-based techniques to inform readers about the field. Your analysis can address, *for example*, the following research questions:

1. What are the most common topics and themes in IS research over the past decade?
2. How have the IS research topics and themes evolved over the past decade?
3. What are the most productive authors and journals in the field of IS?
4. What is the impact of IS research, as measured by citation counts?
5. What are the most frequent keywords in IS and how have they changed over time?
6. Is there a relationship between particular topics and themes and citation counts?
7. What is the relationship between authorship patterns (e.g. single authorship, multi-authorship) and research impact in the IS field?
8. Does the tonality of IS abstracts differ by topic?

In your analysis, you are expected to use NLP to process the abstracts (and any other available metrics/text) of the articles. You will then present your findings in a clear and concise manner, supported by appropriate tables, figures, and visualizations. Alongside your report, you will also submit your analysis code in a zip file. **The assignment should be completed in groups of three (3). The groups must be different to those of assignment 2.**

### Assessment criteria

#### Relevance and Quality of Own Research Questions (5%)

- The research questions should be relevant and aligned with the brief

#### Data Preparation and Cleaning (15%)

- Efficient and effective data cleaning to prepare the dataset for analysis
- Properly documenting the data preparation and cleaning process

#### NLP and Text-Based Analyses (50%)

- Appropriate application of NLP techniques to answer the research questions
- Interpretation and reporting of results in a clear and concise manner
- Properly documenting the NLP and text-based analyses process

#### Visualizations and Figures (15%)

- Use of appropriate and effective visualizations and figures to present findings
- Properly documenting the visualization creation process

#### Report Organization and Writing (10%)

- Clarity and coherence of the written explanation

#### Code and Data Management (5%)

- Clarity and organization of code

**Due Date: Friday 17 May, 23h59 (via SUNLearn link)**