Fine-Grained News Classification

Maia Lunde Gjelsvik

Supervisor: Prof. Enrico Motta

Master's Thesis
Department of Information Science and Media Studies
University of Bergen

May 31, 2023

Acknowledgements

First and foremost, I would like to thank my supervisor, Prof. Enrico Motta. Throughout this thesis, you have motivated me with your engagement and enthusiasm. Working with you has been incredibly fun and interesting. I am forever grateful to you for sharing your knowledge and for all your advice. Good luck with your article and future work!

I am also very grateful to Dr. Angelo Salatino for helping me produce and release the dataset.

I would also like to thank my parents who have given me endless support throughout my life, and especially during my time as a student. You have always been proud of me, which has been a great source of motivation. Uendelig glad i dere begge to.

I would also like to thank my boyfriend, Espen, who has been very patient with me during the most stressful times.

Lastly, I would like to thank all my friends who participated in the empirical study.

Maia Lunde Gjelsvik
Oslo, 28.05.2023

Abstract

This thesis investigates the concepts of fine-grained news classification. To do this, an empirical study in which human annotators categorized news was conducted. The also study consisted of measuring agreement between human annotators and evaluating the precision of the annotations. The study revealed a need for a framework for fine-grained news classification. A framework was then developed and evaluated, producing a complete annotated dataset.

Contents

List of Figures

Chapter 1

Introduction

As digital versions of news items are being published in vast amounts on a daily basis by countless news providers, the potential for computational media analysis is enormous. However, trying to navigate this landscape of news is difficult as the manual categorization conducted by journalists during the publication process tends to provide only generic categories. Given these deficiencies, the search process for similar articles is extremely difficult and can produce noisy results. Additionally, the practice of manual categorization itself is costly and time-consuming. This practice leads to a no-win situation, where readers are unable to conduct effective and accurate search while journalists are required to complete this tedious task. Furthermore, current computational methods for news classifications are not able to provide the fine-grained classifications desired. IPTC NewsCodes provide a taxonomy for news classification, however, this consists of rather coarse-grained and broad terms *IPTC* (2022).

This thesis aims to provide a contribution towards addressing these issues by proposing and validating a model for fine-grained news classification that encompasses the various aspects that can provide the focus of news stories, including entities, events, viewpoints and others.

1.1 Motivation

As the availability and amount of digital content are at exponential growth, there is a need to classify this data to conduct large-scale analysis. The availability of this digital data entails great opportunities for insights and research. Existing approaches to classify digital content, specifically news documents, are significantly limited. Manual classification, which is the most commonly used method by news outlets and journalists, offers no scalability. Additionally, the manual classification practice may lead to biased categorizations. Furthermore, there is no standardized vocabulary between news outlets, leading to inconsistent categorizations and difficulties concerning search. Automatic computational methods offer a more scalable approach but cannot provide precise fine-grained classifications. Topic Modelling is a frequently used machine learning method for news classifi-

cation. However, the methods can only classify at a coarse-grained level. Entity Recognition and Entity Linking can provide useful tools in news classification, however, these fields of research can only partially address the issue of fine-grained news classification, as this activity goes beyond recognising entities in news content. Because of these issues concerning news classification, this thesis aims to develop and validate a framework that can provide fine-grained classification. Furthermore, to develop such a framework, an investigation of what fine-grained classification precedes.

1.2 Problem Statement

To develop a framework for fine-grained news classification, firstly, the concept of fine-grained classification needs to be investigated and clarified. Additionally, which aspects of a news article must be identified to achieve fine-grained classification? To understand the concept of fine-grained classification, an empirical study where human annotators were asked to generate specific annotations was conducted. As mentioned in the previous section, the aim of this thesis is to develop a framework for fine-grained news classification. Furthermore, the framework is required to provide precise and accurate classification. Additionally, the framework must be practical to use, in the sense that it captures the aspects that are present in a news article and can be used and understood easily.

1.3 Research Questions

Through the work that is conducted in the context of this thesis, the following research questions will be answered. To answer these questions, an empirical study with human annotators was conducted. Additionally, the development and validation of a framework was completed.

RQ1: What does it mean to perform fine-grained news classification?

RQ2: Is it possible to produce a framework for fine-grained news classification that is accurate and complete?

1.4 Contribution

Through the work that is presented in this thesis, several contributions have been made. The most noticeable contributions include the following:

- An investigation of what fine-grained news classification includes.

By conducting a preliminary empirical study involving human annotators, the concept of fine-grained classification and what that entails has been investigated and clarified. Furthermore, this study includes other relevant findings, such as an insight into the effectiveness of the solution that is being proposed to the issues related to manual categorization.

- A validated framework for news classification.

  This thesis also contributes with a complete and validated framework for news classification. This framework is developed to capture all important aspects that can be found in any given news article. Furthermore, this framework was developed through an iterative process, which concluded in a finalized framework.

- A dataset that can be used to train and evaluate classification models.

  A result of the validation process is a dataset comprising 224 news articles annotated according to the proposed framework. This dataset can be used to develop and train computational models for fine-grained news classification.

1.5 Thesis outline

Chapter 2 Background: This chapter introduces relevant automatic computational methods and techniques for news classification and relevant research within this field. Furthermore, it discusses existing procedures for news categorization and proposed solutions to the issues that follow manual categorization.

Chapter 3 Methodology: This chapter describes the theory that was used to develop a method of approach. The chapter also details the methodology used in the Empirical study, such as Content analysis and Qualitative Research. Furthermore, it discusses Iterative Design, which is the methodology utilized to develop the Framework. Lastly, it describes Ontology-Driven analysis.

Chapter 4 Methods: This chapter details the empirical study that was conducted as a preliminary analysis. Additionally, it describes how

the framework was developed and the work completed to validate it. Lastly, this chapter details the method of procedure to annotate news articles according to the framework.

Chapter 5 Results: This chapter documents the results that were derived from the empirical study. Furthermore, it details the Framework, which was a result of the iterative design process. Additionally, it discusses the validation process, which included the production of a complete annotated dataset.

Chapter 6 Discussion: This chapter discusses the results presented in chapter 5. Firstly, it details the results from the analysis of the empirical study. Following this, this chapter correlates the concepts presented in the framework to relevant literature and existing approaches. Lastly, this chapter details the contributions made in this thesis and the limitations.

Chapter 7 Conclusion and Future Work: The final chapter discusses the conclusions that can be drawn from the results and discussion in the previous chapters. Additionally, it discuss the future work that can be conducted within the field.

Chapter 2

Background

This chapter details various relevant research and existing methods for news classification. Firstly, the section discusses the relevant automatic approaches to news classification. Additionally, it explores methods for entity recognition and event extraction. Moreover, the chapter provides an outline of relevant ontology-based research. Furthermore, this section discusses research within the area of manual classification and the associated issues. Lastly, this chapter provides a rendition of current solutions to the challenges of manual classification. Lastly, an exemplification of topic assignments from a news outlet will be illustrated.

2.1 Machine Learning

Machine learning is a field in computer science that concerns constructing methods or models that can learn using large amounts of data. The methods then improve the execution of a task. Machine learning is considered to be a part of the Artificial Intelligence field *Alpaydin* (2014). There are several variations of Machine Learning, with Topic Modelling and Neural Networks being the most relevant method used for research-based news classification, and will be explored further in the following section.

2.1.1 Topic Modelling

Topic Modelling is a machine-learning model for unsupervised document classification widely used for news classification and other machine-learning tasks. A topic modelling algorithm aims to uncover patterns or themes in unstructured data. Documents are probability distributions over topics *Allahyari et al.* (2017). The model can then organise the documents based on the themes that have been established. The main advantage of topic modelling is that it is very adaptable, and it can be utilised on different kinds of data *Blei* (2012).

Topic modelling methods offer clustering of documents based on the discovered latent topics, as seen in figure 2.1. In this illustration of a topic

model clustering the color coded clusters represent different broad topics. Some of the latent topics discovered in the green clusters are; Covid, Coronavirus, vaccine, numbers, Omicron, Booster, NHS, Javid and elderly.

The red cluster contains documents where the following latent topics where identified by a topic modelling method; United states, Russia, weapons, Putin, Donbass, Ukraine, threat.

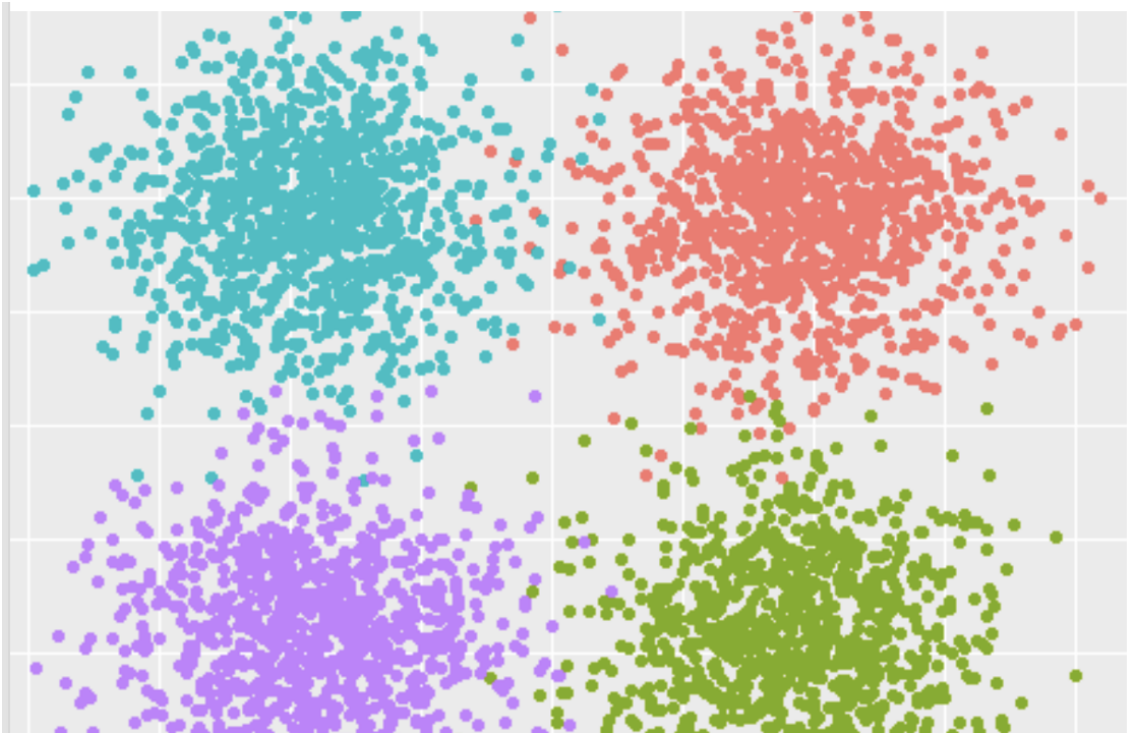

Figure 2.1: Topic Model Clusters

Latent Dirichlet Allocation

Latent Dirichlet Allocation (LDA) is a statistical topic model. The LDA model includes a "bag-of-words" assumption, meaning that the order of the words and grammar are not considered *Blei* (2012). The LDA model calculates the probability of each word in a document belonging to a topic. The words with the highest probability score are then allocated to represent that topic *Blei et al.* (2003). There are other varieties of topic modelling. However, LDA is one of the most commonly used news classification research methods.

In 2016, Jacobi et al. conducted a study where the goal was to classify news using LDA *Jacobi et al.* (2016). Jacobi et al. concluded that LDA could be a sufficient method for classifying news if the goal was to get fast results with little cost. Furthermore, they found that analysing the results was a burden. They also state that LDA would be suitable for preliminary analysis, and a more advanced method would have to follow. The patterns recognised by the LDA model were described as vague and broad.

Troncy et al. explored several topic modelling algorithms to determine the superior method. They explored Latent Dirichlet Allocation (LDA), Non-Negative Matrix Factorization (NMF), Latent Semantic Analysis (LSA), Probabilistic Latent Semantic Analysis (PLSA), Topical Word Embedding (TWE) and Word2Vec. These six algorithms were trained and tested on the same dataset. Following this, they were then evaluated using the same metrics, coherence, diversity, exclusivity, perplexity and runtime.

LSA and Word2Vec performed poorly on all metrics. PLSA performed satisfactorily on coherence and perplexity but inadequately on diversity and exclusivity. Furthermore, LDA, NMF and TWE performed consistently well on all metrics.

Troncy et al. concluded that there is no superior method of topic modelling. Furthermore, they advise future researchers to evaluate several techniques to determine the method most suitable for their project *Harrando et al.* (2021).

Evaluation of Topic Models

Chang et al. conducted a study to evaluate the interpretability of topic models. Interpretability refers to which level of understanding humans have of the topic models. Chang et al. executed a study involving human participants to investigate the interpretability of topic models. The participants were asked to identify the latent themes discovered in a corpus of text documents. They were presented with the top words recognised by the topic model and were asked to render which underlying topic the top words referred to. The results from the study were then analysed to discover patterns. There was a high level of agreement among the participants. However, there was also a substantial level of variation in latent topic identification. Chang et al. concluded that the participants might have varied interpretations of the same topic model output because of preexisting biases.

Their study only involved 30 participants and one topic model. Inciting the question of whether generalisations could be made. Chang et al. stressed the need for further investigation for this reason. They concluded that topic models had great potential to identify general or broad themes in a textual corpus *Chang et al.* (2009).

Similarly to Chang et al., Doogan and Buntine investigated the problem of evaluating the interpretability of topic models. In this context, interpretability refers to which level of understanding humans have of the output from topic models.

In their research, they investigate measures for the evaluation of interpretability. One commonly used method is Topic Coherence, which measures the similarities between the top words belonging to a topic. Doogan and Butine contest the reliability of Topic Coherence because it does not consider word context or semantic correlation between words. Addition-

ally, they investigated the Diversity measure, a method that evaluates to which extent a topic scopes aspects of a domain. They argue that Diversity Measure may not be a good indicator of interpretability because it does not gauge highly diverse topics. Furthermore, they found that traditional evaluation metrics and topic models may not be compatible, as the development of topic models has exceeded the evaluation metrics.

Because Doogan and Buntine concluded that current measures have limited performances, they proposed a new measure, Topical Consistency Measure. Doogan and Buntines proposed measure considers the consistency of keywords belonging to a topic across documents. Additionally, the Topical Consistency Measure evaluates to which extent keywords belonging to a topic co-occur in the same documents. Furthermore, the Topical Consistency Measure assumes a topic is more interpretable if belonging keywords are consistently used in the same context.

They concluded their research by emphasising the importance of interpretability measures for topic models and analysis to be practical and valuable. Although their interpretability measure performed close to human evaluation, Doogan and Buntine emphasise the need for further research *Doogan and Buntine* (2021).

2.1.2 Neural Networks

Neural Network techniques are also a common machine learning method for text classification.

Conneau et al. have, in their study, utilized these techniques in combination with convolutional operations to attempt to improve effectiveness and accuracy when classifying text documents. Additionally, they aimed to develop a scalable method, as traditional text classification models did not offer this feature. Additionally, they developed a model that would capture the complex patterns that traditional models could not. Furthermore, their model accounted for the hierarchical representation present in textual data. They found that their model outperformed several existing methods for text classification on several datasets. However, Natural Network text classification suffers from limited benchmark and training data, leading to difficulties in generalisation. Furthermore, available training data and benchmark datasets offer too coarse-grained annotations, as they are based on standards, such as IPTC.

2.2 Named Entity Recognition and Entity Linking

Named Entity Recognition is the process of identifying entities in a text document. In the context of named entity recognition, what an entity is can vary; for example, an entity can be a person, an organisation, a date or a location. Named Entity Recognition is a very useful tool in research concerning information extraction. Additionally, Named Entity Recognition models are beneficial for utilizing and comprehending unstructured text.

Marrero et al. discuss Named Entity Recognition in their study from 2012. They want Named Entity Recognition to be viewed as a more nuanced instrument. To do this, they discussed the challenges, benefits and potential improvements of Named Entity Recognition.

They found that entity disambiguation was an issue concerning named entity recognition. Entities can be represented under different names, such as aliases, nicknames, abbreviations and acronyms in text documents. Additionally, they found that this was a frequent occurrence. Marrero et al. also discuss the challenges of context and understanding contextual information. A lack of context can become an issue with entity classification and result in inaccurate entity recognition.

They conclude that implementing formal knowledge representation in Named Entity Recognition could improve the issue of entity disambiguation. They also suggested implementing solutions based on Neural Architectures and Deep Learning to achieve non-domain-specific models that generalize and recognise entities across domains. Furthermore, Marrero et al. petition for standardized evaluation metrics. Additionally, they discuss the opportunity for progression in the field of Named Entity Recognition that would follow the development of benchmark datasets *Marrero et al.* (2013).

Entity Linking entails entity recognition and entity disambiguation. Entity Recognition can be performed by using Named Entity Recognition techniques. Entity Disambiguation involves relating the identified entity to another entity of the closest proximity in the knowledge base.

Pontes et al. discuss the challenges correlated to Entity Linking. They found similar issues with this procedure, as Marrero et al. identified in their investigation of Named Entity Recognition. As mentioned, entity linking includes entity recognition. Therefore, entity Linking inherently suffers from many of the same challenges as Named Entity Recognition, such as variation in entity identifiers, ambiguous context and lack of supporting datasets, such as training data. They propose the use of domain expertise to resolve the issue of limited available data. Additionally, specific to the domain of historical documents, Pontes et al. propound the benefit of implementing a historical language model, which could improve entity linking in text documents that are not built on contemporary language *Linhares Pontes et al.* (2020).

In 2015 Shen et al. also discussed the challenges and possible improvements related to Entity Linking. Similarly to Marrero et al. and Pontes et al., they also identified entity and contextual disambiguation as a paramount challenge within the field. Additionally, Shen et al. discussed the matter of scalability of entity linking techniques. As the knowledge base for an entity linking technique is at continuous growth, the models are required to produce accurate results while handling a large number of entities and text doc-

uments. Additionally, Entity Linking models relate entities across multiple knowledge bases, which exacerbates the scalability issue. Furthermore, Entity Linking models are subject to the issue of heterogeneous knowledge bases, where the formal representation can be of different structure, quality and content.

2.3 Event Extraction

Event Extraction consists of identifying and classifying events and additional information about identified events in a text document.

Smith et al. conducted an analysis of the limitations of event extraction. Additionally, their study includes possible implementations for improvement. They point to the issue of limited domain-specific available data, which leads to imprecise and inaccurate event extraction. Additionally, not only is there a lack of domain-specific labelled data, but there is also a limitation of data in several different languages, which excludes many domains that are specific to a certain language. Because of the limitations concerning labelled training data, building a model that can offer accurate and precise event extraction is difficult.

Smithæ et al. also discuss the issues of event extraction models disregarding the inferred or implicit events correlated to another extracted event. While the models cannot identify the inferred events, the event extraction is incomplete. Additionally, Smith et al. are critical of the lack of inclusion of temporal relations in event extraction. They encourage further research within the domain to resolve the current issues *Smith et al.* (2021).

In 2018 Huang et al. aimed to resolve the issue of limited available training data for event extraction. The issue, concerning many domains, entails the lack of labelled training data for event extraction models. Their proposal also aims to provide more precise and accurate performance for event extraction. Huang et al. propose a zero-shot transfer learning framework. Their model is initially trained on a non-domain-specific dataset. By doing this, the model can achieve generic and simple event extraction. The model is then adapted by training on a very small domain-specific dataset. Huang et al. evaluated this approach and found that the method outperforms other methods where domain-specific data is limited. However, Huang et al. discuss the possibility of varying outcomes based on available data from the target domain *Huang et al.* (2018).

Liu et al. have in their study, proposed a method that includes event extraction and machine reading comprehension to address the limitations of the accuracy of existing event extraction methods. The reason for their study is the argument that current event extraction methods are reliant on predetermined patterns. These predetermined patterns limit the methods' ability to capture the variation and complexity present in natural language. Their

event extraction method is formulated as tasks consisting of question answering. Included in their research is an evaluation and analysis of the model's performance. Their method is evaluated using precision and recall. Based on these metrics, they concluded that their model outperformed traditional event extraction methods *Liu et al.* (2020).

2.4 Relevant Ontology-based research

In 2017 Brown et al. developed a comprehensive Ontology to represent complex event information. Their ontology, "The Rich Event Ontology" or REO, was developed to be used in the context of news items. The reason for their project was to create an ontology that would capture not only the events that could be found in a document but the information associated with events. They aimed to simplify the complex and nuanced task of capturing additional information in a formal representation.

REO consist of several classes and their correlations. The ontology includes important aspects of events, such as temporal, causal intentional and spatial features. Moreover, mechanisms to capture the relation between events or dependencies between events and causalities are included in the ontology. Including mechanisms to capture these features of events leads to a nuanced representation of both events and event relationships.

They detail their contribution as an aim for news annotation through a demonstrated annotation process. They illustrate this area of use through document retrieval based on specified events or event features *Brown et al.* (2017).

Borgo et al. have through their research developed an Ontology for Linguistic and Cognitive Engineering, DOLCE. Their research focuses on important linguistic concepts, such as space, objects, events, time and qualities. DOLCE captures these linguistic concepts. Their ontology is not developed for a specific domain and can therefore be adapted and applied under various circumstances *Borgo et al.* (2022).

In 2021 Carriero et al., in their research paper, focused on the formal representation of reoccurring situations or events. Their work aimed to develop an ontology design pattern that would capture recurring situations or events. Their reason for the need for this ontology was the importance of recurring situations in fields such as environmental monitoring, finance and sports. The issues they aimed to resolve were the challenges related to recurring situations were the aspects of constant variation of the patterns and dependencies. Their proposed solution of an ontology was developed to be flexible, which also allows for domain-specific adaptation. In their ontology, they separate the notion of recurring events and recurring situations and include properties to capture the temporal aspects of both. They evaluate the ontology, which indicates good results. Furthermore, they conclude that their ontology is useful and can capture recurring events, situations and

their properties *Carriero et al.* (2021).

2.5 Manual Classification

As mentioned previously, manual classification is the most common practice of news categorization used by journalists and media outlets. The reason for including these categorizations in each article is to simplify search within a topic on the outlet's website. However, because no standard practice has been implemented in the business, IPTC has attempted to create a taxonomy to combat the issue of no shared vocabulary. The NewsCodes taxonomy, developed by IPTC, consists of keywords at a coarse-grained level and will be exemplified in this section *IPTC* (2022). Exemplification of some categories retrieved from news articles published on The Guardian's website will be given.

2.5.1 Issues Concerning Manual Classification

As mentioned, the practice of manual news categorisation is the most common method during the publication process. This method of category annotation is often performed by journalists. Therefore, the annotations assigned to news articles across news outlets are not of a shared vocabulary. Meaning the annotations are dependent on organizational standards. Furthermore, because the process is manually completed by individuals, there is a substantial possibility that annotations not only vary between news outlets but also between journalists. Annotations that are dependent on the person will inherently suffer from personal bias. Furthermore, as mentioned this practice is also very time-consuming and therefore costly *Gallofré Ocaña and Opdahl* (2022).

2.5.2 IPTC NewsCodes

The International Press Telecommunications Council (IPTC) works to supply technical tools and standards for news providers, to improve the flow of information. The IPTC has developed a taxonomy for news categorisation. A taxonomy is a controlled vocabulary, meaning it consists of predefined categories. This controlled vocabulary is a tool used to assign predefined categories to news text. The means of developing this taxonomy was to establish a shared vocabulary between different news providers. Furthermore, it is available in several languages, but mainly for news providers *IPTC* (2022). The taxonomy created by the IPTC contains a relatively broad and unprecise vocabulary. In other words, the categories have a large scope. Figure 2.2 is retrieved from the IPTC web page and illustrates the top level of the IPTC NewsCodes taxonomy.

IPTC Media Topic NewsCodes as of 2022-02-04 (language: en-GB)

Expand All | Collapse All | View language versions: Arabic | Chinese (分类) | Danish | English (en-GB) | English (en-US) | French | German | Norwegian | Portuguese | Portuguese (Brazil) | Spanish | Swedish

MediaTopic ID	Name (en-GB)	Definition (en-GB)
medtop:01000000	arts, culture, entertainment and media	All forms of arts, entertainment, cultural heritage and media
medtop:16000000	conflict, war and peace	Acts of socially or politically motivated protest or violence, military activities, geopolitical conflicts, as well as resolution efforts
medtop:02000000	crime, law and justice	The establishment and/or statement of the rules of behaviour in society, the enforcement of these rules, breaches of the rules, the punishment of offenders and the organisations and bodies involved in these activities
medtop:03000000	disaster, accident and emergency incident	Man made or natural event resulting in loss of life or injury to living creatures and/or damage to inanimate objects or property
medtop:04000000	economy, business and finance	All matters concerning the planning, production and exchange of wealth.
medtop:05000000	education	All aspects of furthering knowledge, formally or informally
medtop:06000000	environment	All aspects of protection, damage, and condition of the ecosystem of the planet earth and its surroundings.
medtop:07000000	health	All aspects pertaining to the physical and mental welfare of living beings
medtop:08000000	human interest	Item that discusses individuals, groups, animals, plants or other objects in an emotional way
medtop:09000000	labour	Social aspects, organisations, rules and conditions affecting the employment of human effort for the generation of wealth or provision of services and the economic support of the unemployed.
medtop:10000000	lifestyle and leisure	Activities undertaken for pleasure, relaxation or recreation outside paid employment, including eating and travel.
medtop:11000000	politics	Local, regional, national and international exercise of power, or struggle for power, and the relationships between governing bodies and states.
medtop:12000000	religion	Belief systems, institutions and people who provide moral guidance to followers
medtop:13000000	science and technology	All aspects pertaining to human understanding of, as well as methodical study and research of natural, formal and social sciences, such as astronomy, linguistics or economics
medtop:14000000	society	The concerns, issues, affairs and institutions relevant to human social interactions, problems and welfare, such as poverty, human rights and family planning
medtop:15000000	sport	Competitive activity or skill that involves physical and/or mental effort and organisations and bodies involved in these activities
medtop:17000000	weather	The study, prediction and reporting of meteorological phenomena

Figure 2.2: IPTC taxonomy top level

Figure 2.3 illustrates all levels of the "disaster, accident and emergency incident" category. As shown in Figure 2.3 a top-level category consists of many sub-categories. However, even with several levels, the hierarchy contains keywords in a general manner.

medtop:03000000	disaster, accident and emergency incident	Man made or natural event resulting in loss of life or injury to living creatures and/or damage to inanimate objects or property
medtop:20000139	accident and emergency incident	A sudden, unexpected event that causes unwanted consequences or requires immediate action
medtop:20001321	drowning	When a person's airpipe and lungs become filled with water causing death or injury, most often caused by falling into or bathing in rivers, lakes or the ocean
medtop:20000140	explosion accident and incident	Explosions caused by an accidental, non-combat, or non-terrorism source. Also use when the source of the explosion is non-specified or unknown
medtop:20000141	industrial accident and incident	Any unplanned event or mishap in an industrial setting that result in injuries to people and damage to property or the environment
medtop:20000142	nuclear accident and incident	Any unplanned event that causes unwanted consequences involving radioactive materials
medtop:20000161	structural failure	The unexpected collapse of a building, bridge or other structure
medtop:20000143	transportation accident and incident	An accident or incident involving one or more vehicles
medtop:20000144	air and space accident and incident	An accident or incident involving aircrafts or spacecrafts
medtop:20000145	maritime accident and incident	An accident or incident involving marine vessels.
medtop:20000146	railway accident and incident	An accident or incident involving trains.
medtop:20000147	road accident and incident	An accident or incident involving road vehicles.
medtop:20001201	wildlife traffic accident	Traffic accidents involving wild animals
medtop:20000148	disaster	A sudden, unplanned event that causes great damage or loss of life, such as an accident or a natural catastrophe
medtop:20000149	famine	Severe lack of food for a large population
medtop:20000150	fire	Fires started by an accidental or unknown source. For wildfires, use wildfire. For fires started with a criminal intent, use arson.
medtop:20001202	wildfire	An uncontrolled fire in an area of combustible vegetation
medtop:20000151	natural disaster	Destructive incidents caused by nature, such as earthquakes and floods
medtop:20000152	drought	A severe lack of water over a period of time
medtop:20000153	earthquake	The shifting of the tectonic plates of the Earth, creating in some cases damage to structures
medtop:20000154	flood	Surfeit of water, caused by heavy rains or melting snow, usually in places where it's not wanted
medtop:20000155	landslide	Sudden dislodging of massive amounts of snow or soil
medtop:20000156	avalanche	A sudden fall of snow, ice, rock, or mud down the side of a mountain
medtop:20000157	meteorological disaster	A weather-related disaster
medtop:20001280	hurricane	A type of storm also called a tropical cyclone, which forms over tropical or subtropical waters
medtop:20000158	windstorms	A storm with very strong wind but little or no rain or snow
medtop:20001353	tsunami	High and powerful ocean waves caused by an underwater land disturbance, such as an earthquake or volcanic eruption, known to cause significant damage and loss when they hit land
medtop:20000159	volcanic eruption	A rupture in the crust of the Earth allowing molten material to escape to the surface
medtop:20000168	emergency response	The planning and efforts made by people and organizations to help victims of a sudden, unplanned event, natural disaster or crisis

Figure 2.3: IPTC taxonomy Disaster category and sub-categories

2.5.3 Topic assignment: The Guardian

As stated, the most common practice for annotating news articles is currently manual classification. The figures below exemplify the topic assigned to articles retrieved from The Guardian.

The first article covers a prediction of a drought in California, USA. The governor of California threatens to introduce mandatory water restrictions if the residents do not reduce their water usage, shown in figure 2.4. The topic assignment for this news article is depicted in figure 2.5

California declares drought emergency across vast swath of state

Majority of counties now under emergency declaration as California faces extensive dry spell and dwindling water supply

Houseboats are dwarfed by the steep banks of Lake Oroville last month in Oroville, California. Photograph: Justin Sullivan/Getty Images

Figure 2.4: The Guardian: California Drought

Figure 2.5: The Guardian Topics California Drought

The second article concerns the Liberal Democrats' criticism of raw sewage being pumped into English bathing waters 25,000 times in 2021. These figures were compiled by Environment Agency.

Figure 2.6: The Guardian: Raw Sewage in English Bathing Waters

Figure 2.7: The Guardian Topics Raw Sewage in English Bathing Waters

In the first news story, the focus is on a particular event, the emergency declaration in California. This is obviously not captured in the news annotation. The second story also focuses on an event, "Raw sewage pumped into English bathing waters", and again this aspect is not captured at all in the annotations

Search using manual categorization

This coarse-grained classification is present in all of the categorizations conducted by The Guardian. For example, the news article about the consequences of the staffing crisis in UK care homes was uncovered after inspections were conducted, shown in image 2.8.

Staffing crisis leaves many English care home residents' basic needs unmet

Vulnerable people left alone in rooms for 24 hours a day and denied showers for a week as number of vacancies grows to 165,000

'They don't have time to talk': residents feel impact of care staffing crisis

Staff shortages have been identified as a key problem in three-quarters of all care homes in England. Photograph: David Levene/The Guardian

Figure 2.8: Staffing crisis leaves many English care home residents basic needs unmet

Topics

Social care

Health / England / Care Quality Commission (CQC) / news

Figure 2.9: Topics Staffing crisis leaves many English care home residents basic needs unmet

This article was assigned the categorizations shown in figure 2.9. As seen, the topics assigned are very coarse-grained and when searching for similar articles using the topics provided by The Guardian, the results are depicted in figure 2.10 from the 'Social Care' topic from approximately the same period as the original article was published.

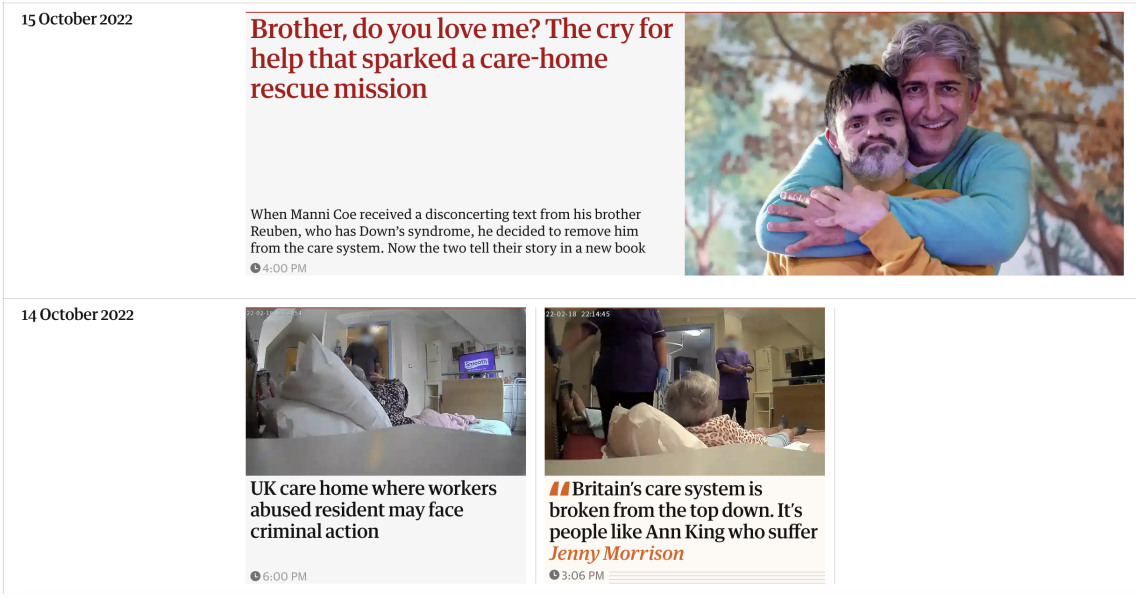

Figure 2.10: Search Results 'Social Care'

Although, some of these articles are about the level of care in UK care homes, they do not cover the staffing crisis in UK care homes, which was the key event in the original news article. The suggested articles under the 'Care Quality Commission (CQC)' topic do not provide other news articles covering the staffing crisis, search results from approximately the same time period shown in figure 2.11.

Figure 2.11: Search Results 'Care Quality Commission (CQC)'

Furthermore, when attempting to search for similar stories using google news, the results do not offer exclusivly articles concerning the exact same topic. Shown in figure 2.12.

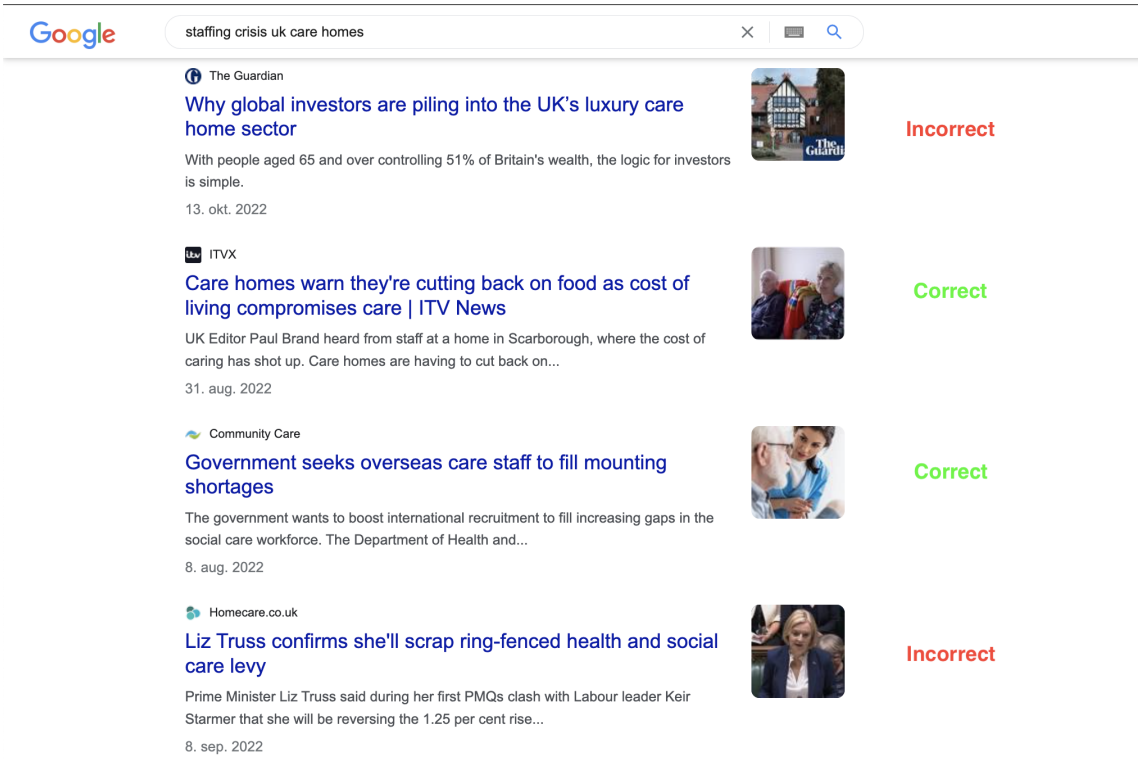

Figure 2.12: Google news search results for staffing crisis in UK care homes

2.6 Inter-rater Agreement Measures

There are several different types of inter-rater agreement measures. The different measures vary in appropriateness depending on the material that is being analyzed. Cohen's Kappa is the most commonly used method to measure inter-rater agreement. However, Cohen's Kappa only accounts for two raters. Cohen's kappa was introduced in 1960 and has since been the basis for new methods that can measure agreement between two or more raters. Fleiss Kappa is an extension of Cohen's kappa, the method was introduced in 1971 and can be used to calculate agreement between more than two raters. Fleiss Kappa measures numerical assignment by raters. Additionally, the categories are mutually exclusive. Fleiss Kappa also accounts for expected agreement. The expected agreement includes measuring the chance of agreements *Fleiss* (1971). Fleiss Kappa is a common method used when measuring agreement in behavioural studies *Laerd Statistics* (2019).

Additionally, Fleiss Kappa is considered a reliable method that can measure agreement for nominal scale data. Furthermore, the method offers reliable results when analysing incomplete or unbalanced data *Laerd Statistics* (2019).

However, Fleiss Kappa does not account for varying reliability and independence in raters. Furthermore, the Fleiss Kappa measure does not con-

sider the magnitude or order of the categories and only evaluates the number of raters for one specific category at a time. While Fleiss Kappa is a commonly used method in many fields of study, it is not suitable for the analysis of continuous data or ordinal data *Laerd Statistics* (2019).

Chapter 3

Methodology

3.1 Empirical research

To complete an Empirical study involving human annotators, empirical research methods were used. These methods are detailed in the following section.

3.1.1 Content analysis

Content analysis is a research method where patterns or themes within a selected corpus are identified. The selected corpus can contain any communication content, such as images, audio recordings or text. Researchers determine a set of representations of features common to the corpus. The content is then systematically analysed and studied according to the determined set of features. From there, patterns and conclusions are yielded *Williamson and Johanson* (2017).

Intercoder Reliability

Intercoder Reliability is a collective term for intercoder agreement measures, like Cohens Kappa or Fleiss Kappa. Intercoder Reliability measures represent the level of agreement between two or more independent raters who are coding the same corpus. Providing clear and refined guidelines or codes to raters is vital to achieving high intercoder reliability. High intercoder reliability can indicate the validity of coding schemes *Williamson and Johanson* (2017).

3.1.2 Qualitative Research

Qualitative research scopes the methods of analysis and study that aim to interpret human behaviour and experience. Qualitative research data can be collected through several courses, such as interviews, observation, and other means of interaction with participants. Qualitative Research aims to draw conclusions about a specific phenomenon *Williamson and Johanson*

(2017).

Qualitative research methods provide many benefits concerning the analysis of human behaviour or experience. Through qualitative research methods, researchers are able to obtain an in-depth insight into a certain perspective. Additionally, qualitative research methods include the context in which a phenomenon occurs, a trait that gives analysis nuance. Moreover, Qualitative research is versatile in the sense that methods can be adapted to be suitable for any domain. Furthermore, based on emerging insights and trends, researchers are able to reevaluate and change their approach. As Qualitative research is very adaptable, emerging insights are a common occurrence.

3.2 Iterative Design

Iterative design is a method within design science research that aims to develop an artefact through several iteration stages. The iterative design includes repeating several tasks for its iteration stage, eventually producing a complete artefact. The first stage involves defining design requirements and the initial design of the artefact. Each stage then consists of further development and design of the artefact and then evaluation. In the following stage, the results of the previous evaluation are considered to improve and further develop the artefact. This process is repeated until the artefact is complete and the evaluation reveals that the artefact is accomplished according to requirements.

The evaluation of the artefact in each stage is vital to produce a complete artefact. Vigorous testing should be conducted at the end of each stage to accomplish the aim of the research. Furthermore, collaboration between researchers is beneficial when conducting an iterative design process. Involving several researchers at each stage will lead to important insight at each iteration. Using an iterative design methodology while developing an artefact can lead to effective and substantial contributions and research findings to the respective discourse. *Williamson and Johanson* (2017)

3.3 Ontology-Driven analysis

The ontology-driven analysis consists of several phases. Firstly, an ontology is developed for a specific domain *Saggion et al.* (2007). An ontology formally represents the knowledge that defines concepts and correlations within a domain. It can be used to outline vocabulary and requirements for reasoning. An ontology consists of classes or concepts that have determined properties. Additionally, ontologies scope the relations between the concepts. Often ontologies are represented in a hierarchical structure *Guarino et al.* (2009).

Utilising ontologies aids in the comprehension of complex data sets. Developing an ontology is, however, very time-consuming. Furthermore, the quality of insights obtained depends on the ontology's quality and the analysed data. Therefore, the ontology must be comprehensive and accurate. The ontology-driven analysis utilises the beneficial structure of ontologies and the ability to classify data meaningfully. Following, data is collected and transformed to conform to the requirements and vocabulary determined in the ontology. Subsequently, the data can then be analysed using a preferred suitable method, like machine learning or statistical analysis *Saggion et al.* (2007).

Chapter 4

Methods

This chapter will detail the methods used in this thesis. Firstly, the methods used in the empirical study consisting of human annotators will be detailed. Furthermore, the development of the framework that followed will be described, including earlier versions of the framework. The finalised version was derived from these iterations but will be explained in the Results chapter. Moreover, the validation process of the framework will be described. Lastly, exemplifications of the procedure to annotate news using the framework will be detailed.

4.1 Empirical Study with Human Annotators

The empirical study consisted of 8 participants who were asked to annotate a corpus of news stories using two contrasting means. The participants were asked to complete two annotation tasks on each news article. The first task was designed to measure agreement between the participants. The second task was to investigate the concept of fine-grained classification.

Before the study was completed, the participants signed a consent form detailing how and why the study would be conducted. Although none of the participant's personal information was used in the study, the consent form gave the participants an in-depth description of how their contribution would be used. The consent form was signed by the participants digitally.

Additionally, the inclusion of a consent form early in the study eliminated the participants who were not serious about partaking. Only one individual who signed the consent form did not complete the study. After the participants had signed the consent form, they received an email containing a description of the two tasks and the corpus.

The consent form and instructions sent to the participants can be located in the appendix.

4.1.1 Data gathering for Empirical study

The corpus consisted of 26 news articles. The participants were asked to annotate 25 news of the documents provided. In order to give some directions as to how the news items could be annotated, the first article of the corpus was annotated by the investigator.

Before data gathering, the investigator set the following requirements for news items to be included. There were several reasons for requirements to be set, for example, varied language proficiency. Additionally, requirements were set in place to ensure as many participants as possible completed the study. Only one person withdrew from the study after signing the consent form and receiving the corpus. The requirements in place for each article in the corpus were the following:

- Because of the inconsistent English proficiency within the group of participants, it was a necessity that the corpus consisted of news items with a low level of complexity and language.

- Additionally, to sustain participants attention and engagement, it was advantageous to eliminate lengthy and extensive news articles in the selection process. Articles that were eliminated on this basis were long and often covered stories that would require prior knowledge about a particular topic to obtain a complete understanding of the news document.

- Furthermore, news articles detailing grim or disturbing events were avoided to ensure the participants would not feel uncomfortable.

- The corpus of news articles should cover various topics. Including several stories about the same topic or events in the corpus could become tedious and repetitive for the participants.

Because of the above requirements, a random selection of data was excluded. Therefore, the news items included in the corpus were read thoroughly and examined according to the requirements.

4.1.2 Provisions

To complete the tasks, participants were provided with an example. The article used for exemplification covered the controversy following a go-

karting event for inmates at Fresnes Prison in France. The example is presented in Figure 4.1.

ID	Headline	IPTC NewsCodes	Self-Generated Topics
1	French justice ministry under pressure to explain jail go-karting	>crime, law and justice >> prison	controversy over go-karting event at Fresnes prison

Figure 4.1: Examplification empirical study

Additionally, the participants were given a link to the IPTC NewsCodes taxonomy for the first task.

The participants were provided PDF files of each article instead of links to The Guardians website. This was because the existing annotations could influence how the participants completed the tasks and could not be disclosed.

4.1.3 Tasks

As mentioned above, the participants were asked to execute two tasks. Each task consisted of news annotation. Additionally, the two tasks were carried out using the same corpus.

Task 1

The participants were provided with a link to the IPTC NewsCodes website to execute the first task. When visiting this website, the participants were presented with the top-level categories. The participants were then required to navigate multiple subcategories within the IPTC NewsCodes taxonomy and identify all appropriate levels associated with the top-level category they found suitable. They were also asked to indicate the different levels in their annotations.

The purpose of this task was to evaluate how the participants would perform annotating news items given guidelines or a taxonomy. Additionally, measuring the agreement between annotators using a taxonomy would provide insight into the relevancy of solving the issues presented previously in the thesis.

Task 2

To complete this task, the participants were given instructions to read the news articles and provide more specific self-generated topics than those provided in Task 1. They were asked to capture the actual events and topics

covered in the story. The participants were not given a format to present the topics identified.

The aim of this task was to investigate the concept of fine-grained classification. Additionally, the annotations from tasks 1 and 2 provided insight into how agreement varies when given a specific method for annotating.

4.1.4 Methods and methodology for analysis

To analyse the results from the empirical study, two different methods were used.

Task 1: Statistical Analysis

To investigate the agreement provided in task 1 between the participants, a statistical interrater agreement method was used. This analysis was conducted using a Python program. The Statsmodels library includes a built-in function for calculating the Fleiss Kappa Score of agreement. Because Fleiss Kappa is originally used for numerical ratings, the labels provided by the annotators were transformed into numerical values using a hash function *Fleiss* (1971).

For analysis of the user study consisting of eight human annotators, it is necessary to utilize a measure that can calculate agreement between several raters. While several methods can be suitable, Fleiss Kappa was used in this project. The formula for Fleiss Kappa is:

$$K = \frac{Po - Pe}{1 - Pe}$$

Where Po is the observed agreement and Pe is the expected agreement.

The score calculated using Fleiss Kappa will be between 0 and 1, where 1 indicates complete agreement beyond chance, and 0 indicates no agreement *Fleiss* (1971). The scores are interpreted as the following:

Fleiss Kappa Value	Description
0.81-1.00	Perfect Agreement
0.61-0.80	Substantial Agreement
0.41-0.60	Moderate Agreement
0.21-0.40	Fair Agreement
0.01-0.20	Slight Agreement
0	Poor Agreement

Table 4.1: Kappa Values

Task 2: Qualitative Analysis

Because the annotations provided by participants in task 2 were differing in vocabulary, preciseness and length, using a statistical analysis method was unattainable. In order to conduct statistical analysis, the annotations would have to be normalised to a very large extent, thus risking the loss of patterns and meaning. Therefore, qualitative analysis was chosen for the analysis of the annotations given in task 2 *Williamson and Johanson* (2017).

4.2 Development of the Framework

Following the empirical study with human annotators, the concept of fine-grained news classification required further investigation. In the study, annotations that managed to capture the entirety of a news article were very lengthy and appeared like paragraphs or summaries. One conclusion drawn from the empirical study was that fine-grained news classification is complicated and requires several collaborative mechanisms. This resulted in the proposal of creating a typology or ontology designed to act as a framework for news classification. The framework was throughout the entire process developed in collaboration with the supervisor of this thesis, Prof. Enrico Motta.

An early version of the framework was developed. The early version of the framework consists of a typology of concepts that can be identified in news articles. The framework was developed through several iterations. The first version of the framework was used to annotate a small number of news articles from Aftenposten, a Norwegian newspaper. Validating the framework as it was being developed revealed any inadequacy or shortcomings.

4.2.1 Early versions of framework

Through the development and validation of the framework, the typologies included were both extended and specified. This section will detail the earlier version of the framework.

The first iteration of the framework consisted of six typologies identified by analysing the corpus collected for the empirical study and the annotations generated by the participants. The first iteration of the framework is depicted in figure 4.2. During this iteration, to evaluate the typology, a small number of articles from Aftenposten were annotated. This was to identify any shortcomings of the framework. After this small validation process was completed, several changes were made to the framework.

During this stage of the development process, the news articles that were annotated were only assigned a typology that was present in that story. The specific typologies were not identified further. The reason this simple an-

notation tactic was chosen was to discover any possible weaknesses of the framework in a short period of time.

Categories	Definition	Example
Specific events		"Salt Bae's London restaurant reports £7m in sales in its first three months"
Composite events	These are events that by definition comprise many sub-events.	a trial
Macro events	These are abstractions of many individual events, which are not necessarily part of a composite event (in contrast with the trial example earlier)	Thousands of people suffering inadequate care in UK's care homes
Prediction of event		
Discussion of event		criticism of go-karting event at Fresnes prison
Entities	These can be people, organizations, countries, fictional characters	such as Mickey Mouse

Figure 4.2: Framework: First iteration

Specific Events refer to events that are not part of a larger parent event or main event. News articles that cover specific events.

Composite Events are events that include other more specific (sub-)events. For example, a trial would be considered a main event, whereas a closing argument and the jury deliberation are examples of sub-events of the trial.

Macro Events are correlated events that refer to the same group of people, topic or issue. For instance, let's consider a news story that talks about multiple drone sightings in Norway. This story aggregates a number of events that have taken place in different locations at different times, thus creating a macro event, which abstracts from the specific aspects of each individual sighting. This notion of macro event is related to the notion of "impact" in journalistic guidelines, in the sense that a story about multiple sightings has more impact than one about an individual sighting.

Prediction of Events Prediction of Events captures the news items that centre around a prediction of coming events. Predictions can be based on studies that based on statistical analysis are able to predict possible future events. This classification is also relevant when the focus of a news story is speculation of future events. Where the speculation is based on previous events that have occurred.

Discussion of Event The classification of Discussion of Events is relevant in news items where a topic is debated by individuals. As the topic of a news item is often an event, many news articles include the discussion of this topic by individuals who have some relation to the events being discussed.

Entities refer to news articles about a person or a business. In this iteration of the framework Entities are exemplified by Mickey Mouse. News articles that contain this aspect can often be interviews, for example, where a famous person talks about their upbringing or personal life. News articles mainly focusing on an entity are not necessarily driven by an event.

4.2.2 Expansion of the Framework

As the validation of the early version of the framework revealed some weaknesses, the framework was expanded. In this version, the classifications detailed in the previous section remained. In this version, more specific classifications concerning entities were added. This was because a number of articles did not have only an entity as the point of convergence but a specific facet of an entity. Additionally, a number of news items included the description of the relation between entities.

The second iteration of the framework is illustrated in figure 4.3

This version of the framework was more similar to the finalised version. However, there are a number of changes made to the complete version. The complete version of the framework will be detailed in the results section.

Categories	Definition	Example
Specific events		"Salt Bae's London restaurant reports £7m in sales in its first three months"
Composite events	These are events that by definition comprise many sub-events.	a trial
Macro events	These are abstractions of many individual events, which are not necessarily part of a composite event (in contrast with the trial example earlier)	Thousands of people suffering inadequate care in UK's care homes
Prediction of event		
Discussion of event		criticism of go-karting event at Fresnes prison
Entities	These can be people, organizations, countries, fictional characters	such as Mickey Mouse
Relations between entities		as in the article about the Honduran Maradona
Entity facet	Articles about an aspect of an entity	

Figure 4.3: Framework: Second iteration

It is not uncommon for news items to cover the **Relation between entities**. Although the profile of the entities determines the newsworthiness, the relation between two or more entities can often be the point of convergence.

Entity Facet details the features of an entity that can be the point of convergence in news stories. Entity Facets can be a number of features, for example, if an individual's political ideology is the focus of attention in a news story.

4.3 Validation of Framework

As mentioned previously, the framework was further developed after the earliest iterations by validation. The validation provided insights about which classes of categories were incomplete and if the framework was insufficient. Early in the validation process, the framework was finalised. Following the completion of the framework, a proper validation process was conducted. A result of this validation was a dataset consisting of 224 annotated stories.

4.3.1 Data collection for Validation of Framework

The corpus used to validate the framework consists of 224 news articles retrieved from two news outlets, Aftenposten and The Guardian. Aftenposten is a Norwegian newspaper, while The Guardian is a British newspaper. Both news outlets cover nationwide news and international news. The articles were collected by visiting news outlets' websites on particular differing days and collecting links for all the stories published that day. This ensured that the retrieval process was not influenced by selection or bias. Additionally, articles from Aftenposten were not collected on the same days as articles from The Guardian. The reason for this was to eliminate the possibility of several documents covering identical topics or events, as both news outlets include the cover of international events.

The news items collected from Aftenposten were collected over several days in October of 2022. More specifically, 18.10.2022, 20.10.2022, 21.10.2022, 23.10.2022 and 24.10.2022. By collecting all articles published by Aftenposten on the mentioned days, the corpus consisted of 105 articles from this specific news outlet.

The news items collected from The Guardian were retrieved on three subsequent days, 25.10.2022, 26.10.2022 and 27.10.2022. During this time period, 107 news articles were collected and 98 of those news articles were used. The reason for eliminating 9 of the collected articles was various. Some of the items collected were not actual stories but a headline and pictures. For the purpose of annotating the content of the news article, these pictures were not suitable.

Furthermore, the 26 stories used in the empirical study using human annotators were also included in the corpus. These stories were, at the time of collection, highly influenced by selection.

Because of an imbalance in the number of publications each day by each

news outlet, the collection of Aftenposten news articles required more days than The Guardian. The aim was to accumulate a substantial and approximately equal amount of articles from each news outlet,

4.3.2 Annotating news articles

This section will detail the method of procedure to annotate news articles according to the finalized version of the framework. This section will exemplify and give a reason for the detection of the typologies.

As mentioned previously, the framework was designed through several iterations. The finalized version is depicted in figure 4.4. The finalized version of the framework and the components it consists of will be discussed in detail in the Results chapter. While using the framework there is always a possibility of interpretation. However, the interpretation process is facilitated by the journalistic writing style used in the news sector, which emphasises the use of a lead paragraph to summarise the main focus of an article *University of Arizona Global Campus Writing Center*.

Category	Typologies	Definition	Example
Events	Atomic Events	Individual Events	"Salt Bae's London restaurant reports £7m in sales in its first three months"
	Negative Events	Expression of agency by not doing something	Several countries fail to submit reevaluated climate plans
	Dependent Events	An event that would not have happened had it not been for a key event (Atomic, Negative)	Companies cut ties with Kanye West after antisemitic statements
	Collection of Events	Several key events	Champions league roundup
	Composite Events	Consists of several sub-events	Ecclestone pleads not guilty in fraud trial
Situation		A situation classifies the state of affairs	Third of young women and girls in UK can't access free period products
Prediction of Events		Speculation or prediction of future coming events	Prediction of refugee crisis because of war in Ukraine
Entity	Entity Aspect	An aspect or facet of an Entity	'I want to keep being the first': Hideo Kojima on seven years as an independent game developer' Aspect: Hideo Kojima Career aspirations
	Relation between Entities	Newsworthy relation between two or more entities	Joe Biden and Rishi Sunak, political friendship

Figure 4.4: Schematic version of finalized framework

To exemplify, instances from the corpus will be used. The events referenced in parentheses convey dependencies between the events. For example, in the annotations where Event1 reference Situation1 in parentheses, Event1 is dependent on Situation1. Concerning the viewpoint, the format is "Which event the viewpoint is geared towards, Who made the viewpoint, the viewpoint".

Example 1: Kanye West reportedly no longer a billionaire as companies cut ties

A news article covering Kanye West's finances declining after companies cut ties with the artist following antisemitic and racist statements was published by The Guardian in October 2022. This article is included in the corpus ID:115.

Kanye West reportedly no longer a billionaire as companies cut ties

For years critics have denounced Ye for his rightwing views and comments – but only now are they costing him his career

The ending of Ye's relationship with Adidas has reportedly cost him his billionaire status. Photograph: Mark J Rebilas/USA Today Sports

Figure 4.5: The Guardian: Kanye West reportedly no longer a billionaire as companies cut ties

The first paragraphs in this article, depicted in figure 4.6, indicate the main event and topic covered in the article. The main event that spiked the controversy and newsworthiness of the story is the controversial statements made by Kanye West on numerous occasions.

In the span of two weeks, Kanye West has lost his talent representation, connections to major fashion houses and other lucrative relationships over recent anti-Black and antisemitic comments.

Figure 4.6: The Guardian: Kanye West reportedly no longer a billionaire as companies cut ties

The information presented in the first paragraphs, illustrated in figure 4.6, derives a key event which instigates and is a driving force for the other events mentioned:

Event1: Kanye West expresses antisemitic and racist statements;

> As sports brand Adidas ended its estimated €250m partnership with West on Tuesday, reportedly costing the Black American rapper his billionaire status, many are asking if the fashion and music mogul's actions have ended his decades-long career.

Figure 4.7: The Guardian: Kanye West reportedly no longer a billionaire as companies cut ties

From the information presented in the second paragraph, depicted in figure 4.7, it is also possible to identify dependent events that occurred because the first event did:

Event2: Companies cut ties with the artist (Event1);

Event3: Kanye West's finances decline (Event2);

The controversial statements are exemplified multiple times later in the articles. Additionally, are other controversies involving the artist mentioned. However, annotating each statement as an event or viewpoint would not be necessary when the main topic of this article is the consequences of the controversies in which Kanye West is involved.

> Ye's comments and actions have not only taken a toll on the artist's social standing - they have also hurt his financial standing palpably.
>
> As of Tuesday Ye was no longer featured on the Forbes billionaires list as the ranks of those who were done working with the mogul swelled.

Figure 4.8: The Guardian: Kanye West reportedly no longer a billionaire as companies cut ties

In addition to detecting events, the framework is developed to capture entities and aspects of entities. The final section of the news article solidifies that this news article is about an entity, Kanye West, and an aspect of this entity. This derives the following annotation:

Entity: Kanye West;

Entity Aspect: Finances;

To summarize, the complete annotations of the news article would be the following:

Event1: Kanye West expresses antisemitic and racist statements;

Event2: Companies cut ties with the artist (Event1);

Event3: Kanye West's finances decline (Event2);

Entity: Kanye West;

Entity Aspect: Finances;


```
"115": {
    "id": "115",
    "title": "Kanye West reportedly no longer a billionaire as companies cut ties",
    "url": "https://www.theguardian.com/music/2022/oct/25/kanye-west-ye-billionaire-companies-cut-ties-adidas",
    "summary": "Kanye West\u2019s finances decline after several companies cut ties with artist",
    "iptc": [
        "arts",
        "culture",
        "entertainment and media"
    ],
    "events": {
        "News115-Event1": {
            "id": "News115-Event1",
            "event": "Kanye West expresses antisemitic and racis statements"
        }
    },
    "dependent_events": {
        "News115-Event2": {
            "id": "News115-Event2",
            "event": "Companies cut ties with artist",
            "depends_on": "News115-Event1"
        },
        "News115-Event3": {
            "id": "News115-Event3",
            "event": "Kanye West finance decline",
            "depends_on": "News115-Event2"
        }
    },
    "entities": [
        "Kanye West:"
    ],
    "entity aspects": [
        "finances"
    ]
},
```


Figure 4.9: Dataset: Instance 115

Example 2: Testimony of boy who witnessed mothers murder leads to life sentence in India

In October 2022, The Guardian published a news article about a boy in India testifying against his father and paternal grandparents for the murder of his mother. This article is included in the corpus under ID:131.

Testimony of boy who witnessed mother's murder leads to life sentence in India

Court sentenced boy's father and grandparents for killing his mother over demands for higher dowry

A protest after a woman was allegedly murdered by her in-laws for dowry in Beawar, India in October 2021. The National Crime Records Bureau of India said more than 7,100 dowry deaths were recorded in 2019. Photograph: Sumit Saraswat/Pacific Press/REX/Shutterstock

Figure 4.10: The Guardian: Testimony of boy who witnessed mothers murder leads to life sentence in India

Both the headline and first section of the article convey the key event that acts as the driving force for the remaining events to occur, shown in figure 4.10 and figure 4.11. The key event in this article would be the murder of the boy's mother.

> The testimony of an Indian child who witnessed his father and grandparents murder his mother in the living room of his home when he was four years old has led to their imprisonment for life.

Figure 4.11: The Guardian: Testimony of boy who witnessed mothers murder leads to life sentence in India

Additionally, this section identifies a dependent event: the boy witnessing his mother's murder. Deriving the annotations:

Event1: Boys father and paternal grandparents murder his mother in India;

Event2: Boy witnesses his father and paternal grandparents murder his mother in India (Event1);

Furthermore, one aspect that is not mentioned in the headline or first section is the occurrence of the murder victim's accusation before passing. The boy's mother was able to provide police with a statement accusing the boy's father and paternal grandparents, illustrated in 4.12.

> The case against the accused was strong. There was Sadhna's dying declaration to the police while she lay in hospital with almost 100% burns. "Only her face remained," said Amar Singh Tomar, a district government counsel.
>
> "Luckily, the doctors rushed to get the police to her bedside in time for her to give a statement accusing her husband and in-laws of setting her on fire, Tomar said. "She died minutes later."

Figure 4.12: The Guardian: Testimony of boy who witnessed mothers murder leads to life sentence in India

This derives the annotation:

Event3: Mother accuse boy's father and paternal grandparents before dying (Event1);

The previous events lead to the trial of the father and paternal grandparents. Furthermore, a trial is itself an event, which consists of several sub-events. One of the sub-events heavily focused on in this article is the boy's testimony, as seen in figure 4.13.

> "Aryan told the court that his father had come home drunk and started beating mummy. His grandparents joined in. Then they got some kerosene from the kitchen, which they threw on mummy. His grandfather said: 'Go on, light the match,' to Aryan's father. He saw it all, he saw his mother being burned alive when he was so small," said Tomar.

Figure 4.13: The Guardian: Testimony of boy who witnessed mothers murder leads to life sentence in India

This information derives the annotation:

Event3: trial against father and paternal grandparents (Event1);

And as mentioned previously, the following sub-event.

Event4: Boy gives testimony against father and paternal grandparents (Event3);

Lastly, the final event to occur is the sentence. Based on the mother's statement before dying and the boy's testimony, the boy's father and paternal grandparents received a life sentence for the murder. This information is depicted in the headline of the story and figure 4.10. The sentence delivery is a sub-event of the trial. This derives the annotation:

Event5: Father and grandparents receive life sentence (Event3);

To summarize, the full annotation of this news article is as follows:

Event1: Boys father and paternal grandparents murder his mother in India;

Event2: Boy witnesses his father and paternal grandparents murder his mother in India (Event1);

Event3: trial against father and paternal grandparents (Event1);

And as mentioned previously, the following sub-events.

Event4: Boy gives testimony against father and paternal grandparents (Event3);

Event5: Father and grandparents receive life sentence (Event3);

```
"events": {
    "News131-Event1": {
        "id": "News131-Event1",
        "event": "Boy\u2019s father and paternal grandparents murder his mother in India"
    }
},
"dependent_events": {
    "News131-Event2": {
        "id": "News131-Event2",
        "event": "Boy witness his father and paternal grandparents murder his mother in India",
        "depends_on": "News131-Event1"
    },
    },
    "News131-Event3": {
        "id": "News131-Event3",
        "event": "'Mother accuse boy's father and paternal grandparents before dying'",
        "depends_on": "News131-Event1"
    },
    "News131-Event4": {
        "id": "News131-Event4",
        "event": "trial against father and paternal grandparents",
        "depends_on": "News131-Event1"
    }
},
"sub_events": {
    "News131-Event5": {
        "id": "News131-Event5",
        "event": "Boy gives testimony against father and paternal grandparents",
        "depends_on": "News131-Event2"
    },
    "News131-Event6": {
        "id": "News131-Event6",
        "event": "Father and grandparents receive life sentence",
        "depends_on": "News131-Event5"
```


Figure 4.14: Dataset: Instance 131

Chapter 5

Results

This Chapter will detail the results and analysis. Firstly, the results from the analysis of the empirical study with human annotators will be revealed. Both the computational analysis conducted to measure agreement between participants in task 1, and the manual qualitative analysis of the topics generated by the participants in task 2.

Furthermore, a complete representation of the framework will be given. Additionally, the dataset, which is a result of the validation of the framework, will be detailed.

A discussion of the results and which indications can be drawn will be given in the following chapter.

5.1 Analysis of Empirical study

An empirical study with human annotators was conducted early in the thesis work. The study consisted of the participants completing two annotation tasks. The following section will detail the analysis of the annotations provided for each task. Task 1 was analysed computationally using an inter-rater agreement measure. Task 2 was analysed qualitatively by focusing on three main areas.

5.1.1 Task 1: Computational agreement

As mentioned in previous chapters, an empirical study which included an investigation of the level of agreement between human annotators, was conducted. In this study, the participants were asked to annotate news using IPTC NewsCodes. Following the collection of annotated news stories according to the IPTC NewsCodes, the results were analysed using computational methods like Fleiss Kappa. Fleiss Kappa is a statistical method to measure inter-rater agreement between two or more raters *Fleiss* (1971). The measure is detailed in the Background Chapter.

As mentioned, to conduct this statistical analysis of agreement between raters. Python and Python libraries were utilized. This was to achieve a precise score that could be used on a small sample of the annotations and the complete retrieved information. Furthermore, the statsmodels library was utilized, specifically the built-in function for calculating Fleiss Kappa *Seabold and Perktold* (2010).

The text values provided by the participants were cleaned to ensure consistency in formatting. The text values were also transformed into numerical values using a hash function.

The Fleiss Kappa agreement score that was given was set to 0.62 after the computational analysis. This score is classified as substantial agreement *Fleiss* (1971). The annotations were also subject to a simple short analysis to verify that the Fleiss Kappa Score seemed accurate. Based on this manual verification, the score appeared correct.

The level of agreement can indicate that annotations based on a typology framework can be useful and effective. However, the taxonomy used in this analysis does not offer fine-grained classification or annotation of news stories.

5.1.2 Task 2: Qualitative Analysis

The qualitative analysis consisted of several areas of consideration. However, the main purpose was to investigate the agreement between annotators. Other areas of analysis consist of evaluating the preciseness of the annotations and investigating bias in the annotations.

Observed agreement

As the agreement between participants in textual documents with a variation in language, length, and preciseness is very difficult to calculate computationally, the investigator evaluated the annotations on similarity. Generally, there was a high variation in formatting, length and inclusion of information in the annotations.

Although the annotations are not computationally evaluated, based on comparisons between the annotations, there was a lower level of agreement. The participants tended to identify different aspects as the focus of the news article. Furthermore, the annotations have different levels of how much information was included. Meaning the participants had different views of what fine-grained and specific annotations are.

Concerning the news item about Nicholas Cage's statement about actors being able to use firearms following the accidental fatal shooting involving Alec Baldwin. The statements were made while he was attending a roundtable discussion about banning real firearms from movie sets. It was clear that the participants had different views about what the article focused on.

Annotation 1: In comments responding to the fatal shooting on the set of Rust, Nicholas Cage states that movie stars needs to know how to use a gun

Annotation 2: Actor gives his point of view on how to prevent future fatal accidents on movie sets

Annotation 3: Banning firearms on stage

Annotation 4: After a tragic shooting incident, a famous actor believes that it is a part of the profession as an actor to be able to handle weapons.

Annotation 5: Nicolas cage asks for better gun safety in movie industry

Annotation 6: Use of firearms on filmsets, actors need to know how to use guns to ensure on-set safety

Annotation 6: Nicolas Cage requires weapons training for actors

A few of the participants identified which actor gave the statement that is being discussed. The identification of the person giving the viewpoint indicates that the participant views not only the statement as important but the individual as important as well. The annotations reveal disagreement about whether the identification of the actor is important information to achieve fine-grained classification. Additionally, the actual viewpoint of Nicholas Cage was also interpreted differently; some participants relate his viewpoint to gun safety and weapon training. In contrast, others render the statement as an aspect of the actor's profession. One participant identified the topic of the roundtable discussion: whether firearms should be banned from movie sets. The statement was given in the context of the fatal accidental shooting, and there is an evident disagreement between the annotators on whether the context of the statement is vital for fine-grained classification. Although all of the participants have identified at least one aspect of the article, there is a high level of disagreement concerning the focus of the news item. This level of disagreement was present in the majority of the annotations.

Precision in annotations

The precision of the annotations varied to a very large extent. Some of the participants utilized keywords for some stories, while others submitted summary-like annotations. For example, in the story where James Sinclair is convicted of murder and attempted murder, some annotations are extremely precise, while others are not. Below are examples of the more precise annotations provided.

Annotation 1: Man found guilty of murdering a woman at a bus stop in East Ham and attempted murder of another woman gets 41 years in prison

Annotation 2: James Sinclair brutally attacked two women. The lead investigator encourages women to report suspicious behaviour.

As mentioned, some of the annotations provided were not as precise as previously exemplified.

Annotation 3: Murder, womens safety

Annotation 4: Man commits violent crime and is punished

Annotation 5: Murder

The annotations depicted above offer very little specific information about the event. These annotations could not directly correlate to this event, as opposed to the events provided in the first exemplification. This shows a varying understanding of what specific topics entail and what the concept of fine-grained classification means. This level of preciseness variation was present in many of the annotated stories. This contrast was not as apparent in the news stories that could be considered less complicated.

In the news item that covers female celebrities embracing their grey hair, and the positive attention that followed on social media, the dissimilarities were not as present. In these annotations, most participants included the aftermath that played out on social media; however, not all.

Annotation 1: The hashtag 'grayhairdontcare' popular on Instagram and TikTok to show how women embrace their grey hair

Annotation 2: Women impower other women through hashtag greyhairdontcare to wear and grow out their naturally grey hair.

Annotation 3: Celebrities show off their grey hairs and with it creates a movement to normalize it and work against ageism

The annotations above include the movement that was created on social media after female celebrities started growing their grey hair. To achieve fine-grained classification, this information is very important. The consequences of the driving force event is the focus of the news item. The annotations below have only focused on the embracing of gray hair, rather than the movement that has been created.

Annotation 4: Celebrities embracing grey hair

Annotation 5: More women are embracing their grey hair instead of dyeing it

Annotation 6: Celebrities normalizes natural aging in which they embrace their grey hairs on red carpets

Annotation 7: Women embracing their body

This story is the news item with one of the highest preciseness levels across annotators. However, a few annotators could not identify the two key elements.

Concerning the story of a firework festival being held in Germany amid a drought and the controversy that followed, the participants achieved a high level of precision in their annotations. The key event of the news item was the firework festival being arranged despite a drought, however, the article heavily focuses on the controversy surrounding the preparation of the festival, which included spraying down dry areas to prevent fire. The annotations provided by the participants are shown below:

Annotation 1: Rhine in Flames under criticism in Germany for still taking place despite the water levels continue to drop

Annotation 2: Controversy over usage of water to entertainment when

Annotation 3: Wasting natural resources in the face of a climate catastrophe

Annotation 4: After a long summer of drought and reduced water levels, a firework festival receives great criticism.

Annotation 5: Controversy over water consumption of firework display amidst draught in Germany

Annotation 6: Critique of fireworks during drought

Annotation 7: Rhine firework display taking place despite drought

As seen in the annotations, the level of precision is quite high. Below are the annotations according to the framework. The events references in parentheses convey dependencies between the events, for example in the annotations below Event1 reference Situation1, meaning Event1 is dependent on Situation1. Concerning the viewpoint, the format is "Which event the viewpoint is geared towards, Who made the viewpoint, the viewpoint".

Situation1: Drought in Germany;

Event1: Residents in drought area asked to save water (Situation1);

Event2: Rhine in Flames firework festival staged;

Event3: Slopes surrounding Rhine hosed down in preparation (Event2);

Viewpoint: Event3, Member of German parliament, Festival should have been cancelled instead of wasting water;

Comparing the annotations made by the participants and the annotations according to the framework to measure precision reveals a relatively high level. Annotations 1, 2, 4, 5 and 6 all identified several key events covered in the news item. They all identified the focus of the news item, which is the controversy that followed the water usage that was necessary to stage the festival. While annotation 7 identified both the firework festival and the drought. The information included in annotation 3 is of the lowest level of precision, where a firework festival is referred to as entertainment and the drought is referred to as a climate catastrophe. Annotation 3 uses broad and general terms to depict the events in the news article. Although there are some variations of precision, the overall level is quite high, when compared to the benchmark annotation.

Possible bias in annotations

As mentioned previously, while annotations are performed by individuals the presence of personal bias is inevitable *Gallofré Ocaña and Opdahl* (2022).

As the participants of the empirical study are friends of the investigator, possible bias is feasible to detect.

For example, one of the stories included in the corpus is about a comedian, who is also a healthcare worker, winning the best joke award at the Edinburgh Festival Fringe. Only two of the participants chose to include his other profession in their annotations. One of these participants is a nursing student, while the other is a partner of said nursing student. Their annotations are depicted below:

Partner of Nursing Student: care worker wins funniest joke for the second time at Edinburgh festival fringe

Nursing Student: Once again, a health-care worker wins best joke award at Edinburgh festival fringe

Other annotations did not include this information, some even eluded the key event of the victory and mentioned only the Edinburgh festival fringe.

Examples below:

Example 1: Joke contest in Edinburgh

Example 2: Man wins award for best joke at festival

Example 3: Light humor, characteristic festival, dry humor, lifestyle as an ordinary worker as well as an comedian wins prize for best dry humor joke.

Example 4: Best comedian at Edinburgh Fringe Festival

Although this is a clear example of information inclusion based on personal bias, it is the only identifiable instance found without further investigation into the participants. As the investigation of bias in annotations was not the primary mean for the study, it was not accounted for when designing the tasks. Furthermore, it would involve a much more extensive process, possibly including interviews to identify the thought processes of each participant for each annotation.

5.2 Validation of framework

As seen in the exemplification of the method of procedure to conduct fine-grained news annotation illustrated in the methods chapter, the validation was conducted by doing this procedure on 224 news articles. The aim of the validation was to verify that the framework was complete. Furthermore, it was accurate and could be used effectively.

The validation was initially conducted in a Word document, consisting of a table with one news article and correlated information on each row. As shown in figure 5.1.

As depicted in figure 5.1, additional information was included in the annotations and dataset. The additional information consists of an ID, headline, link to the article on the publishing outlet's website, a short description of the news item and the Top Level IPTC categorization. This information was included to ensure the dataset would be accessible. Additionally, including this information would give future users the ability to further investigate a certain instance and comprehend the annotations that were given.

ID	Headline	Link	Short description	IPTC	Fine-grained classification
101	Russia steps up Ukraine 'dirty bomb' claim in letter delivered to UN	https://www.theguardian.com/world/2022/oct/25/russia-to-raise-ukraine-dirty-bomb-claim-at-un-security-council	A. Russia sends letter to UN about the Ukrainian 'dirty bomb' B. Several countries and professionals dismiss Russia's letter to UN concerning Ukraine supposed planned 'dirty bomb'	conflict, war and peace	**Categorical** **Event:** War in Ukraine **Dependent event (level 1)**: Russia sends letter to UN about the Ukrainian 'dirty bomb' **Dependent event (level 2)**: Several countries dismiss Russia's letter to UN about the Ukrainian 'dirty bomb'
102	Putin's wild claims of a dirty bomb show just how badly his army is faring Hamish de Bretton-Gordon	https://www.theguardian.com/commentisfree/2022/oct/25/putin-dirty-bomb-war-russian-leader-tactics	A Letter to UN about dirty bomb proves Putins warfare not successful B Putin's unconventional Warfare is not as successful as planned despite being tested by Assad in Syria	conflict, war and peace	**Event**: War in Ukraine **Dependent event:** Russian warfare not performing as expected. **Viewpoint (Hamish de Bretton-Gordon):** 'Dirty Bomb' letter from Russia to UN proves Russian warfare not performing as expected.

Figure 5.1: Annotations from word

The result of the validation process is a complete dataset published and available for other researchers.

5.2.1 The Dataset

As mentioned previously, the validation of the framework resulted in a dataset consisting of 224 news articles. The dataset is published to Figshare, for researchers to access and use. The dataset is available at: https://figshare.com/articles/dataset/Fine_grained_annotations_json/22785944.

For future usage the documents containing the annotations required to be converted to a machine-readable format, JSON was chosen. To do this, the word annotations were transferred to an Excel document, where each row was a news article and each typology was set to a column. Angelo

Salantino from UK Open University then developed a Python program to convert the Excel file to JSON.

5.3 Finalized Version of the Framework

This section will detail the finalized version of the framework. The framework consists of several typologies and sub-types of instances intended to capture the contents of news articles.

5.3.1 Events

Most news items are derived from events that have occurred or currently occurring *Sjøvaag and Kvalheim* (2019). Events is a general term and, therefore, in the framework, divided into more specific types of events, Atomic Events, Negative Events, Dependent Events, Collection of Events and Composite Events. However, the broad notion of Events is also classified. This is to associate news articles that cover different aspects of the same key event.

Atomic events

Atomic events refer to the most basic form of classification. This classification applies to events that do not necessarily have newsworthy sub-events.

Negative Events

While news items often discuss events that have occurred, they can also discuss the absence of an expected event *Sjøvaag and Kvalheim* (2019). The absence of an event itself has to be newsworthy. A negative event captures the notion of expressing agency by not performing an act or creating an event. When individuals refrain from completing an action that is expected of them, it is a form of expressing agency. When this expression of agency is newsworthy, it can be classified as a negative event *Payton* (2018).

Dependent Events

The notions of background and context are used in journalism to indicate additional elements that are necessary to explain, for example, the event that is the focus of a news story. Hence, Dependent Events are events that, from a journalistic point of view, cannot be discussed without reference to one or more other events. For instance, an article about a verdict in a trial needs to say something about the trial itself. While a sub-event (e.g., verdict) always depends on the parent event (e.g., trial), also other forms of event dependencies exist. For example, a trial can only take place if a referral to trial is issued.

Collection of Events

News articles that cover multiple events that are in some capacity related to each other are classified as a collection of events. For example, a news article that gives a synopsis of recent football matches and results or the one described earlier that talks about multiple drone sightings in Norway. In the former case, each football match is discussed as separate event. Each match would then be an event in the same collection.

Composite Events

Composite events are events that comprise of sub-events that may themselves be the focus of a news story. For example, as already discussed, a trial is a composite event composed of many sub-events. The events of a trial are predetermined and always consist of smaller sub-events, such as closing statements and sentencing.

Additionally, war is an example of a larger event which entails many sub-events. However, war is more unpredictable, as the sub-events are not predetermined, as it is in a trial.

5.3.2 Situations

A situation classifies a newsworthy state of affairs. While the other classification centre around the notion of an event occurring, a situation is often a consequence of an event. However, a situation can be the driving force of other dependent events, for example, the article that details the books that are missing from Peterborough Library. This news item focuses on the situation of the missing books and a decline in borrowers following the pandemic. Because of the large number of missing books and the decline in the number of borrowers, the library removed fines in the hope that the borrowers and books would return.

5.3.3 Prediction of Events

While many news stories cover events that have previously occurred or are occurring currently, many news items discuss the prediction of events. Either a study has been released including numbers that indicate possible coming states or speculation of an event that can occur.

Not all predictions are based on studies or numbers released. Some predictions are based on previous events, for example, the prediction of the selling price in an impending auction of a recently recovered baroque painting. In this news article, the prediction was based on the selling prices of other previously auctioned baroque paintings.

5.3.4 Viewpoints and debate

News articles often centre around events, however, it is very common for journalists to include viewpoints from individuals who are in some capacity a party of interest. Often viewpoints are a direct response to the event.

For example, in the story about a Melbourne street artist being charged with criminal damage after painting a mural of a deceased cricket player on an abandoned building. Following the trial, the street artist criticised Victoria Police for pursuing charges. In the story of Conor Benn being relinquished his boxing licence, the boxer was later interviewed and claimed the British Boxing Board of Control conducted their investigation with a biased procedure.

As seen in the previous example, viewpoints can often be made by individuals involved in a certain event. However, many news articles include expert commentary on an event or situation. The individuals who contribute with viewpoints are experts within a field of study relevant to an event or situation.

However, viewpoints frequently respond to another individual's previous statement or viewpoint. If an article includes several opposing viewpoints, the basis for the viewpoints becomes a topic for debate.

5.3.5 Entity

As discussed previously events are usually the driving force of news items, however often a person or organisation can be the topic of a news article. Entity Recognition has for many years been a topic of research, and there are established automatic methods to complete such identification. However, just identifying an Entity in a news article would not provide the precise annotations that the framework aims to produce.

According to the framework, the entities in a story are only identified if it is the focal point of the news item. For example, if a third party gives a viewpoint on a topic, the individual providing commentary is the focus of the news story, rather than the viewpoint itself. Therefore the Entity classification in this framework is only used where the individual or organisation is at the centre of the story.

Entity Aspect

In addition to classifying newsworthy entities that are the centre of a news story, the framework accounts for aspects of the entity that are vital in the news item. Entities themselves can be newsworthy, but often they are supported by an entity aspect to induce newsworthiness. An entity aspect can, similar to entities, be various facets of an entity. For example, an aspect can be an individual's medical diagnosis or career. Such as an interview The Guardian conducted with Hannah Fry where the focus of the news item was her cancer diagnosis and career as a mathematician.

Relation between Entities

Another facet of entities that can be newsworthy and are often detailed in news items is the relations between two or more entities. News stories about relations between entities can often appear as if the article is actually about one of the entities, such as the news article about the Honduran Maradonna. However, the focus of the relation can also be obvious, such as the news item detailing the relationship between Italian prime minister Giorgia Meloni and another politician and previous prime minister, Silvio Berlusconi. This story details a supposed power struggle between the two individuals.

5.3.6 Categorical Issues

As the framework strives to produce accurate and very fine-grained classifications, it is also valuable to identify broader topics that are present in news items. In particular, topics that tend to be routinely newsworthy, independently of specific circumstances, such as a country's economic situation. Categorical Issues capture broad topics and make it possible to establish links between news items focusing on different events that are somewhat related. For example, as tensions have risen between NATO and Russia recently, many events are occurring related to this tumultuous relationship. For example, Britain requested to fly over Russian territory briefly, and to this, Russia responded that this action would be considered a sign of provocation. The frequent sightings of Russian drones in Norway is another event that is directly correlated to the relationship between NATO and Russia. However, the two events mentioned would require the categorical issue classification to be connected in a search process. Both stories have a categorical issue classification, Relations between NATO and Russia.

Chapter 6

Discussion

This chapter will discuss the results that were given in the previous chapter. Firstly, the results from the analysis of the empirical study will be discussed. The topics of the discussion will be the agreement calculated and observed and how precise the annotations were. Furthermore, whether bias was present in the annotations.

Following this, the concepts established in the framework will be discussed and correlated to relevant research within several fields, such as event extraction and entity recognition.

Furthermore, a discussion of how the work related to this thesis has contributed to the discourses relevant. Such as the areas of use of both the framework and dataset.

Lastly, the limitations that were imposed on this thesis will be rendered.

6.1 Analysis Empirical study

The following section will discuss the findings of the analysis of the empirical study yielded.

6.1.1 Agreement between Human annotators

As the analysis of the annotations collected in the empirical study revealed, annotators had a higher level of agreement when using the IPTC NewsCodes Taxonomy. Using an inter-rater agreement measure, the agreement was calculated to be a Fleiss Kappa score of 0.62, which indicated a substantial agreement. Additionally, when asked to provide self-generated specific annotations for the same corpus, the level of disagreement was high. Participants had vast opinions of what fine-grained specific annotations included. Based on these calculations and observed agreement from the self-generated annotations, the conclusion that a framework is practical

and useful for the manual annotation of news can be drawn. However, the framework utilized has to produce fine-grained classification, unlike IPTC NewsCodes. The annotations produced by using the IPTC NewsCodes taxonomy are very generic *IPTC* (2022). Additionally, to further support this statement, a large number of the annotations generated purely by the participants were very precise. Although there were some differentials concerning preciseness, many of the annotations were quite precise at a superior level.

6.1.2 Precision and bias in annotations

The results from the analysis of the empirical study reveal at least one instance of bias present in the annotations generated by the participants. However, because it is only one instance identified, it is difficult to generalize that there was a general bias present by the participants while annotating. However, because an action is performed by an individual, the presence of bias is highly likely to occur.

The precision, however, was very varied but at an overall level quite high. A portion of the annotators identified all relevant aspects of a news article, while some identified irrelevant aspects.

6.2 Concepts of the Framework

The concepts that are established in the framework are detailed above, this section will discuss the concepts and relate them to existing theory and work.

Events are the driving force of news stories and therefore are the greater part of the framework typologies related to events. Although, not all events that occur at any given time are covered by the media. The events must possess a certain level of newsworthiness *Sjøvaag and Kvalheim* (2019). As Borgo et al. concluded in their research, an ontology benefits from linguistic concepts such as events and qualities *Borgo et al.* (2022). Qualities in this context can be the relation between events, such as dependencies.

Research within Event Extraction is based on identifying events in text documents and extracting the event itself and relevant supporting information. However, this action of event extraction does not capture all mechanisms that are required to perform fine-grained classification. The concepts of atomic events, collection of events, composite events, and the notion of dependencies established in the framework are not present in standard Event Recognition research. Additionally, event extraction is hindered by the predetermined patterns that cannot capture the variation and complexity in natural language *Liu et al.* (2020). However, event extraction is an important field of research, and techniques can be very useful in achieving fine-grained classification.

Named Entity Recognition and Entity Linking are other fields of research that can offer valuable insights into fine-grained classification *Linhares Pontes et al.* (2020) *Marrero et al.* (2013). However, these techniques alone cannot provide the fine-grained classification desired. Techniques derived from Named Entity Recognition and Entity Linking can support computational fine-grained classification models.

Several key events have a Unifying Factor, a concept identified by Carriero et al. *Carriero et al.* (2021). This concept is related to the notion of a collection of events. For the collection of events classification to be applied, a news article must contain several key events with the same feature for them to be considered a collection. For example, if an article summarizes the recent results of football matches in the Champions League, each match is an atomic event, and the collection is Champions League Football matches.

Payton discusses the expression of agency by not performing an action *Payton* (2018). His understanding of the absence of events and actions can be correlated to the concept of negative events. Where an event not occurring becomes newsworthy. For a negative event to be a topic of discussion or news, there needs to be an element of omission, as not performing an action itself is not an expression of anything. Not performing an action needs to be characterized by omission, where there is a deliberate decision not to perform the act. The omission could also become apparent if an action is expected and an agent fails to perform it *Payton* (2018).

In the context of the framework, viewpoints are considered abstractions of positions. Similarly, to the collection of events, the concept of a unifying factor is relevant in the discussion of viewpoints as well. While viewpoints are considered abstractions of positions, these positions have a unifying factor *Carriero et al.* (2021). All abstractions of positions are derived from one common topic or event. The driving force of the debate.

Even though the framework offers several established and clear typologies, an interpretation process while annotating is still present, however, the framework is consistent with the journalistic writing style *University of Arizona Global Campus Writing Center*. This journalistic writing style facilitates and limits this interpretation process that will occur between different annotators. Furthermore, the framework is consistent with the method of procedure media analysis utilizes while analysing news.

The proposed framework does allow for a certain interpretation process. This interpretation process is however facilitated by the strict journalistic writing style that is present in every media outlet. This journalistic writing style includes several important concepts, such as simplicity and brevity, precision, factuality and objectivity, fairness and balance. Furthermore, when writing a news item, the journalists are required to build the article

conforming to the inverted pyramid, where the events and topics covered, are presented according to newsworthiness. The most important events and facts are always presented first and the additional information is presented in descending order according to the importance concerning the story *University of Arizona Global Campus Writing Center*.

6.3 Contribution

As mentioned previously the aim of this thesis was to make several contributions to the discourse of fine-grained news classification.

6.3.1 Level of agreement between annotators

By conducting the empirical study with annotators, a measure of agreement between human annotators was computed. This study contributes to the discourse of news categorization in the sense that it has identified the disagreement present between annotators without utilizing a framework. Furthermore, the detection of the high level of agreement between annotators given a framework indicated that a framework is beneficial for further research within the field of fine-grained news categorization.

6.3.2 Meaning of Fine-Grained News Classification

One of the more vital contributions this thesis has made is clarifying fine-grained news classification and what it entails. As mentioned previously, fine-grained news classification requires several mechanisms to work together to achieve accurate and precise annotations. The mechanisms presented in the framework are not necessarily all present in every news item, but very rarely only one occurs. Therefore, it is important to consider multiple computational mechanisms to support fine-grained news classification. Additionally, when conducting the annotation of news items, there must not be aspects that are not captured because of a deficiency in the framework. The framework presented in this thesis is able to meet these requirements, and this conclusion is made based on the validation of the framework.

To provide a framework for fine-grained news classification it is important to consider how news items can differ from outlet to outlet as well. In the Guardian news articles including viewpoints from an individual was much more common than in the news articles retrieved from Aftenposten. Furthermore, the different mechanisms would be more useful depending on the outlet in question and how they produce and structure the news articles.

6.3.3 Framework: Areas of Use

The complete framework has several potential areas of use. The following paragraphs will detail what the contribution of the framework entails.

Journalists are often bound by organizational standards when categorizing news, and because there is no shared practice between news outlets, the framework can contribute to the resolution of the issues that follow. Additionally, as found in the empirical study, annotations can be influenced by personal bias, such as personal relationships and other factors. A business standard rather than countless organizational standards would be very beneficial for search purposes. Additionally, having a business standard practice could resolve the issue of biased news categorization.

Therefore, the framework could be implemented for manual categorization in the publication process.

Furthermore, the development and implementation of an automatic computational model based on the framework for news categorization would be very beneficial for news outlets. The bias categorization could be eliminated. Furthermore, manual classification is not scalable, and therefore, categorizing a large corpus of news would be very time-consuming.

Additionally, some of the typologies identified could be separated, and the concepts could be applied to other areas of research. Such as the concepts that are established related to events in the framework could be translated and used in event extraction research. The concept of dependencies established in this framework could lead to novel Event Extraction research. Furthermore, the separation between dependent events and sub-events can also be useful to achieve more nuanced and precise event extraction. Although in the framework and news production, the event identification is heavily influenced by newsworthiness, these concepts of dependencies can be utilized in domains of event extraction that are not news related.

Furthermore, the concepts related to entities presented in the framework could be useful for the areas of research concerning entity recognition and entity linking. Classifying the entity aspects presented in a text document could provide a more extensive knowledge base for entity recognition and entity linking studies. As relations between entities are similar to entity linking techniques, how the relationships are identified and the formal representation used in the framework can prove useful. Furthermore, the identification of which relation the entities have would provide the knowledge base with further insight.

The Framework could also provide insight to research within Topic Modelling. This method of classification does not alone provide fine-grained classification, techniques from this field can be used as a tool. Furthermore, the inclusion of the Categorical Issue classification can provide clustering

methods, such as topic modelling, more precise clusters, that are not bound by general and vague topics.

6.3.4 Dataset: Areas of Use

One of the main areas of use for the dataset is to be used as a training dataset for a news classification model. Future work within this field consists of developing automatic models that are able to classify news as it is done in the dataset. This model would then be trained on the manually annotated dataset produced in this thesis. Furthermore, the dataset can be used as an evaluation method while developing automatic fine-grained news classification models.

As mentioned previously, topic models are one of the most common methods for news classification in current research. Topic models are able to identify latent topics in text documents, however, the results of these models are coarse-grained. The use of the dataset in future computational methods would be superior to the coarse-grained classification that has been achieved thus far. One of the contributions of this thesis is therefore a useful data set that can be used as a benchmark and gold standard for future fine-grained news classification. Therefore, the dataset lays a good foundation for further research within the field of fine-grained news classification.

As mentioned in the previous section, the framework and concepts established can be useful for research within entity linking. The entities identified by using the framework for annotation, which are included in the dataset, can be used in formal entity representation. For example, the entities and entity relations could be added to a knowledge base.

As mentioned in the previous section, the framework could provide useful insights and concepts to research concerning event extraction. As discussed by Huang et al., there is a lack of available labelled training data *Huang et al.* (2018). Extracting the events identified in the corpus and relevant related data could be used as labelled training data for event extraction. Although it would be useful training data for event extraction within the news domain, it could also be used for training generic non-domain-specific models.

6.4 Limitations

Because of the limited number of participants included in the empirical study, the data that was derived also suffer from limitations. As only 8 participants agreed to complete the study, the generalization of results could be considered weak and require more human annotators to verify completely. A reason for this was very few people accepted to participate, learning that the tasks required to complete were time-consuming.

Furthermore, validating the framework would prove even more useful had it involved several human annotators using it to categorize news. However, because of time-constrains, completing this validation was not possible. Furthermore, conducting a validation using several human annotators and a large corpus would involve funding and possibly utilizing a tool like Amazon Turk.

Chapter 7

Conclusions and Future Work

This chapter will present the conclusions that have been drawn based on the discussion and results presented in the previous chapter. Additionally, the future work that is needed will be discussed.

7.1 Conclusion

After completing the work that is included in the thesis, several conclusions can be drawn based on the results and discussions presented in the previous sections.

The preliminary empirical study that was conducted revealed a high level of agreement between human annotators when using the IPTC NewsCodes, and a low level of agreement between annotators when asked to generate their own topics. Moreover, the empirical study results indicate that using a framework for annotations is beneficial. However, the taxonomy provided by IPTC is too coarse-grained and does not facilitate the production of specific annotations. Furthermore, from the empirical study, tremendous insights were made concerning the concept of fine-grained classification. It was clear that fine-grained classification requires several mechanisms to work together.

These mechanisms were captured in the proposed framework, which was developed through several iterations before being completed. In each iteration stage, the framework was validated, until the validation process revealed no shortcomings. The final iteration was then validated by conducting annotations of a corpus consisting of 224 news articles. All aspects present in these news articles were covered by the framework, which leads to the conclusion that the framework is complete and can facilitate fine-grained and accurate news classification. A result of the validation process was a complete dataset that was published in a JSON format on Figshare. This dataset is now available for other researchers and can be used as a training dataset on an automatic fine-grained classification model.

7.2 Research Questions

In the introduction, two research questions were presented. As the thesis work is concluded these questions have been answered to a great extent.

RQ1: What does it mean to perform fine-grained news classification?

As established, performing fine-grained news classification requires several mechanisms. In order to capture all aspects of a news item, these mechanisms have to be very thorough in the definition. Furthermore, fine-grained news classification requires a practice that can not be subject to bias, as the manual categorization methods existing are.

RQ2: Is it possible to produce a framework for fine-grained news classification that is accurate and complete?

Based on the validation of the framework, conducted by annotating news items based on the established concepts, a conclusion can be made regarding this research question. The validation of the framework reveals that the typology is complete, as we were able to cover all relevant fine-grained topics in the news corpus. To prove accuracy comprehensively would require another empirical evaluation, which is outside the scope of this thesis. However, a preliminary evaluation conducted by the supervisor and members of his research team has provided an initial validation of the accuracy of the annotated corpus.

7.3 Future Work

As mentioned in the previous chapter, this thesis has several contributions. However, there is still work that has to be completed, within the field of fine-grained news classification. A suggestion for complete validation of the framework would be to conduct a validation process using several human annotators. Conducting this process would also expand the dataset, given that the human annotators can produce correct annotations. Moreover, having a larger dataset would be beneficial for training models.

Furthermore, developing a model that is trained on the dataset would be considered a great progression in the domain of fine-grained news classification. It would be a monumental contribution to the discourse, allowing for a more in-depth and specific analysis of news.

Appendix A

Appendix

Hello!

I have just started writing my master thesis in Information Science. As a part of the master thesis, I am conducting an empirical study to measure agreement between human annotators in the context of topic assignments.

Attached to this email is a folder consisting of 25 news articles and a form. Please use the PDF documents and refrain from visiting the website for news articles. The news articles have a number in the filename which corresponds to a row in the form. Please make sure the correct topics are filed in the correct row, column. The study will be conducted in English.

After reading each article, please complete the following tasks;

1. Visit https://www.iptc.org/std/NewsCodes/treeview/mediatopic/mediatopic-en-GB.html and select the topics you find suitable for the news article from the top level to the lowest level. This link includes several top-level topics, to access the lower levels, press the triangle. Indicate the levels using > and ».
2. Generate your own more specific topics. The topics can be short sentences, keywords or expressions. The self-generated topics can be as many or as few words as you like, as long as you feel it captures the contents of the news articles. Try asking your selves what the news article is really about.

For example:

News article 1: "French justice ministry under pressure to explain jail go-karting

- IPTC topics: "crime, law and justice, prison
- Self-generated topics: controversy over go-karting event at Fresnes prison

Informed Consent Form

Measuring agreement of news classification between human annotators
Principal Investigator (PI): Maia Lunde Gjelsvik
Contact: mgj027@uib.no
Supervisor: Prof Enrico Motta

You are invited to be part of an empirical study in conjunction with my master's thesis, where the objective is to measure to which degree human annotators agree on manual news classification. Please read through the information provided in this document, and contact the PI if anything is unclear.

General Information about the study
In the study, the participants are asked to annotate news articles in two different ways. The participants are asked to read news articles provided by the PI. Following, the participants will select topics they find suitable to describe each article from IPTC NewsCodes. Additionally, the participants are asked to generate their own topics for each news article.
The purpose of this study is to measure the agreement of news classification between human annotators. In order to investigate this, the participants will be provided with news articles that they will annotate in different ways.

Note that, no personal information will be stored, and it will not be possible to identify you based on the collected data.

What will I be asked to do?
You will receive all relevant documents by email or by any other preferred means.
The participants of the study are asked to read news articles provided by the PI. Following, the participants will select topics they find suitable to describe each article from IPTC NewsCodes, a link is provided. Additionally, the participants are asked to generate their own topics for each news article.

How will the data I provide be used?
The findings of this study will be used in the PI's master's thesis.

Your right to withdraw from the study
You have the right to withdraw your participation at any time before the study commences. To do this, notify the PI via email. You have the right to ask for the data provided by you to be removed. However, after the thesis has been submitted it is not possible to withdraw your participation.

How do I agree to take part?
To participate in our study, you will need to sign and return the consent form by email before the study starts.

We appreciate your time and thank you for taking part in this study.

Is the participant illiterate?

Ja

Name of Participant

Fornavn Etternavn

Dato

Måned Dag År

Name of Reseracher

Fornavn Etternavn

Dato

Måned Dag År

Bibliography

Allahyari, M., S. Pouriyeh, K. Kochut, and H. Arabnia (2017), A knowledge-based topic modeling approach for automatic topic labeling. 2.1.1

Alpaydin, E. (2014), *Introduction to Machine Learning*. 2.1

Blei, D. (2012), Latent Dirichlet Allocation, *Communications of the ACM*, *55*(4), 993–1022. 2.1.1, 2.1.1

Blei, D., A. Ng, and M. Jordan (2003), Latent Dirichlet Allocation, *Journal of Machine Learning Research*, (3), 993–1022. 2.1.1

Borgo, S., R. Ferrario, A. Gangemi, N. Guarino, C. Masolo, D. Porello, E. Sanfilippo, and L. Vieu (2022), Dolce: A descriptive ontology for linguistic and cognitive engineering. 2.4, 6.2

Brown, S., C. Bonial, L. Obrst, and M. Palmer (2017), The rich event ontology, in *Proceedings of the Events and Stories in the News Workshop*, pp. 87–97, Association for Computational Linguistics, Vancouver, Canada, doi:10.18653/v1/W17-2712. 2.4

Carriero, V. A., A. Gangemi, A. G. Nuzzolese, and V. Presutti (2021), An ontology design pattern for representing recurrent situations., *In Advances in Pattern-Based Ontology Engineering*, pp. 166–182. 2.4, 6.2

Chang, J., S. Gerrish, C. Wang, J. Boyd-graber, and D. Blei (2009), Reading tea leaves: How humans interpret topic models, in *Advances in Neural Information Processing Systems*, vol. 22, edited by Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams, and A. Culotta, Curran Associates, Inc. 2.1.1

Doogan, C., and W. Buntine (2021), Topic model or topic twaddle? re-evaluating semantic interpretability measures, in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 3824–3848, Association for Computational Linguistics, Online, doi:10.18653/v1/2021.naacl-main.300. 2.1.1

Fleiss, J. L. (1971), Measuring nominal scale agreement among many raters, *Psychological Bulletin*, *76*(5), 378–382. 2.6, 4.1.4, 5.1.1

Gallofré Ocaña, M., and A. Opdahl (2022), Supporting newsrooms with journalistic knowledge graph platforms: Current state and future directions, *Technologies*, *10*(3), 68, doi:10.3390/technologies10030068. 2.5.1, 5.1.2

Guarino, N., D. Oberle, and S. Staab (2009), *What Is an Ontology?*, pp. 1–17, doi:10.1007/978-3-540-92673-3_0. 3.3

Harrando, I., P. Lisena, and R. Troncy (2021), Apples to apples: A systematic evaluation of topic models, in *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)*, pp. 483–493, INCOMA Ltd., Held Online. 2.1.1

Huang, L., H. Ji, K. Cho, I. Dagan, S. Riedel, and C. Voss (2018), Zero-shot transfer learning for event extraction, in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 2160–2170, Association for Computational Linguistics, Melbourne, Australia, doi:10.18653/v1/P18-1201. 2.3, 6.3.4

IPTC (2022), NewsCodes. 1, 2.5, 2.5.2, 6.1.1

Jacobi, C., W. von Atteeldt, and K. Welbers (2016), QUANTITATIVE ANALYSIS OF LARGE AMOUNTS OF JOURNALISTIC TEXTS USING TOPIC MODELLING, *Digital Journalism*, *4*(1), 89–106, doi:10.1080/21670811.2015.1111968. 2.1.1

Laerd Statistics (2019), Fleiss' kappa using SPSS Statistics. 2.6

Linhares Pontes, E., L. A. Cabrera-Diego, J. G. Moreno, E. Boros, A. Hamdi, N. Sidère, M. Coustaty, and A. Doucet (2020), Entity linking for historical documents: Challenges and solutions, in *Digital Libraries at Times of Massive Societal Transition*, edited by E. Ishita, N. L. S. Pang, and L. Zhou, pp. 215–231, Springer International Publishing. 2.2, 6.2

Liu, J., Y. Chen, K. Liu, W. Bi, and X. Liu (2020), Event extraction as machine reading comprehension, in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1641–1651, Association for Computational Linguistics, Online, doi:10.18653/v1/2020.emnlp-main.128. 2.3, 6.2

Marrero, M., J. Urbano, S. Sánchez-Cuadrado, J. Morato, and J. M. Gómez-Berbís (2013), Named entity recognition: Fallacies, challenges and opportunities, *Computer Standards Interfaces*, *35*(5), 482–489, doi:https://doi.org/10.1016/j.csi.2012.09.004. 2.2, 6.2

Payton, J. D. (2018), How to identify negative actions with positive events., *Australasian Journal of Philosophy*, *96*(1), 87–101. 5.3.1, 6.2

Saggion, Horacio, Funk, Adam, Maynard, Diana, Bontcheva, and Kalina (2007), Ontology-based information extraction for business intelligence, pp. 843–856, doi:10.1007/978-3-540-76298-0_61. 3.3

Seabold, S., and J. Perktold (2010), Statsmodels: Econometric and statistical modeling with python, in *9th Python in Science Conference*. 5.1.1

Sjøvaag, H., and N. Kvalheim (2019), Eventless news: Blindspots in journalism and the 'long tail' of news content., *Journal of Applied Journalism and Media Studies*, *8*(3), 291–310. 5.3.1, 5.3.1, 6.2

Smith, J., L. Johnson, K. Brown, and M. Lee (2021), Limitations of Event Extraction: A Critical Analysis, *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2.3

University of Arizona Global Campus Writing Center (), Journalistic writing. 4.3.2, 6.2

Williamson, K., and G. Johanson (2017), *Research Methods. Information, Systems and Contexts*. 3.1.1, 3.1.1, 3.1.2, 3.2, 4.1.4