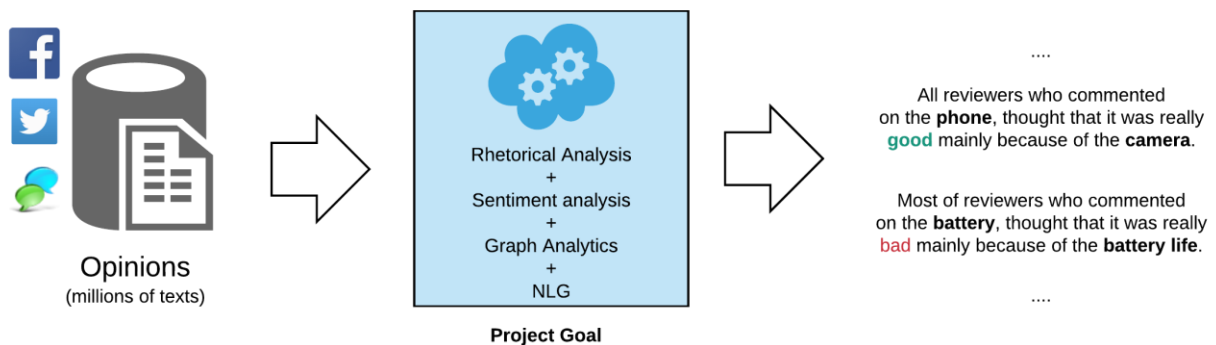


Modern society is an information society, bombarded from all sides by an increasing number of different pieces of information. The 21st century has brought us the rapid development of media such as found on the internet. This change has caused the transfer of many areas of our lives to virtual reality. New forms of communication have been established. Their development has created the need for analysis of related data. Nowadays, unstructured information is available in digital form, but how can we analyse and summarize billions of newly created texts that appear daily on the internet? Natural language analysis techniques, statistics and machine learning have emerged as tools to help us. In recent years, particular attention has focussed on sentiment analysis. This area is defined as the study of opinions expressed by people as well as attitudes and emotions about a particular topic, product, event, or person. Sentiment analysis determines the polarisation of the text. It answers the question as to whether a particular text is a positive, negative, or neutral one.

Sentiment assignment can be made on the level of (1) the whole document, (2) the individual sentences, or (what is the most interesting question to me and an extremely difficult scientific task) (3) on the level of aspects. Aspect-based sentiment analysis has been growing rapidly in recent years. It aims to predict the user's attitude to a specific feature of a product or service. As a result of such an analysis, we can aggregate the opinions not only at the level of whole texts but also specify that e.g. the phone's screen attracts positive opinions, but the battery is often judged negatively. This type of analysis is much more complex because it requires more advanced knowledge representation. In addition, the documents often consist of multiple sentences, and if you say that the document is positive that provides only partial information in describing this phenomenon. Currently, the literature is related only to initial work in this area. This subject is incredibly complex due to problems as to how to combine certain sentiment orientations and their aspects (features, attributes). There exists an opportunity for research in this domain in order to create a model that can automatically discriminate at such a deep level of text analysis. It is especially needed for various domains – and domains are understood here as thematically related datasets, e.g. texts related to music, electronics, books, politics, clothes, hotels and so on.



The subject of my research fills a gap in aspect-based sentiment analysis and aims to build a comprehensive set of techniques for preparing and analysing texts containing opinion and generating user-friendly descriptive reports in natural language. Currently, existing solutions offer analysis on the level of entire documents, and if you go deeper to the level of individual product features, they are only superficial and poorly prepared for the analysis of large volumes of data. This can especially be seen in scientific articles where the analysis is carried out on a few hundred reviews only. It is worth mentioning that this task is extremely problematic because of the huge diversity of languages and the difficulty of building a single solution that can cover all the languages used in the world. Natural language analysis often requires additional pre-processing steps, especially at the stage of preparing the data for analysis, and steps specific for each language. Large differences can be seen in the analysis of the Polish language (a highly inflected language) and English (a grammatically simpler one). I propose a solution that will cover the Polish and English languages. This type of analysis requires a series of tests for both languages and then we can select the most appropriate solutions for each of them.