

Financial markets process information available to investors, who interact by trading assets and discover prices of assets. Scientific methods of analyzing prices of financial assets allow researchers to assess the economic importance of various types of information, their comparison and applications in forecasting. Information technologies change the scene, however, by allowing quicker access to an ever-growing body of information. Classic methods of financial analysis use only a fraction of the information available, because we lack the methods for integrating information. Unstructured data, such as text included in mandatory corporate reports, are a particular challenge. The problem is important, as recent years have seen an increase in skepticism about the information role of financial markets. At the same time, corporations are required to publish longer and longer reports that are difficult to read and understand.

This project aims to study how textual information evolves over time, what affects its development and what significance it carries for the financial markets. Textual information is less studied than numerical financial information, perhaps because people in finance care more about numbers than words. This project innovates in both the methods of study and the data to offer new, international evidence about textual information. We want to find out what patterns exist in financial texts and how they evolve. Next, we plan to see what motivates corporations to use or change these patterns. Finally, we will try to find out how useful these texts are for financial markets.

To make this project possible, we combine the theory of finance with linguistics, and we reach for computational methods of automatic textual analysis. From finance, we know we need to convert textual data into a signal that carries information and measure that information. We construct a large dataset of corporate reports and organize it following the procedures set by corpus linguistics.

From linguistics, we take the approach to thinking about text called the ATTITUDE framework. For example, we focus our study on evaluative expressions. These are sentences, in which the management of a corporation share their opinion about an event, a corporate result, or some activity. This way, we can identify one type of useful information in a report.

From computational linguistics, we take the tools that help computer algorithms process text. We reach to the CLARIN center at Wroclaw University of Technology which offers a research infrastructure for studying languages. We also make use of WordNet, a large database of language. WordNet organizes thousands of words around meanings and links them in a network. It gives computers a way to „think about” or process text. The best thing is, WordNets for different languages are connected, allowing us to combine data from many countries.

Our project offers new evidence useful for regulators who set reporting requirements and for management who want their texts to reach investors in the most effective way.