

**FACULTY
OF MATHEMATICS
AND PHYSICS**
Charles University

BACHELOR THESIS

Václav Stibor

Financial News Sentiment Analysis

Department of Software Engineering

Supervisor of the bachelor thesis: **Supername Supersurname**

Study programme: **study programme**

Study branch: **study branch**

Prague **YEAR**

I declare that I carried out this bachelor thesis independently, and only with the cited sources, literature and other professional sources. It has not been used to obtain another or the same degree.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In date

Author's signature

Dedication. It is nice to say thanks to supervisors, friends, family, book authors and food providers.

Title: Financial News Sentiment Analysis

Author: Václav Stibor

Department: Department of Software Engineering

Supervisor: **Supername Supersurname, department**

Abstract: One key skill required to make good investments in the stock market is being able to correctly analyze news related to the finance and the business sector. Which company is diversifying its sectors or which company is showing signs of heading towards bankruptcy? You need to keep yourself updated with every little deal and fallout happening in the market. Financial news can be a little tricky to understand especially for those who are new to the financial world.

Keywords: **key words**

Contents

Introduction	2
1 Theoretical Background	3
1.1 Sentiment Analysis Basics	3
1.1.1 Levels of Sentiment Analysis	3
1.1.2 Workflow of Sentiment Analysis	6
1.2 Named Entity Recognition	6
1.3 Time Series Forecasting Integration	6
2 Related work	7
2.1 Existing Application Overview	7
2.1.1 Bloomberg Terminal	7
2.2 Predicting Stock Market Behaviour	10
2.3 Entity-level Sentiment Analysis	10
2.4 Mining dynamic Social Networks	11
3 Textual data	12
3.1 Aspects for considerations	12
3.2 Data sources	13
3.2.1 Web Scraping	13
3.2.2 RSS Feeds	13
3.2.3 News publisher's APIs	13
3.2.4 Third party data providers	13
3.3 First party data providers	13
3.3.1 The New York Times	13
3.3.2 The Guardian	13
3.4 Third party data providers	13
3.4.1 Alpha Ventage	13
4 Architecture	15
Conclusion	16
Bibliography	17
A Using CoolThesisSoftware	20

Introduction

In today's era of information explosion and constant flow of information, it becomes more time-consuming to keep track of associations and deeply understand the published content through media and online news, primarily when investing in a specific area. For instance, the investment in a company like Apple Inc. requires acquiring and processing a wide range of available information with significant effort and dedication in studying articles and other sources. At the same time, publicly available information resources such as news articles and tools like sentiment analysis allow us to transfer real-world context into the digital environment and use it for our benefit.

Sentiment analysis, the ability to identify and evaluate the emotional charge of content, has evolved into a crucial instrument for comprehending opinions, attitudes, and the general atmosphere surrounding various topics. This work focuses on developing an application that allows users to visualize connections between companies and news articles using a knowledge graph network and the impact of news sentiment on a company's stock price, *even in real time*.

Many experiments are currently being conducted based on historical data to examine the effect of sentiment, but not on current data, despite the rather promising results on datasets. The absence of such an application motivates this thesis. An application that extracts actual data from news articles for sentiment analysis and subsequently evaluates the future impact of that sentiment on a company's stock price.

This thesis will discuss the technical aspects of sentiment analysis and implementing an application that conveys this information to users as recently as possible. The aim is to provide users with a tool that allows them to actively monitor and analyze the flow of information about emotional overtones as one of the key identifiers in trading decisions. *The thesis will be structured as follows. Chapter 1 will discuss the theoretical background behind the stock market. Chapter 2 will give an overview of data sources and the data itself. Chapter 3 will discuss the sentiment analysis and design of the application. Chapter 4 will discuss the implementation of the application. Chapter 5 will discuss the evaluation of the application. Chapter 6 will discuss the conclusion and future work.*

(Q-0) Zeptat se - Jak je to s použitím 1. osoby množného čísla a pasivními formami. Tj. používání "We will, we discuss..." a "It will be discussed, it will be implemented...".

1. Theoretical Background

Since the application's core is sentiment analysis, it is necessary to define the basic concepts. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed non risus. Suspendisse lectus tortor, dignissim sit amet, adipiscing nec, ultricies sed, dolor. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed non risus. Suspendisse lectus tortor, dignissim sit amet, adipiscing nec, ultricies sed, dolor. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed non risus. Suspendisse lectus tortor, dignissim sit amet, adipiscing nec, ultricies sed, dolor. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed non risus.

1.1 Sentiment Analysis Basics

Sentiment analysis or opinion mining is a subfield of **NLP** that aims to identify and extract opinions and emotions from a text. The goal is to determine the author's attitude towards a particular topic or the overall contextual polarity of various document levels. We measure the text's polarity using a numerical scale ranging from -1 to 1. The low-end score of the scale signifies a negative sentiment, zero represents neutrality, and the high-end score indicates a positive sentiment. This scale effectively estimates the degree of negativity or positivity in the text's tone.

The extraction of opinions and emotions has applications in various areas, from product reviews to political events. Hence, it is imperative to work in different domains (see Piryani et al., 2017). Because of cross-domain and cross-language, two of the most general issues in sentiment analysis, this thesis will focus only on the financial domain in English. Nevertheless, domain-specific sentiment analysis achieves remarkable accuracy while staying highly domain-sensitive, as shown Saunders, 2020. To delve deeper into cross issues, Liu provides further details in his book *Sentiment analysis and opinion mining*.

(Q-1.1) Zeptat se - nebo je lepší odkaz "in his work Lei, 2022" v context citatce v Bibliography, kde odkazují na konkrétní strany z jeho knihy, které se daným problémem odkazují.

1.1.1 Levels of Sentiment Analysis

Sentiment analysis has been studied at several levels of granularity: Document-level, Sentence-level, Phrase-level, and Entity-level¹, as illustrated in Fig. 1.1

¹Entities are sometimes referred to as targets, hence Target-level or Target-based sentiment analysis.

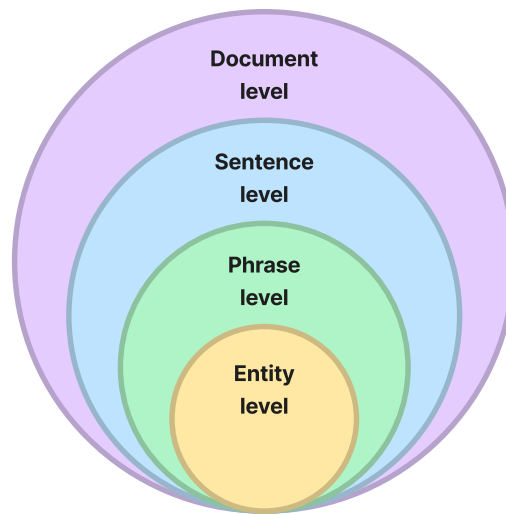


Figure 1.1 Levels of sentiment analysis (inspired by Wankhade et al., 2022).

(Q-1.1.1-1) Zeptat se - Je třeba citovat?

Document-level

Document-level sentiment analysis is the most straight level. The task is to determine the overall emotional context of the entire document, such as a chapter, article, or review, whether or not involving a study of entities or aspects. This level gives us a general assessment of whether the content is more likely to be positive, negative, or neutral.

Sentence-level

Sentiment analysis at the sentence level focuses on individual sentences within the text. We observe the polarity of each sentence autonomously, employing the same methodologies utilized at the document level but with an increased volume of training data and enhanced processing resources. This level is more challenging than the document level because it requires a more in-depth understanding of the text.

Phrase-level

Phrase-level sentiment analysis examines sentiment within smaller linguistic units such as phrases or sentence members. Thus, it can better reveal the emotional charge in specific parts of sentences. Additionally, this level is more challenging than the sentence level because it requires a more detailed understanding of the text.

Entity-level

The most elaborative level is entity-level sentiment analysis, where we study sentiment associated with specific entities mentioned in the text. This level provides a detailed look at the expressed polarity of certain products, individuals, or organizations. One of the main tasks in this scope is the named entity recognition, which will be discussed later.

Some researchers classify the last level as the aspect-level, as noted by Wankhade et al., [2022] or a more detailed entity-level version called the feature-level proposed by Mary et al., [2017]. While both approaches aim to evaluate sentiment towards specific aspects, they differ in their task approach. Relationships between these levels are illustrated in Fig. 1.2.

In the first case, aspects are considered without directly mentioning entities in the text. We are not interested in the entities since the investigated textual data are commonly associated with them², such as reviews. The study conducted by Wang et al., [2019] analyzed sentiment at the aspect level within restaurant reviews. It primarily examines aspects such as food, price, service, and others. In the feature-based approach, aspects are commonly associated with an entity's features by connecting the entity and its aspects in text. To illustrate, consider the sentence:

(Q-1.1.1-2) Zeptat se - Nebo zaměnit "investigated textual data" za "(input) data"?

"The battery life of this phone is excellent, but the camera is not good."

At the feature level, we identify *the battery life* and *camera* as specific features of entity *the phone*, allowing us to determine the polarity of each entity's feature.

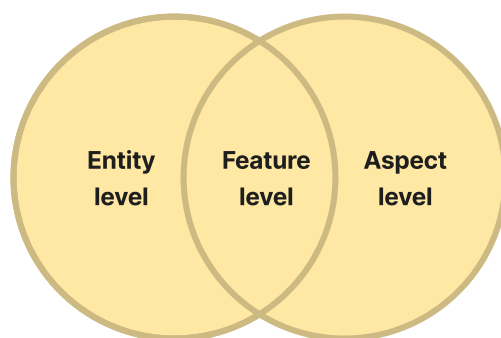


Figure 1.2 Comprehensive overview of the last level.

The term entity-level sentiment analysis is frequently employed in literature, and some studies consider it synonymous with targeted sentiment analysis,

²Entities are not handled in this case, but we provide them here for a better understanding.

as discussed Rønningstad et al., [2022] in the terminology review. For our purposes, entity-level sentiment analysis better captures the aggregate, document-wide approach, where a single entity can be associated with multiple targets in different sentences, discerning it from traditional target-level sentiment analysis.

However, this thesis primarily focuses on entity-level sentiment analysis, excluding consideration of the entity's features. This decision is motivated by treating the mentioned companies in news articles as entities rather than delving into their specific aspects. Additionally, entity and aspect extraction as separate tasks are complex and challenging, given that the methods and facets employed for recognition differ due to their distinct characteristics (Liu, [2015]; Zhang et al., [2014]).

TODO: Pokud nenarazím na článek, který by to vyvrátil. Navíc se zkoumáním aspektů by přibýlo spousty práce.

1.1.2 Workflow of Sentiment Analysis

The sentiment analysis process can be divided into three main steps: data retrieval, preprocessing, and analysis. The following sections will discuss these steps in more detail.

1.2 Named Entity Recognition

We will focus on the named entity recognition, also known as entity extraction, in the context of news. Named entity recognition is a subtask of NLP with a focus on identifying and classifying named entities in text into predefined categories such as the names of organizations, persons, locations, expressions of times, quantities, monetary values, and so on. Named entity recognition is a crucial step in entity-level sentiment analysis, as it allows us to identify the sentiment associated with specific entities mentioned in the text. And so on...

1.3 Time Series Forecasting Integration

We will focus on integrating time series forecasting in the context of news.

(Q-1.3) Zeptat se - Bude třeba úvod do stock marketu? Podle mě ne a zbytečně by se to míchalo do (technology) teoretického pozadí.

2. Related work

While examining online applications with similar characteristics, a common challenge arises from the need for more transparency regarding the techniques applied in sentiment analysis. Specifically, employed methods and the details of the models remain undisclosed. We acknowledge this limitation due to the proprietary nature of the software, mainly as it is an essential part of the business model. Hence, this chapter is organized into two primary sections.

The initial part provides an overview of existing applications. At the same time, the second section will concern the most recent and relevant research on predicting stock market behaviour using data mining techniques and news sentiment analysis. Additionally, it explores research studies on entity-level sentiment analysis over news articles and its application in the stock market.

2.1 Existing Application Overview

Sentiment analysis applications accessible to the public are typically based on investigating social network posts, with StockTwist^[1] as a notable illustration. According to Reuters, StockTwits in 2022 boasts more than six million registered users and one million monthly active users, underscoring its prominent user base. With the growing volume of social media contributions, it is difficult to determine which post will prompt action. In our case, these posts do not constitute highly relevant data since our interest is in news articles containing a more significant amount of information. Even if the selection of accounts is limited to informative sources, News organizations' social media accounts only link to their articles, usually accompanied by a headline or lead paragraph. These posts are not enough for our purposes, as we need the entire article to perform better context for entity-level sentiment analysis (discussed further in Chapter ^[3]).

(Q-2.1) Zeptat se - je třeba uvádět odkaz na webový článek?

2.1.1 Bloomberg Terminal

The only software similar to the one created is a module in the so-called Bloomberg Terminal^[2] from Bloomberg L.P.. This software system provides investors with analytical tools over financial data, including sentiment analysis of news articles and posts on social network X, formerly known as Twitter. The cost of

¹<https://stocktwits.com>

²<https://www.bloomberg.com/professional/solution/bloomberg-terminal/>

TODO: a nebo the ?

(Q-2.1.1-3) Zeptat se - Mohu přehodit jména autorů? Protože jako poslední autor zmiňovaného white paperu je Bloomberg L.P. a rád bych an to kladl důraz. To znamená, aby se při citatci objevilo Bloomberg L.P. namísto Cui

a Bloomberg Terminal depends on the required specific features and services. A standard subscription typically amounts to approximately \$24,000 per year. It is a very complex and powerful software, but it is not accessible to the general public.

Bloomberg's work (Bloomberg L.P. et al., 2024) describes two types of sentiment analysis: story-level and company-level sentiment, utilizing a suite of SML techniques. Classification engines are trained on labeled datasets containing news articles and social media posts. Reportedly, the labeling process is based on the question³

"If an investor having a long position in the security mentioned were to read this news or tweet, is he/she bullish, bearish or neutral on his/her holdings?"

Once model training is completed, the models are employed to analyze recently published posts and articles associated with organizations, seeking distinctive sentiment signals related to the business and finance domain. As mentioned above, sentiment is divided into two levels:

Story-level Sentiment score value of articles and posts is calculated after arrival in real time. The calculation includes score and confidence, where the score has one of three options: positive, negative, or neutral, each described by a numerical value from the set $\{-1, 0, 1\}$. The confidence is defined by a value ranging from 0 to 1, demonstrating the intensity of the sentiment, which can be interpreted as the probability of being positive, negative, or neutral. Hence, the story-level sentiment ranges from -1 to 1 . For both, we get the following equation:

$$\text{Story-level}_c^{\text{Articles}} = S_c^a C_c^a, \quad a \in P(c) \quad (2.1)$$

$$\text{Story-level}_c^{\text{Posts}} = S_c^p C_c^p, \quad p \in P(c) \quad (2.2)$$

where a represents an article and p represents a post from $P(c)$, the set of published articles and posts referencing company c . S_c^a and S_c^p are the sentiment polarity scores of article a and post p that reference company c . C_c^a and C_c^p are the confidences of article a and post p that reference company c .

Company-level Sentiment score value is then calculated as the confidence-weighted average of the story-level sentiment scores, incorporating all relevant news articles and social media posts mentioning the company.

$$\text{Company-level}_{c,t}^{\text{Articles}} = \frac{\sum_{a \in P(c,T)} S_c^a C_c^a}{N_{c,T}^a}, \quad T \in [t_b, t] \quad (2.3)$$

³This information is difficult to verify due to the unavailability of the dataset.

$$\text{Company-level}_{c,t}^{\text{Posts}} = \frac{\sum_{p \in P(c,T)} S_c^p C_c^p}{N_{c,T}^p}, \quad T \in [t_b, t] \quad (2.4)$$

where a represents an article and p represents a post from $P(c, T)$, the set of published articles and posts referencing company c during period T . Period T is a time interval of length t_b to t , where t_b is the time constant of the beginning. $N_{c,T}^a$ and $N_{c,T}^p$ are the number of articles and posts referencing company c during period T . In this way, the company-level sentiment is calculated as follows:

$$\text{Company-level}_{c,t} = \text{Company-level}_{c,t}^{\text{Articles}} + \text{Company-level}_{c,t}^{\text{Posts}} \quad (2.5)$$

Intraday Company-level sentiment score for news articles is recalculated every two minutes, utilizing an eight-hour rolling window. The sentiment score for social network posts is recalculated every minute, employing a 30-minute rolling window. Due to the previous definitions, we can express these by a simple substitution of t_b depending on the rolling window as follows:

$$\text{Intraday Company-level}_{c,t}^{\text{Articles}} = \frac{\sum_{a \in P(c,T)} S_c^a C_c^a}{N_{c,T}^a}, \quad T \in [t - 8, t] \quad (2.6)$$

$$\text{Intraday Company-level}_{c,t}^{\text{Posts}} = \frac{\sum_{p \in P(c,T)} S_c^p C_c^p}{N_{c,T}^p}, \quad T \in [t - 0.5, t] \quad (2.7)$$

$$\text{Intraday Company-level}_{c,t} = \text{Company-level}_{c,t}^{\text{Articles}} + \text{Company-level}_{c,t}^{\text{Posts}} \quad (2.8)$$

Daily company-level sentiment scores are published every morning about 10 minutes before the market opens. The calculation is determined as a confidence-weighted average of sentiment scores derived from the story-level sentiments of news and social media posts over the past 24 hours as follows:

$$\text{Daily Company-level}_{c,t} = \frac{\sum_{d \in P(c,T)} S_c^d C_c^d}{N_{c,T}}, \quad T \in [t - 24, t] \quad (2.9)$$

where document d represents a news article or social media post, the sentiment polarity score of document d referencing company c is denoted as S_c^d , and the confidence associated with this reference is represented by C_c^d . The set $P(c, T)$ encompasses non-neutral documents referencing company c published within the last 24 hours. $N_{c,T}$ expresses the count of non-neutral documents referencing company c during period T . This approach is further explored in terms of the informational role of social media by Gu et al. (2020).

2.2 Predicting Stock Market Behaviour

Several approaches for predicting stock market behaviour and price trends have been proposed, utilizing sentiment analysis of financial news and historical stock prices. Several studies prove a strong correlation between financial news sentiment and stock prices (Li et al., 2014) (Wan et al., 2021). Due to the nature of unstructured textual data, predicting stock market behaviour is a challenging task.

Khedr et al. (2017) focuses on creating an effective model for forecasting future trends in the stock market, using sentiment analysis of financial news and historical stock prices. The model achieves more accurate results than previous works by considering different market and company news types combined with historical stock prices. The experiments utilize datasets from three companies: Yahoo Inc., Facebook Inc.⁴, and Microsoft Corporation. The authors use well-known and informative news sources such as Reuters, The Wall Street Journal, and Nasdaq. The first step of sentiment analysis to get the text polarity using a Naive Bayes classifier was shown to achieve accuracy from 72.73% to 86.21%, while the second step, which combines news sentiment with historical prices, improved prediction accuracy up to 89.80%. Moreover, that is why we find this study motivating and inspiring.

2.3 Entity-level Sentiment Analysis

This section will discuss the most recent and relevant research conducted in entity-level sentiment analysis and its application in news articles, including named entity recognition over news articles.

Zhao et al., 2021 employed RoBERTa, a Robustly Optimized BERT Pretraining Approach (Liu et al., 2019b), to propose a sentiment analysis and entity detection strategy in financial text mining and public opinion analysis in social media. In the first step, sentiment analysis, mainly focusing on negative polarity, is performed. Then, entity detection is considered in different granularities with MRC, Machine Reading Comprehension (Liu et al., 2019a), or sentence-matching tasks. As a result, this study serves entity detection differently than traditional Named Entity Recognition. The authors claim that the proposed method outperforms traditional sentiment analysis and entity detection methods. The authors also emphasize the importance of the financial domain, where the sentiment of a single entity can significantly affect the stock price.

⁴Known as Meta since 2021

Named entity recognition plays a significant role in entity-level sentiment analysis, as discussed in Section [2.3](#). Vybírám články, které stojí za zmínku.

2.4 Mining dynamic Social Networks

This section will discuss interesting research Jin et al., [2012](#) dealing with mining dynamic social networks and its application in the stock market.

Ještě si nejsem jistý, zda-li je tato sekce potřebná.

Conclusion

In the conclusion, you should summarize what was achieved by the thesis. In a few paragraphs, try to answer the following:

- Was the problem stated in the introduction solved? (Ideally include a list of successfully achieved goals.)
- What is the quality of the result? Is the problem solved for good and the mankind does not need to ever think about it again, or just partially improved upon? (Is the incompleteness caused by overwhelming problem complexity that would be out of thesis scope, or any theoretical reasons, such as computational hardness?)
- Does the result have any practical applications that improve upon something realistic?
- Is there any good future development or research direction that could further improve the results of this thesis? (This is often summarized in a separate subsection called 'Future work'.)

This is quite common.

Bibliography

- Pirayani, R. et al. (2017). *Analytical mapping of opinion mining and sentiment analysis research during 2000–2015*. In: *Information Processing & Management* 53.1, pp. 122–150. ISSN: 0306-4573. DOI: <https://doi.org/10.1016/j.ipm.2016.07.001>. URL: <https://www.sciencedirect.com/science/article/pii/S030645731630245X>.
- Saunders, Danielle (2020). *Domain adaptation for neural machine translation*. PhD thesis. Apollo - University of Cambridge Repository. DOI: [10.17863/CAM.66458](https://doi.org/10.17863/CAM.66458). URL: <https://www.repository.cam.ac.uk/handle/1810/319335>.
- Liu, Bing (2022). *Sentiment analysis and opinion mining*. In: Springer Nature. Chap. 3, pp. 31–36.
- Wankhade, Mayur et al. (2022). *A survey on sentiment analysis methods, applications, and challenges*. In: *Artificial Intelligence Review* 55.7, pp. 5731–5780. ISSN: 1573-7462. DOI: [10.1007/s10462-022-10144-1](https://doi.org/10.1007/s10462-022-10144-1). URL: <https://doi.org/10.1007/s10462-022-10144-1>.
- Mary, A. Jenifer Jothi et al. (2017). *Jen-Ton: A framework to enhance the accuracy of aspect level sentiment analysis in big data*. In: *2017 International Conference on Inventive Computing and Informatics (ICICI)*, pp. 452–457. URL: <https://api.semanticscholar.org/CorpusID:44112128>.
- Wang, Yequan et al. (2019). *Aspect-level Sentiment Analysis using AS-Capsules*. In: *The World Wide Web Conference. WWW '19*. San Francisco, CA, USA: Association for Computing Machinery, 2033–2044. ISBN: 9781450366748. DOI: [10.1145/3308558.3313750](https://doi.org/10.1145/3308558.3313750). URL: <https://doi.org/10.1145/3308558.3313750>.
- Rønningstad, Egil et al. (Oct. 2022). *Entity-Level Sentiment Analysis (ELSA): An Exploratory Task Survey*. In: *Proceedings of the 29th International Conference on Computational Linguistics*. Ed. by Nicoletta Calzolari et al. Gyeongju, Republic of Korea: International Committee on Computational Linguistics, pp. 6773–6783. URL: <https://aclanthology.org/2022.coling-1.589>.
- Liu, Bing (2015). *Aspect and Entity Extraction*. In: *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. Cambridge University Press, 137–188.
- Zhang, Lei et al. (2014). *Aspect and Entity Extraction for Opinion Mining*. In: *Data Mining and Knowledge Discovery for Big Data: Methodologies, Challenge and Opportunities*. Ed. by Wesley W. Chu. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 1–40. ISBN: 978-3-642-40837-3. DOI: [10.1007/978-3-642-40837-3_1](https://doi.org/10.1007/978-3-642-40837-3_1). URL: https://doi.org/10.1007/978-3-642-40837-3_1.

- Bloomberg L.P. et al. (2024). *Embedded Value in Bloomberg News & Social Sentiment Data*. Received via email from Bloomberg L.P. on 2.2.2024. URL: https://data.bloomberglp.com/promo/sites/12/725454457_EDFSentimentWP.pdf.
- Gu, Chen et al. (2020). *Informational role of social media: Evidence from Twitter sentiment*. In: *Journal of Banking & Finance* 121, p. 105969. ISSN: 0378-4266. DOI: <https://doi.org/10.1016/j.jbankfin.2020.105969>, URL: <https://www.sciencedirect.com/science/article/pii/S0378426620302314>.
- Li, Xiaodong et al. (2014). *News impact on stock price return via sentiment analysis*. In: *Knowledge-Based Systems* 69, pp. 14–23. ISSN: 0950-7051. DOI: <https://doi.org/10.1016/j.knosys.2014.04.022>, URL: <https://www.sciencedirect.com/science/article/pii/S0950705114001440>.
- Wan, Xingchen et al. (2021). *Sentiment correlation in financial news networks and associated market movements*. In: *Scientific Reports* 11.1, p. 3062. ISSN: 2045-2322. DOI: [10.1038/s41598-021-82338-6](https://doi.org/10.1038/s41598-021-82338-6), URL: <https://doi.org/10.1038/s41598-021-82338-6>.
- Khedr, Ayman E et al. (2017). *Predicting stock market behavior using data mining technique and news sentiment analysis*. In: *International Journal of Intelligent Systems and Applications* 9.7, p. 22.
- Zhao, Lingyun et al. (2021). *A BERT based Sentiment Analysis and Key Entity Detection Approach for Online Financial Texts*. In: *2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, pp. 1233–1238. DOI: [10.1109/CSCWD49262.2021.9437616](https://doi.org/10.1109/CSCWD49262.2021.9437616).
- Liu, Yinhan et al. (2019b). *RoBERTa: A Robustly Optimized BERT Pretraining Approach*. arXiv: [1907.11692 \[cs.CL\]](https://arxiv.org/abs/1907.11692).
- Liu, Shanshan et al. (2019a). *Neural Machine Reading Comprehension: Methods and Trends*. In: *Applied Sciences* 9.18. ISSN: 2076-3417. DOI: [10.3390/app9183698](https://doi.org/10.3390/app9183698), URL: <https://www.mdpi.com/2076-3417/9/18/3698>.
- Jin, Yingzi et al. (2012). *Mining dynamic social networks from public news articles for company value prediction*. In: *Social Network Analysis and Mining* 2.3, pp. 217–228. ISSN: 1869-5469. DOI: [10.1007/s13278-011-0045-5](https://doi.org/10.1007/s13278-011-0045-5), URL: <https://doi.org/10.1007/s13278-011-0045-5>.

Acronyms

BERT Bidirectional Encoder Representations from Transformers. 10

NLP Natural Language Processing. 3, 6

SML Supervised Machine Learning. 8

A. Using CoolThesisSoftware

Use this appendix to tell the readers (specifically the reviewer) how to use your software. A very reduced example follows; expand as necessary. Description of the program usage (e.g., how to process some example data) should be included as well.

To compile and run the software, you need dependencies XXX and YYY and a C compiler. On Debian-based Linux systems (such as Ubuntu), you may install these dependencies with APT:

```
apt-get install \  
  libsuperdependency-dev \  
  libanotherdependency-dev \  
  build-essential
```

To unpack and compile the software, proceed as follows:

```
unzip coolsoft.zip  
cd coolsoft  
./configure  
make
```

The program can be used as a C++ library, the simplest use is demonstrated in listing [1](#). A demonstration program that processes demonstration data is available in directory demo/, you can run the program on a demonstration dataset as follows:

```
cd demo/  
./bin/cool_process_data data/demo1
```

After the program starts, control the data avenger with standard WSAD controls.

Listing 1 Example program.

```
#include <CoolSoft.h>
#include <iostream>

int main() {
    int i;
    if(i = cool::ProcessAllData()) // returns 0 on error
        std::cout << i << std::endl;
    else
        std::cerr << "error!" << std::endl;
    return 0;
}
```
