



Contribution ID: 124

Type: **not specified**

Investigation of corporate sustainability reporting through automated text analysis

Wednesday, 18 September 2024 14:20 (20 minutes)

In order to evaluate the performance of companies, the focus is shifting from purely quantitative (financial) information to qualitative (textual) information. Corporate annual reports are comprehensive documents designed to inform investors and other stakeholders about a company's performance in the past year and its goals for the coming years. We have focused on the corporate sustainability reporting of FTSE 350 companies in the period 2012–2021. The lack of standardization and structuring of non-financial reporting makes such an analysis difficult.

We extracted all text from the non-financial sections of the annual reports using the pdf2txt tool and filtered it to retain only structurally correct sentences. We then identified sentences related to sustainability using a pre-trained sentence classifier (manual annotation). The content of these sentences was analyzed using the RoBERTa model, which was adapted to the financial domain. Using a hierarchical clustering algorithm, we identified 30 interpretable sustainability-related topics and 6–9 higher-level clusters of sustainability concepts.

For each report and each year, we calculated the proportion of topics within the report. The development of sustainability topics over time shows that external events and new reporting standards influence the overall content of the annual reports. In addition, we clustered the reports hierarchically based on the proportion of topics and identified 6 types of reports. The analysis showed that external events had the greatest influence on the structure of the individual reports.

Type of presentation

Talk

Classification

Mainly application

Keywords

sustainability, annual reports, text classification

Primary authors: FERJANČIČ, Urša (University of Ljubljana, School of Economics and Business); ICHEV, Riste (University of Ljubljana, School of Economics and Business); LONČARSKI, Igor (University of Ljubljana, School of Economics and Business); MONTARIOL, Syrielle (Jozef Stefan Institute); PELICON, Andraž (Jozef Stefan Institute); POLLAK, Senja (Jozef Stefan Institute); SITAR ŠUŠTAR, Katarina (University of Ljubljana, School of Economics and Business); TOMAN, Aleš (University of Ljubljana, School of Economics and Business); VALENTINČIČ, Aljoša (University of Ljubljana, School of Economics and Business); ŽNIDARŠIČ, Martin (Jozef Stefan Institute)

Presenter: TOMAN, Aleš (University of Ljubljana, School of Economics and Business)

Session Classification: PCA and mining

Track Classification: Data Mining and Warehousing