Analyzing Bias in a Knowledge-Aware News Recommendation System

Bachelor Thesis

by

Hannah Greven

Degree Course: Informatik B. Sc.

Matriculation Number: [redacted]

Institute of Applied Informatics and Formal Description Methods (AIFB)

KIT Department of Economics and Management

Advisor:	Prof. Dr. Ralf Reussner
Second Advisor:	Prof. Dr. Harald Sack
Supervisor:	Dr. Mehwish Alam
Submitted:	October 15, 2021

Prologue

I want to thank my supervisor, Dr. Mehwish Alam, for her guidance throughout this thesis.

I also thank Prof. Dr. Sack for the acceptance and supervision of this thesis as second advisor.

Thank you to Prof. Dr. Reussner for the advocacy for software ethics in the 2021 summer term SWT II lecture and the supervision of this thesis.

My special thanks to Luca, who, in loving generosity, lent me his PC. Otherwise, this thesis would not have been possible.

Last but not least, I must express my deepest gratitude to my friends and family who have supported me throughout the past years of study. This achievement would not have been possible without them.

Zusammenfassung

Nachrichtenempfehlungssysteme prägen zunehmend den Online-Nachrichtenkonsum und damit das Weltbild vieler Leser. In Nachrichtenempfehlungssystemen werden zunehmend häufiger Deep Learning Modelle eingesetzt. In diesen Modellen können sich unerwünschte Verzerrungen in den Trainingsdaten als Bias im Modell widerspiegeln. In den letzten Jahren gab es einen starken Anstieg der Forschung zum Thema Bias in Empfehlungssystemen. Vorangegangene Forschung untersucht Bias in Wort-, Text- und Wissensgraph-Embeddings. Literatur, die sich mit der Analyse von Bias in Forschungsmodellen zur Nachrichtenempfehlung befasst, ist noch spärlich. Bias-Analysen sind jedoch unerlässlich, um Forschungsmodelle verantwortungsbewusst in praktische Nachrichtenempfehlungssysteme umzusetzen. In dieser Arbeit untersuchen wir Bias im KRED Nachrichtenempfehlungsmodell, einem inhaltsbasierten Nachrichtenempfehlungsmodell, das zusätzliches Hintergrundwissen aus einem Wissensgraphen nutzt und für personalisierte Nachrichtenempfehlungen auf dem MIND-Datensatz trainiert wurde. Zur Untersuchung der Diversität politischer Ansichten in den Empfehlungen erstellen wir einen Korpus mit Nachrichtenartikeln zum politischen divers diskutierten Thema "Migration in der EU". Mithilfe dieses Artikelkorpus, sechs Rezeptionsprofilen und synthetisch generiertem Nutzerverhalten analysieren wir die Diversität empfohlener Artikel. Die Analyse zeigt, dass KRED Artikel aus Nachrichtenquellen, die häufiger in der Lesehistorie des Nutzers vorkommen, bevorzugt empfiehlt. Unabhängig vom Rezeptionsprofil des Nutzers bevorzugt das KRED NRS im Artikelkorpus unfair Nachrichtenartikel von Breitbart (politisch rechte Quelle).

Abstract

News recommendation systems (NRS) are increasingly shaping online news consumption and thus the world perception of many readers. In NRS, state-of-the-art deep learning models are becoming increasingly popular. In these models, unwanted biases in the training data can lead to bias in the model. In recent years, there has been a surge of research on bias in recommendation systems examining bias in word, text, and knowledge graph embeddings. Literature on analyzing biases in research models proposed for news recommendation is still scarce. To translate scientific research models into responsible NRS bias analysis is indispensable. This thesis investigates bias in the KRED NRS, a content-based knowledge-aware NRS that uses a knowledge graph as side information. To analyze KRED's recommendations regarding political bias in exposure, we create a corpus of news articles on the politically divisive topic of migration in the EU. Using this article corpus, six news reception profiles, and synthetically generated user behavior, we analyze the diversity of recommended articles. The analysis shows that the KRED NRS preferably recommends news from news outlets more frequently contained in the user‘s reading history. Independent from the user‘s news reception profile, the KRED NRS unfairly favors news articles from Breitbart (right-wing political bias) in the article corpus.

Contents

List of Tables

1 Introduction

This chapter provides an introduction to the problem field of responsible news recommendation. Motivated by the introductory overview on technical and beyond technical challenges in news recommendation, the research goal of this thesis follows.

1.1 Beyond technical challenges in news recommendation

Recommendation systems are increasingly shaping our online experience through services from numerous providers such as Amazon, Google, Netflix, and Spotify. The trends towards a more digital, mobile, and platform-dominated media landscape, already established in the e-commerce, film, and music industries, are also emerging for news [40]. Publishers of established magazines have been experiencing declines in print revenues for years. The Internet, where anyone can publish news, offers users real-time on-demand access to a variety of news. Established publishing houses are gradually expanding their digital services as well. The result is a broad online media landscape of diverse news blogs and websites. With millions of news items published online each day [52], providers are competing for the users' attention online. This results in a flood of news that can be overwhelming for users. To counteract the flood of information and to keep users engaged with the content, personalization through recommendation systems is increasingly employed on news websites and platforms.

While print media acted as information gatekeepers and embodied the role of the "Fourth Estate" of democracy [8], this role is increasingly shifting to online news outlets, where recommendation systems influence the user's exposure to content. The user's exposure to news shapes the user‘s world perception. With many users, personalized news recommendation thus also has an influence on society at large. As the Reuters News Report 2020 has found, especially younger generations (age under 35) broadly use online news as a source of information [39].

From the prior description of the news recommendation scenario, four stakeholders derive:

- News provider: Anyone publishing news of any kind (see Figure 1, top left).
- News consumers: Users of news outlets and employed News Recommendation Systems (NRS) (see Figure 1, top right).
- NRS developers: Developers that develop, operate and maintain the NRS (see Figure 1, top middle).
- Society: An aggregate of people typically sharing spacial territory, cultural values and political authority (see Figure 1, bottom middle).

Figure 1 provides a summarizing overview of the stakeholders and their interaction in the news recommendation scenario. In the following, the focus is on the NRS and challenges for its developers.

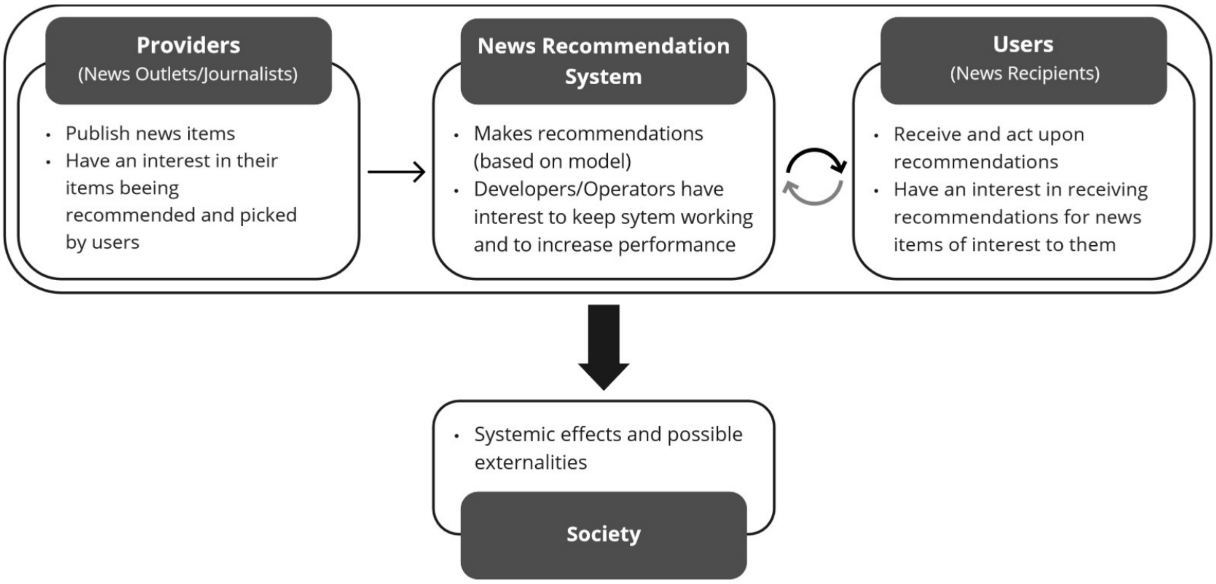


Figure 1: Multi-stakeholder model for recommendation systems from [37] adapted for the scenario of news recommendation.

For developers of news recommendation systems, challenges arise of both technical and beyond technical nature. First, it is necessary to develop a model that accurately predicts the user's interests. Artificial neural networks allow learning patterns from news-user interaction logs. Technically challenging is the solution to the so-called cold start problem. Without known previous user interactions with news articles, such models can not predict the user's interests. In addition, there are challenges in terms of scalability, the large and rapidly changing amount of current news, as well as the detection of short- and long-term user interests. Beyond these technical challenges, privacy is another issue that developers of news recommendation systems must consider. The data processed is sensitive personal data according to the European General Data Protection Regulation (GDPR) [43]. Further restrictions specifically for platforms employing recommendation systems are proposed in the European Digital Service Act (DSA) [11]. In addition to privacy concerns, lawmakers and researchers are also concerned about the potential impact of personalized news recommendations in terms of selective exposure and potential polarization, which are critical to consider for society at large. There is controversy on whether recommendation systems contribute to so-called filter bubbles and echo chambers [38, 64, 42]. Clearly, with the control over news recommendation algorithms comes great social power and responsibility [22]. After all, data-driven recommendations can have potentially invasive, manipulative effects on user privacy, autonomy, and self-determination [16].

These effects do not even have to be intentional. Recommendation algorithms can also unintentionally influence users through biased exposure, as happened, for example, in the case of Youtube's recommendation algorithm [46]. Over time, the recommended content became more sensational and extreme as it kept users on the platform. This ultimately lead to an increased reach for content with extreme political opinions.

Such effects often go unnoticed because recommendation models are primarily trained and evaluated on predictive accuracy. In recent years, the realization that other recommendation qualities can also have a significant impact on the overall quality of a recommendation system has shifted the focus of recommendation system research to a set of goals beyond accuracy [27]. Nevertheless, accuracy is still the dominant metric for evaluating and comparing recommendation systems.

In the case of Youtube's recommendation algorithm, a pre-deployment analysis of content exposure regarding politically biased content would have been beneficial to evaluate the social responsibility of the recommendation algorithm. For news recommendation algorithms in general, such a pre-deployment analysis would be beneficial, as public awareness of media bias is still limited [15] and those who are aware of media bias prefer news that reflects a range of views and let them decide for themselves [40]. Therefore, it is important to not only compare the accuracy of news recommendation models but also to analyze news recommendation systems in terms of exposure, diversity, and the prominence of politically biased content to assess system responsibility.

1.2 Research goal

This thesis intends to contribute to system responsibility assessment of proposed scientific research models for news recommendation. The goal is to analyze a personalized state-of-the-art knowledge-aware NRS regarding political bias in exposure.

(I) Establish an article corpus of articles on the divisive topic of "Migration in the European Union (EU)" that covers different political opinions.

(II) Analyze the systemic effects regarding political bias and exposure diversity in the system's recommendations.

(III) Answer the research question: Does news recommendation with the chosen knowledge-aware NRS amplify political bias from previously read articles?

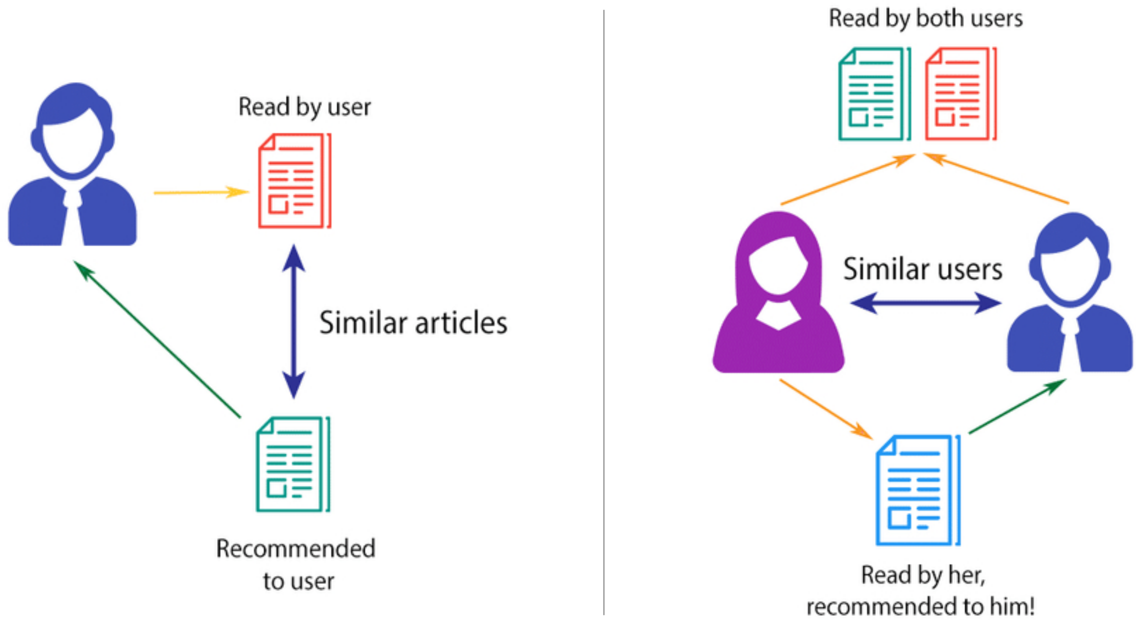
(IV) Discuss possible explanations for the observed systemic effects in form of an exploratory literature review.

2 Background

To give an overview on the subject of knowledge-aware NRS, the sections in this chapter introduce types of recommendation systems, key applications for news recommendation, different forms of article representations along the concept of named entities, and deep learning for news recommendation.

2.1 Types of news recommendation systems

At the core of each recommendation system is the recommendation algorithm. For its development, user and item content and their underlying interactions are essential inputs. Types of recommendation algorithms traditionally used in recommendation systems can be classified as content-based filtering (CBF), collaborative filtering (CF), and hybrid approaches [45]. In CBF algorithms, the algorithm recommends articles by comparing the user profile and the item profile based on the content in a shared attribute space. In contrast, CF algorithms do not use the news item content. Here, the algorithm recommends news items based on the ratings or interactions of users with similar user profiles.



(a) Content-based filtering (b) Collaborative Filtering

Figure 2: Collaborative and content-based filtering as published in [23].

Hybrid approaches combine both CBF and CF approaches to benefit from the best properties of both approaches in solving the recommendation task. In the following chapters, this thesis focuses on content-based NRS. The decisive factor for this decision is the direct influence of content features in content-based NRS.

2.2 News recommendation applications

The types of NRS described in section 2.1 apply personalized news recommendation in the form of user-to-item recommendation. This thesis focuses on content-based user-to-item recommendation. Approaches other than user-to-item recommendation are item-to-item recommendation, popularity-based recommendation, topic-based recommendation, and context-based recommendation (e.g. location-aware and time-aware news recommendations). An industrial NRS has to combine various of those key applications to mitigate challenges specific to the news recommendation domain like the cold start problem.

2.3 Article representation and named entities

Content-based user-to-item recommendation is based on news article content, which naturally consists of unstructured text. Text form is highly inefficient for automatic processing of a vast and rapidly changing news catalog, which is typical for the news domain. To this end, news recommendation systems build on Natural Language Processing (NLP) methods. For effective processing, text embedding methods must first translate news articles from text form into high-dimensional vector spaces. The resulting vector representations are called embeddings.

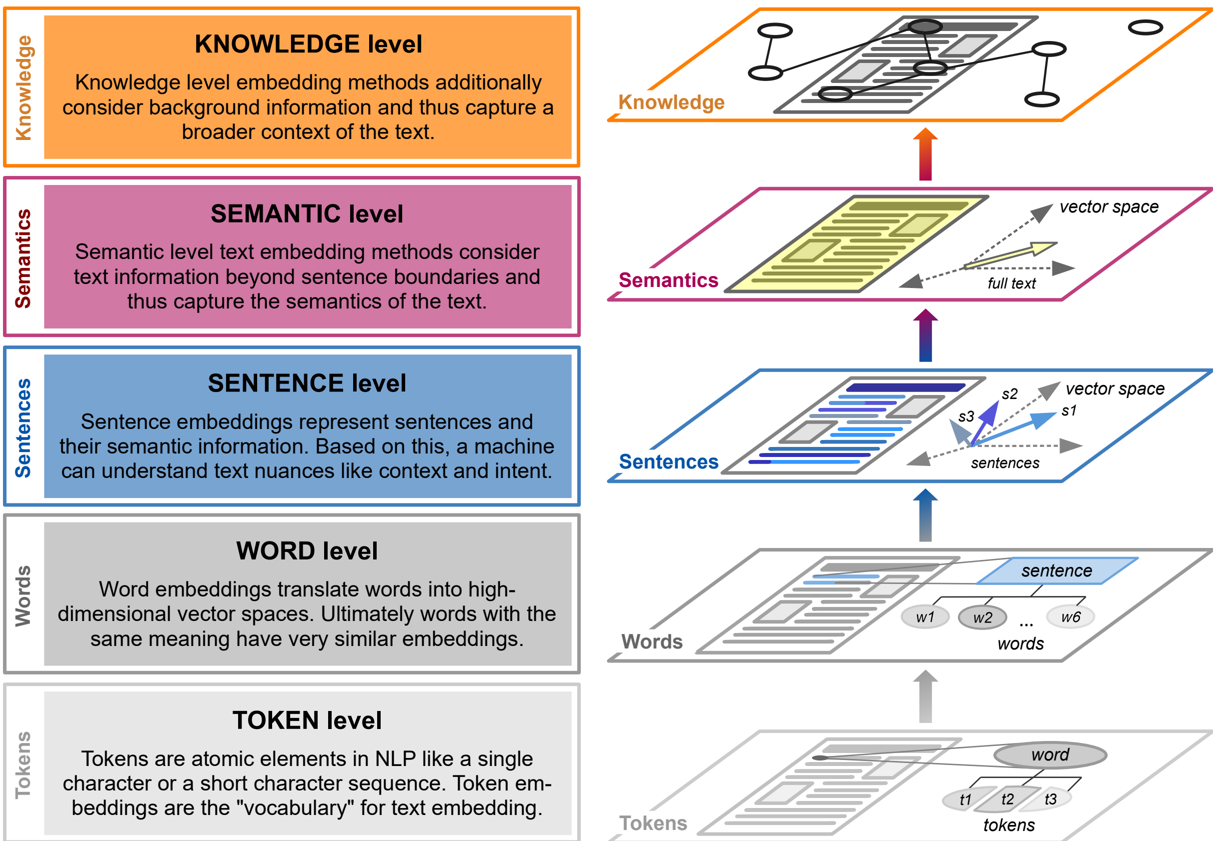


Figure 3: Types of text embeddings organized by level of semantics they capture.

Figure 3 shows the different types of text embeddings. The three base levels comprise embedding methods for lower-level text elements such as tokens, words, and sentences. Embeddings of these elements are not yet capable to capture complex text semantics. But with increasing level, the semantics contained in the embeddings increase. Calculating text embeddings based on text information that extends beyond sentence boundaries, but does not contain any information other than the text, yields semantic-level text representations. Semantic-level text representations capture complex text semantics but do not include contextual information. To capture superordinate text meaning which goes beyond pure text semantics requires additional knowledge in the embedding process. Adding background knowledge ultimately leads to knowledge-level text representations.

Background knowledge is beneficial for text comprehension - for humans and NLP alike. News articles typically contain named entities such as important people, places, or organizations. These entities convey the core message of any news article. Accordingly, background knowledge about those entities naturally helps in understanding the content.

All entities and their relations, real-world and fictional, can be modeled in a knowledge graph. Entities can be conceptualized as vertices and the relationships as edges between them. Taking this notion further, a knowledge graph can be formally defined as follows.

Definition 1. A knowledge graph is a collection of a entity-relation-entity triples $\mathcal{G} = \{(h, r, t) \mid h, t \in \mathcal{E}, r \in \mathcal{R}\}$, where $\mathcal{E}$ and $\mathcal{R}$ represent the set of entities and relations respectively and (h, r, t) indicates a relation r from head entity h to tail entity t.

The inherent connections between entities in the text and neighboring entities in the knowledge graph can help understand the context of a news article.

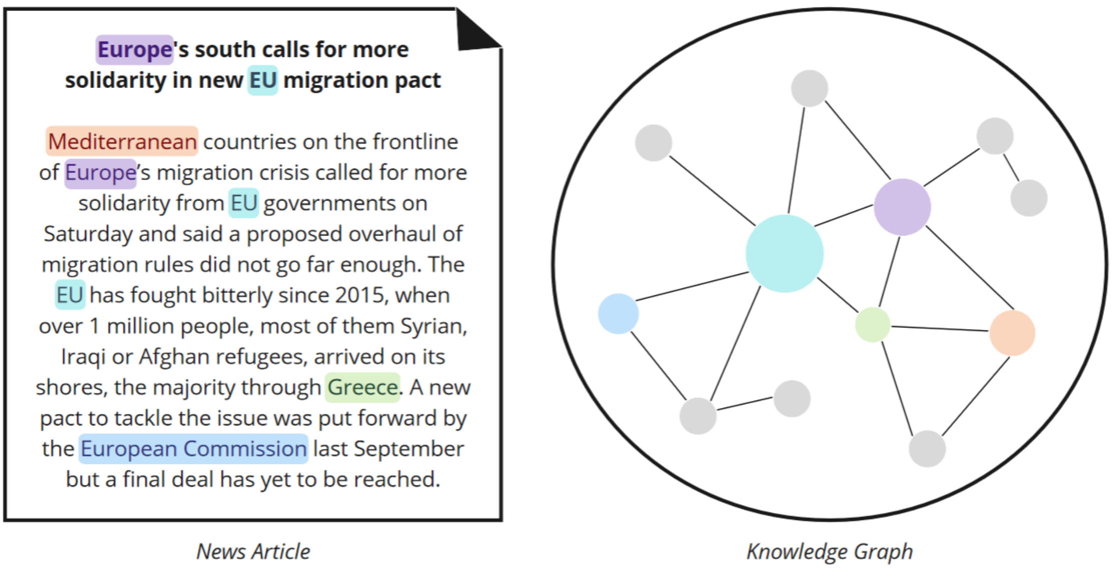


Figure 4: Named entities in a news article[1] and their association with a knowledge graph.

[1] `www.reuters.com/article/idUSKBN2BC0JY` (Last accessed: 2021-08-03)

Figure 4 visualizes the associations between mentioned named entities in a news article and the entities in the knowledge graph by color. Gray vertices represent neighboring entities not mentioned in the text.

Incorporating additional knowledge from a knowledge graph is an advancing field of study in recommendation systems. Those approaches make use of NLP methods. Named Entity Recognition (NER) methods provide the ability to extract named entities from the text representation of the news articles. Named Entity Linking (NEL) methods associate recognized entities with the corresponding entities in a knowledge graph. Same as for text, entities and their relations can be embedded in high-dimensional vector spaces. It is possible to combine entity, relationship, and text vectors employing linear algebra. Including embeddings of neighboring entities from the knowledge graph yields knowledge-level representations of the news article, which allow the recommendation system to leverage the additional knowledge. Recommendation systems that incorporate information from a knowledge graph are called knowledge-aware recommendation systems.

2.4 Deep learning for news recommendation

Recently, modern Deep Learning (DL) approaches have become more and more popular. Deep Learning refers to Machine Learning (ML) algorithms in the field of Artificial Intelligence (AI) that are typically based on Deep Neural Networks (DNNs). Figure 5 visualizes the relation of AI, ML, DL, and DNNs.

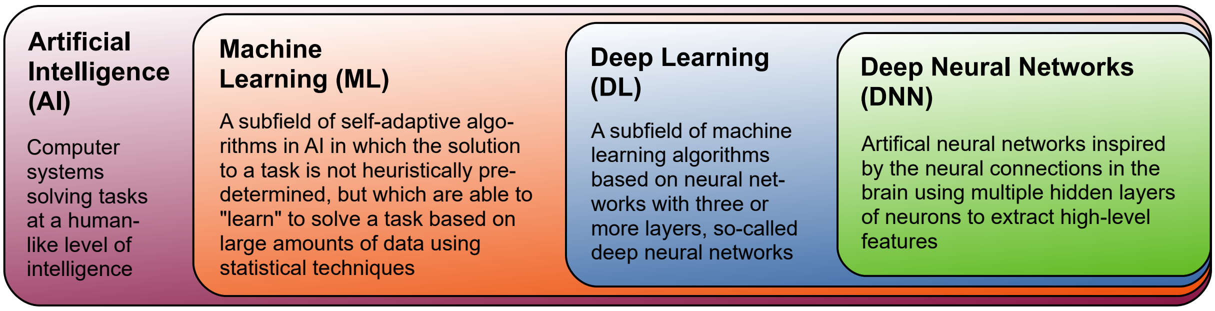


Figure 5: Relation of AI, ML, DL, and DNNs.

DNNs are artificial Neural Networks (NNs) inspired by the neural structures in the human brain. DNNs use at least three layers of neurons to progressively extract higher-level features from the raw input. For example, in text processing, lower layers may identify tokens, while higher layers may identify the semantic concepts relevant to a human such as sentence or text semantics.

DNNs consist of at least three layers: an input layer, an output layer, and at least one, typically multiple, hidden layer(s) between the input and output layer. Each layer comprises a set of nodes, so-called neurons. Depending on the task of the neural network, the neurons of successive layers are connected differently. Connections between neurons are called edges and resemble synapses in the brain.

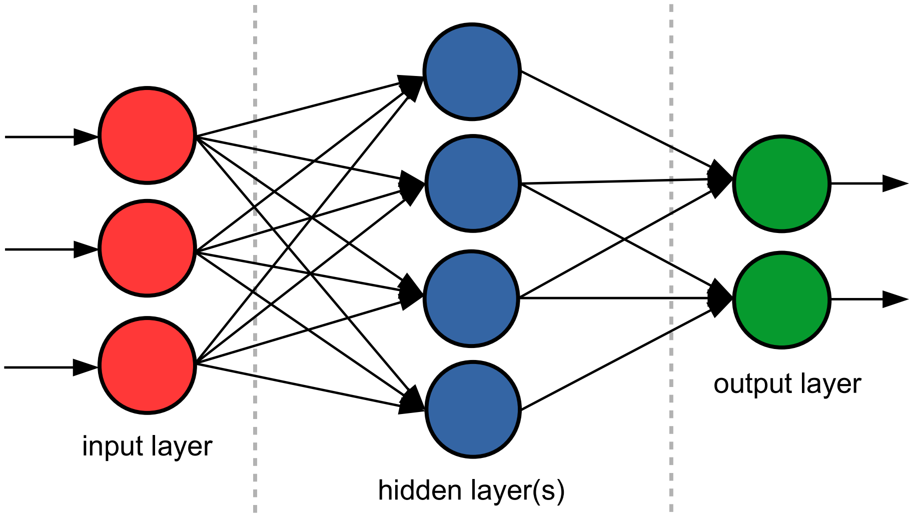


Figure 6: Schematic sketch of a simple deep neural network.

Figure 7 shows the mathematical model of neural activity that constitutes the basis of DNNs. Both neurons and edges have associated weights, which are learned from iterative adaptation in the training process. Neurons represent mathematical, non-linear functions. Each neuron takes its inputs, multiplies them by the associated edge weights, and sums them up. The resulting weighted sum of inputs is processed with the neuron's activation function. The resulting value is passed to neurons in the next layer afterward.

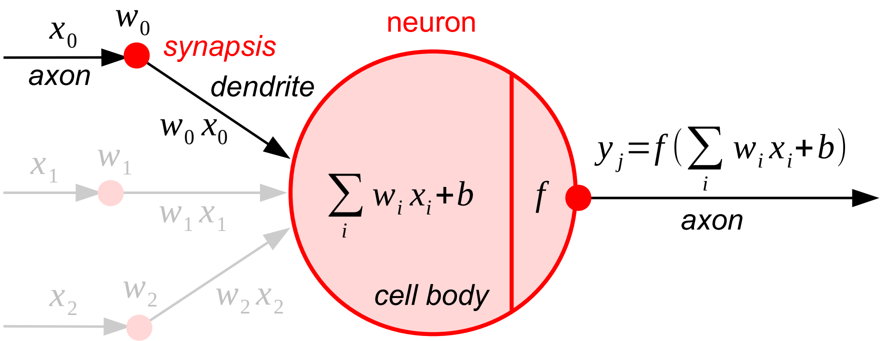


Figure 7: Mathematical model of neural activity adapted from [32], whereas x_0, x_1, x_2 are input features, w_0, w_1, w_2 and b are weights, f is the activation function, and y_i denotes the neuron's output.

By training on large amounts of data, the neural network learns its target function. News recommendation models learn functions that predict the users preference for future news items. With the notation suggested by Guo et al. [21], deep learning approaches for content-based news recommendation can formally be described as follows:

1. The recommendation system calculates representation u_i and v_j for the target user i and the candidate news item j.
2. The system learns a scoring function $f\colon u_i \times v_j \rightarrow \hat{y}_{ij} \in [0,1]$ that models the user's preference for a news item.
3. With this function, the system estimates the user's preference for each news item.
4. The system generates the recommendation by sorting the news items according to the preference scores calculated.

3 State of the art

This chapter briefly outlines several state-of-the-art NRS models for content-based knowledge-aware news recommendation. It also provides an overview on bias in content-based, knowledge-aware NRS, and introduces the NRS evaluation metrics relevant to this thesis.

3.1 Content-based knowledge-aware news recommendation

In recent years, generating recommendations with a knowledge graph as side information has attracted great interest. As explained in section 2.3 article representation is essential in news recommendation. Knowledge-aware recommendation systems leverage additional information about the relations between entities in the text and their context provided by a knowledge graph. The added knowledge helps to improve the accuracy and diversity of personalized recommendations and additionally provides rationale for explainability [21].

Several knowledge graphs have been proposed in the last decade, like Google's Knowledge Graph [50] and Microsoft's Satori [18] but also public knowledge graphs, such as DBpedia [31], WikiData [54], and YAGO [51].

Wang et al. [56] were the first to use knowledge graph embeddings in news recommendation. They propose a deep knowledge-aware network (DKN) built on a knowledge-aware Convolutional Neural Network (CNN) to fuse additional knowledge with semantic level article representations for knowledge level news representations. An attention mechanism models the dynamic change in user preferences. Wang et al. propose RippleNet, a framework for user-preference propagation over a knowledge graph based on the idea of ripples created by raindrops hitting the water [55]. Ripples are activated from entities in articles the user clicked and include neighbors of mentioned entities in ripple sets. Inverting this concept, they also propose a mechanism to compute vector embeddings for entities containing information on the neighboring entities.

To bridge the gap from entity representations including the entities neighborhood to news article representations including neighboring entity information, Liu et al. propose Knowledge aware Representation Enhancement for Documents (KRED) [34]. In contrast to DKN, which relies on Kim-CNN, a specific type of text embedding model, KRED is built to enhance arbitrary document representations with neighboring entity information. To share information across different news recommendation applications, KRED can be trained in a multi-task framework. Building multiple news recommendation applications on a shared model increases accuracy and efficiency at the same time. The ability to enhance arbitrary document representations allows KRED to take full advantage of pre-trained models available in the age of the "pre-trained and finetuning" paradigm.

Until a few years ago, training neural networks was only possible for research institutes and technical organizations, as training is a very time- and resource-consuming task depending on the data and the target task. Nowadays parallel computation capabilities of modern graphic processing units are more powerful and pre-trained models are available even for language-related tasks. With KRED any pre-trained text embedding model can be used as initial document representations, for example Bidirectional Encoder Representations from Transformers (BERT) [14], a model for text embedding that yields state-of-the-art results in a wide variety of NLP tasks. To our knowledge, KRED was evaluated on accuracy metrics only and has not been studied regarding bias yet.

3.2 Bias in recommendation systems

Most of the research on recommendation systems focuses on the development of machine learning models that provide more accurate recommendations based on user behavior data, including [56, 55, 34]. Since 2016, the issue of bias in recommendation systems has received more scientific attention [9]. Major conferences such as SIGIR and RecSys dedicated sessions to bias in recommendation systems in 2020 [12, 41].

First, it is necessary to explain where biases in recommendation systems arise. Recommendation systems are trained on user behavior data, which is observational. As a consequence, there may be various biases in the data, stemming from population and user preference imbalances for instance. A recommendation system learns to derive recommendations from this user behavior data as explained in 2.1. Thus, bias in the user behavior data also introduces bias to the recommendation system. In the interaction with the feedback loop of a recommendation system, systemic biases can lead to bias amplification (see Figure 8). Bias relevant to content-based knowledge-aware NRS is explained below.

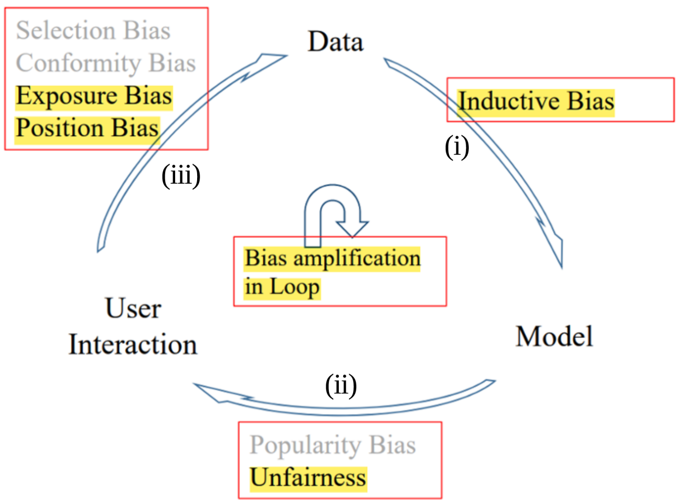


Figure 8: Bias in the feedback loop of a recommendation system adopted from [9]. Bias relevant for content-based knowledge-aware NRS is highlighted in yellow.

User behavior data is collected from implicit click feedback, assuming the user's interest in clicked items. As a recommendation system aims to reduce information overload, it exposes users to only a part of all items. No interaction with an item does not necessarily represent negative preference, which is referred to as exposure bias [9] (see Figure 8 (iii)). Furthermore, studies have shown that users are more likely to click on items positioned higher in the recommendation list regardless of the item's actual relevance to the user [1]. The bias introduced from this phenomenon is called position bias (see Figure 8 (iii)). A model makes certain assumptions to learn the target function and to generalize beyond training data. This bias is called inductive bias, a bias deliberately induced and therefore desirable [9] (see Figure 8 (i)). Another type of bias that has gained attention over the last few years is unfairness (see Figure 8, (ii)), prejudice or favoritism towards an individual or a group in the recommendation results.

Ideally, to study the effect of the feedback loop on bias in recommendation systems, online testing is conducted on a real platform with a steady stream of data. Lacking access to real platforms for experimentation, Mansoury et al. propose offline experiments to simulate and analyze the recommendation system process in [36]. Over 20 iterations, they simulate the user's interaction with the recommendation system. In each iteration they add selected news items from the recommendation list to the user's behavior profile.

3.3 Metrics for NRS evaluation

Metrics for NRS performance evaluation approach from either the classification or ranking point of view. An NRS firstly classifies news as relevant or not relevant to the user and secondly ranks news classified as relevant to the user according to their relevance.

For classification tasks, the confusion matrix in figure 9 is a tool to describe classification performance. It captures how often the model classified an item as positive or negative (relevant or not relevant respectively), correctly or incorrectly.

		Actual Label	
		Positive	Negative
Classification	Positive	True positive (TP)	False positive (FP)
	Negative	False negative (FN)	True negative (TN)

Figure 9: Confusion matrix for classification results.

Many NRS evaluation metrics build on $|TP|$, $|TN|$, $|FP|$, and $|FN|$. Accuracy can be defined as

$$ACC = \frac{|TP| + |TN|}{|TP| + |TN| + |FP| + |FN|}. \tag{3.1}$$

Further metrics derived from the confusion matrix are the Positive Predictive Value (also known as precision) and the True Positive Rate (TPR) (also known as recall or sensitivity).

$$PPV = \frac{|TP|}{|TP| + |FP|}, \tag{3.2}$$

$$TPR = \frac{|TP|}{|TP| + |FN|}. \tag{3.3}$$

True Negative Rate (TNR) and False Positive Rate (FPR) are measures for the system specificity closely related.

$$TNR = \frac{|TN|}{|TN| + |FP|}, \tag{3.4}$$

$$FPR = \frac{|FP|}{|FP| + |TN|},\ FPR = 1 - TNR. \tag{3.5}$$

With TPR on the y-axis plotted over FPR on the x-axis, the Receiver Operating Characteristic (ROC) is a graph that shows the model's classification performance for all classification thresholds. The associated Area Under the ROC Curve (AUC) is a metric preferable to the accuracy measure defined in 3.1 as it is independent of the classification threshold, the mean error decreases with increasing test-sample size and it is invariant to the a priori class probabilities [7, 24].

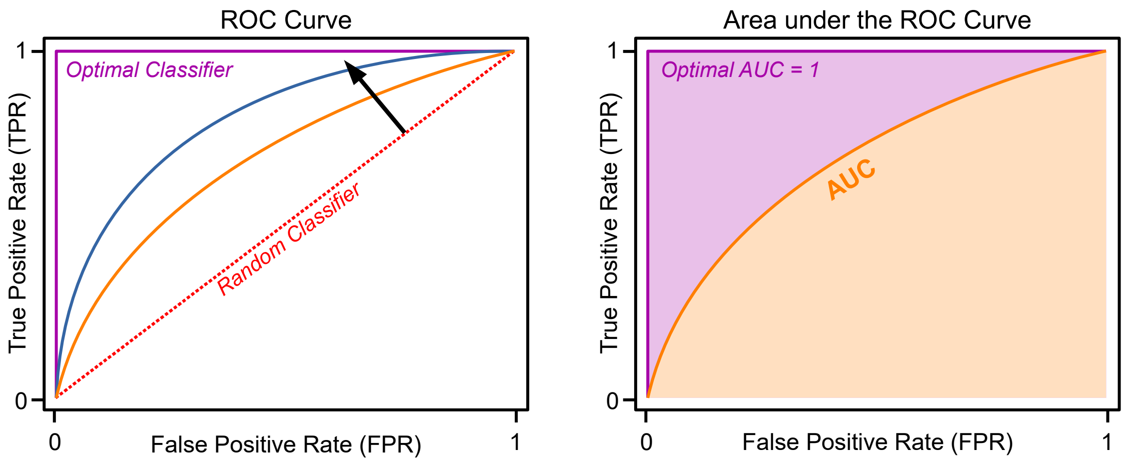


Figure 10: Characteristics of ROC curve and AUC.

With the strengths of the AUC score for evaluating classification performance outlined, we now shift the focus from classification to a metric for ranking performance.

A metric commonly used to measure the performance of an NRS in ranking is its Normalized Discounted Cumulative Gain (NDCG) [26]. The cumulative gain takes into account that highly relevant documents are more valuable than marginally relevant documents. A discount function accounts for greater rankings being less valuable to the user because they appear lower in the list and the user is therefore being less likely to investigate these items. Normalization relativizes the score so that the ideal score is represented by the value 1. Adopting the notation from [57] the NDCG can be formally defined as follows.

For a ranking function f on a dataset S_n of size n, the relevance $y^f_{(r)}$ associated with the item ranked at position r by f, standard discount function $D(r) = \frac{1}{log(1+r)}$ and ideal score $IDCG_D(S_n) = max_{f'} \sum_{r=1}^{n} y^{f'}_{(r)} D(r)$ the NDCG score:

$$NDCG_D(f, S_n) = \frac{DCG_D(f, S_n)}{IDCG_D(S_n)}, \; DCG_D(f, S_n) = \sum_{r=1}^{n} y^f_{(r)} D(r) \tag{3.6}$$

More recent scientific work suggests that a variety of other metrics beyond accuracy exists that should be considered when evaluating recommendation systems [20]. Figure 11 shows a compact overview of recommendation objectives beyond accuracy provided in "A Tutorial on Advances in Bias-aware Recommendation on the Web" at WSDM 2021 [4]. The tutorial's topic shows that metrics beyond accuracy are directly related to bias as they measure system objectives strongly influenced by bias.

Figure 11: Objectives in recommendation systems beyond accuracy adopted from [4].

Kaminkas et al. review definitions of the most popular beyond-accuracy metrics diversity, serendipity, novelty, and coverage in [27]. Silveira et al. collect and define further metrics beyond accuracy and mathematical evaluation functions in [49]. There is some literature in NRS on how to improve individual system goals of the above [44, 58, 61]. Nevertheless, most of the newly developed scientific approaches to news recommendation are still evaluated primarily on their accuracy.

In summary, article representation is essential to any NRS. Different works on content-based, knowledge-aware NRS mostly build on a specific article representation model. In contrast, KRED proposes a solution to enhance arbitrary document representations with additional knowledge from a knowledge graph.

Bias is a factor not to neglect in NRS. Several state-of-the-art NRS exist, which are built to mitigate a specific bias [44, 58, 61]. However, many proposals do not consider bias explicitly so far and are still evaluated and compared primarily on metrics regarding predictive accuracy. Metrics beyond accuracy offer comparability of system objectives strongly influenced by bias. As long as those are not used in the design, evaluation, and comparison of a NRS besides accuracy metrics, a research gap remains in analyzing NRS evaluated primarily on accuracy regarding the impact of systemic bias. This step is indispensable to translate research models into practical news recommendation systems. In this thesis, we want to address this shortcoming in analyzing a state-of-the-art knowledge-aware NRS regarding systemic political bias in exposure.

4 Methodology

This chapter explains the methods used for analyzing the behavior of a state-of-the-art knowledge-aware NRS regarding political bias inherent in the user's reading history and news article data. We choose to analyze KRED [34], because it is one of the state-of-the-art algorithms based on knowledge-graphs. Figure 12 provides a schematic overview.

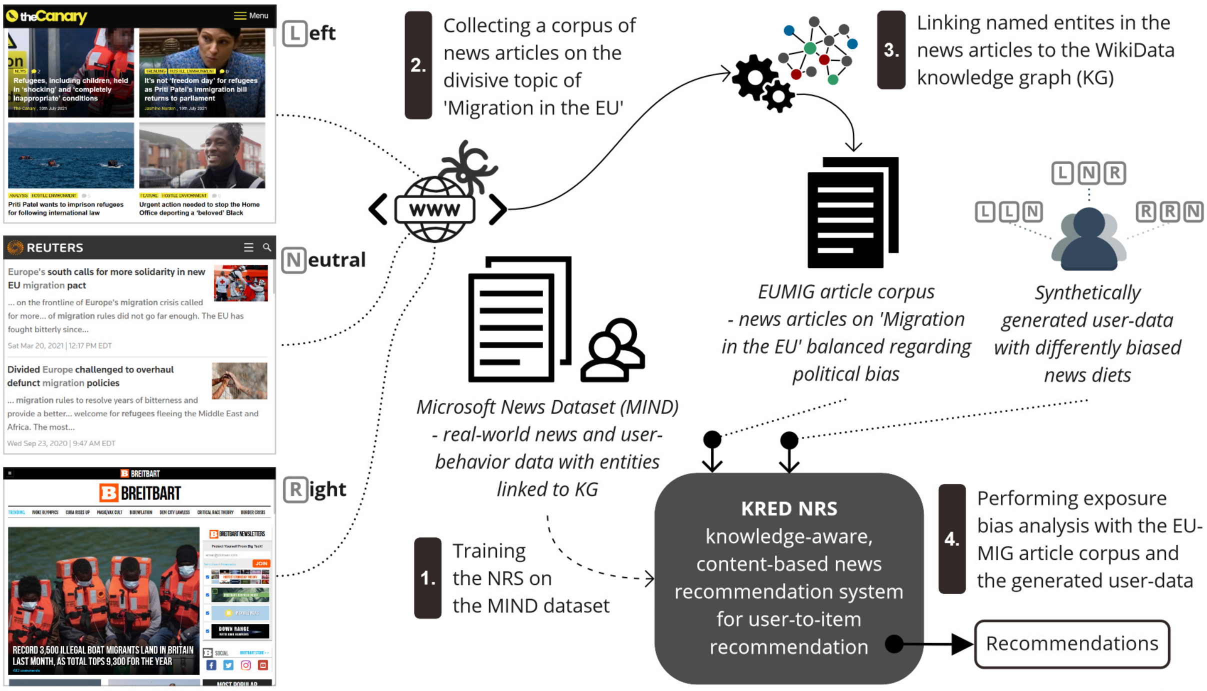


Figure 12: Schematic overview of the methodology employed in this thesis.

The methodology in this thesis comprises four steps: model training, data collection, entity linking, and bias analysis. In ***model training***, we train KRED for personal news recommendation on the Microsoft News Dataset (MIND) [59], a observational real-world dataset established for news recommendation research. In ***data collection*** the challenge lies in establishing a corpus of news articles in the format of the MIND. This corpus has to contain political bias as we want to analyze political bias in exposure. Therefore, we collect a corpus of news articles from news websites with different political opinions on the divisive topic of "Migration in the EU", in the following called the EUMIG article corpus. Afterwards, ***entity linking*** prepares the EUMIG article corpus for use in a knowledge-aware setting. To that end, mentioned entities in the news articles need to be recognized and linked to a knowledge graph. For the ***bias analysis***, we define six exemplary news reception profiles, that differ in the user's news diet regarding exposure to political bias. Lacking access to real-world user data, we synthesize exemplary user-behavior data from those profiles and news articles in the EUMIG article corpus. Finally, using the synthetic user-behavior data and the EUMIG article corpus, the KRED NRS is analyzed regarding political bias in exposure. The following sections explain each step in more detail.

4.1 Model training

We choose to analyze the knowledge-aware state-of-the-art news recommendation model KRED, because of its merits mentioned in section 3.1. As a first step, training is necessary as KRED is not available pre-trained. The model consists of two parts. Three layers fusing semantic-level embeddings of the news articles with information from the knowledge graph build the KRED core. On top, further layers implement different NRS applications.

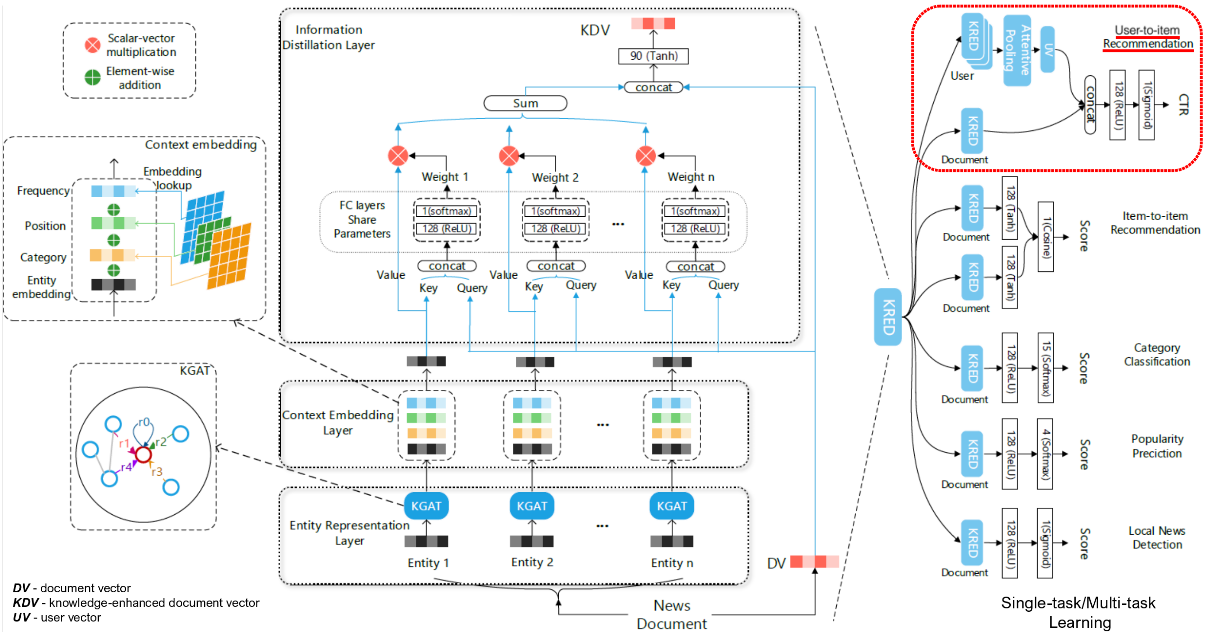


Figure 13: Overview of the KRED model adopted from [34]. In this thesis, the focus is on user-to-item recommendation, which is marked red.

The core model consists of entity representation, context embedding, and information distillation layer. These layers enhance the news article embeddings with context. KRED uses a knowledge graph, which provides context for entities mentioned in the text from neighboring entities. The entity representation layer fuses entity embeddings from mentioned entities with the entity embeddings from their 1-hop neighborhood. Weights determine the amount of information a neighboring entity propagates in the process.

Contextual information such as an entity's position and frequency in the text along with the entity's category indicates the importance of an entity in the news article. If an entity appears in the news title and several times throughout the news body, it is usually more important than an entity mentioned only once in the text. Knowing whether an entity is a person, a place, or an organization increases the accuracy of the model. For each entity, the context embedding layer encodes position, frequency, and category information as vectors. Those are summed up with the entity embedding calculated in the entity embedding layer.

For news recommendations, the single entity representations need to be merged into the article representation vector. Besides the meaning of an entity, co-occurring entities and the article topic strongly determine an entity's importance in the article context. Therefore, the information distillation layer merges all entity embeddings using weights to model the importance. The resulting entity information vector is concatenated with the original article embedding. As original article embeddings we use BERT [14] embeddings in this thesis.

In the training process, the KRED model learns parameters for the weighting functions mentioned above and parameters for the NRS task-specific layers. Interested in the systemic effects of personalized news recommendations, we train the NRS on the task of user-to-item recommendation. As training data, we use the MIND dataset [59] as suggested in [34]. MIND provides user behavior in the form of click histories and impression logs, news article data including information on mentioned entities, and embeddings of entities and relations for a subgraph of WikiData learned with TransE [5]. TransE is an embedding method for multi-relational data such as knowledge graphs.

Training on MIND-large, we obtain a model trained on real-world article and user data that allows us to evaluate how bias inherent in the user's read history influences recommendations with KRED. Due to licensing restrictions article bodies are excluded from the MIND dataset. As MIND-large contains about 160,000 news articles, scraping article bodies and linking mentioned entities would exceed the scope of this bachelor thesis, the model is trained on the provided title and abstract information of the articles.

We have found that at least 14 GB of RAM is required for training. Training 5 epochs with a batch size of 64 using GPU NVIDIA GeForce RTX 2070 SUPER and CPU AMD Ryzen 7 3700X 8x 3.6 GHz took 15 days and resulted in an AUC score of 0.676661171922004 and a NDCG of 0.40928869967161396. Table 1 compares the performance of the original KRED-model [34] and the working model trained for this thesis in the user-to-item recommendation task on MIND-large. Henceforth the KRED working model trained in this step will be referred to as the KRED NRS.

Model	**AUC**	**NDCG@10**
KRED(original)	0.6910	0.2684
KRED(working model)	0.6767	0.4093

Table 1: AUC and NDGC@10 performance of original and working KRED model.

4.2 Data collection

To test the KRED NRS behavior regarding political bias in exposure, we collected an article corpus of real-world news articles on the divisive topic of “Migration in the EU“ containing articles from outlets known for their political bias. Assuming the outlet's political bias as political bias for its articles is common practice in bias and factuality detection [2]. We intend to adopt this approach in this thesis. In general, the political bias of articles from one outlet does not necessarily have to be consistent. Thus, for this thesis news outlets with strong and consistent political bias are selected.

4.2.1 Outlet selection

As news outlets, we chose The Canary[2], a left-wing British news website, Reuters[3], an international news agency committed to unbiased reporting[4], and Breitbart[5], a right-wing American news website with a London/Europe branch.

Figure 14: Outlets in the article corpus and their MBFC bias classification.

The considerations leading to this choice are threefold:

1. We chose outlets publishing English articles because there is more research on news recommendation systems for the English language.
2. All of those outlets publish to a readership in the United Kingdom, where the coverage of the European refugee crisis was the most negative and the most polarised according to a content analysis of the European press reporting on the 2014-15 refugee crisis [3].
3. Those outlets cover the spectrum of political bias left, least-biased, and right according to the Media Bias/Fact Check (MBFC)[6], a website that determines the bias of news outlets through a combination of objective measures and subjective analysis.

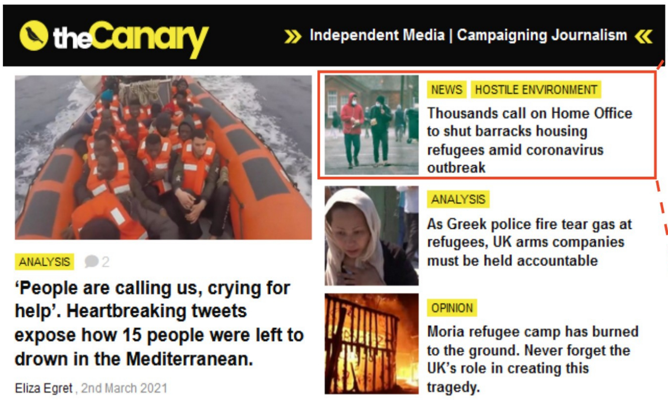
4.2.2 Article scraping

On the websites of the news outlets, the article text is embedded in HyperText Markup Language (HTML) files. To work with the articles in a NRS, the article texts must be extracted from the HTML files. For this purpose, crawlers are used that follow the URLs on the website and pass relevant URLs to a scraper, which extracts the article content.

[2] `www.thecanary.co.uk` (Last accessed: 2021-08-03)
[3] `www.reuters.com` (Last accessed: 2021-08-03)
[4] `www.thomsonreuters.com/en/about-us/trust-principles.html` (Last accessed: 2021-08-18)
[5] `www.breitbart.com` (Last accessed: 2021-08-03)
[6] `www.mediabiasfactcheck.com` (Last accessed: 2021-08-12)

Articles from Reuters were scraped via a crawl by search approach querying for the keywords Europe, migration, and refugee ("Europe" AND "migration" AND "refugee") to scrape the results. For Breitbart, a topic-related news feed exists[7], which was used for article retrieval from this outlet. For The Canary, both querying for the keywords Europe, migration, and refugee ("Europe" AND "migration" AND "refugee") and related topic pages[8,9,10] were used to retrieve articles on migration in the EU.

Figure 15 shows a typical news feed on the website of a news outlet taking The Canary as an example. Since the HTML parse tree is different for each news website, to extract the article text from the HTML file on the news outlet's website, a scraper was written for each news outlet using the Python libraries Requests[11], Newspaper3k[12], and BeautifulSoup[13].



Title	Thousands call on Home Office to shut barracks housing refugees amid coronavirus outbreak
Category	News
Abstract	120 people have reportedly tested positive for coronavirus (Covid-19) at a military barracks housing refugees in Kent. The outbreak has sparked fears of a health crisis.
Body	Around 400 people are living at Napier Barracks in Folkestone, Kent, which means that over 1 in 4 residents could be infected. The virus outbreak has added to concerns over conditions at the barracks...

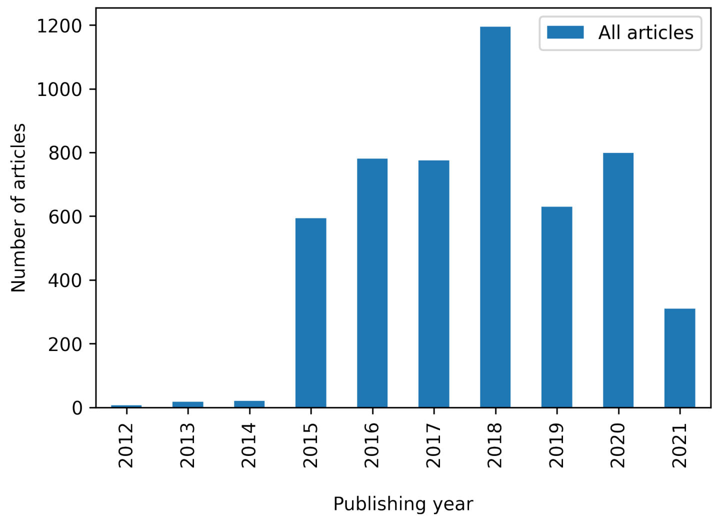
Figure 15: Example article feed for The Canary and content of an example news article.

The scrapers extract the following information from the HTML file of each article: Title, category, abstract, body, and page-URL. Also, the publish date is extracted as meta-information for statistics on the collected article corpus. In the following, we will refer to the article corpus collected in this step as the EUMIG article corpus.

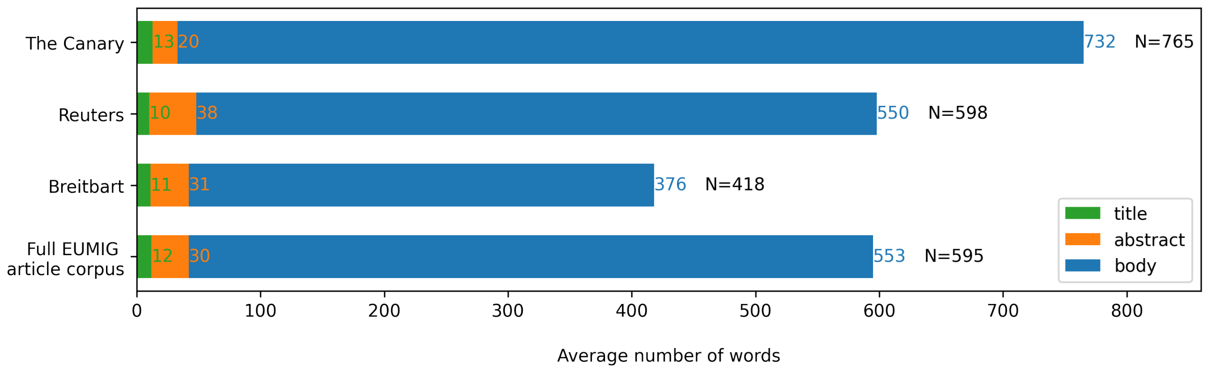
4.2.3 Statistics on the EUMIG article corpus

For each outlet, 1709 news articles were scraped, leading to a bias-balanced article corpus of 5127 articles. The publish dates of the articles range from 2012 to 2021. The average word count per article (title, abstract, and body) in the EUMIG article corpus is 595. Figure 16 shows statistics on the EUMIG article corpus in more detail.

[7] `www.breitbart.com/tag/europe-migrant-crisis` (Last accessed: 2021-08-12)
[8] `www.thecanary.co/topics/hostile-environment/` (Last accessed: 2021-08-12)
[9] `www.thecanary.co/topics/deportations/` (Last accessed: 2021-08-12)
[10] `www.thecanary.co/topics/calais/` (Last accessed: 2021-08-12)
[11] `www.github.com/psf/requests` (Last accessed: 2021-08-18)
[12] `www.github.com/codelucas/newspaper` (Last accessed: 2021-08-18)
[13] `www.crummy.com/software/BeautifulSoup/` (Last access: 2021-08-18)



(a) Time distribution.



(b) Average word count in the articles, denoted by N, for each of the news outlets and the whole EUMIG article corpus subdivided in average word counts for title, abstract, and body.

Figure 16: Statistics on the EUMIG article corpus.

4.3 Data preparation

To analyze KRED with the EUMIG article corpus, the EUMIG article corpus must replicate the data structure of the MIND dataset on which KRED was trained. To this end, the articles in the EUMIG article corpus are supplemented with information about the entities they contain by the means of named entity recognition and linking.

4.3.1 Named entity recognition and linking

KRED needs information about entities mentioned in the news articles to enhance the article representations with additional knowledge from the knowledge graph. For entity linking, we use the modular and lightweight Radboud Entity Linker (REL) proposed by Hulst et al. [25]. We chose REL over the BLINK entity linker [60] for its lower computational cost and interchangeability of the NER-model and the knowledge graph. For named entity recognition we use flair ner-english-ontonotes-large[14], an 18-class model for NER. Unlike many other state-of-the-art approaches to NER, which typically process text at the sentence level, this NER model allows entities to be captured across sentences at the document level [47]. The higher number of classes allows to capture more entities but requires the exclusion of some of the classes from linking since they naturally have no counterpart in the knowledge graph. Figure 17 shows the entity classes. Classes marked red are excluded. Entities with classes marked green are considered for linking.

class	meaning
CARDINAL	cardinal value
DATE	date value
EVENT	event name
FAC	building name
GPE	geo-political entity
LANGUAGE	language name
LAW	law name
LOC	location name
MONEY	money name
NORP	affiliation
ORDINAL	ordinal value
ORG	organization name
PERCENT	percent value
PERSON	person name
PRODUCT	product name
QUANTITY	quantity value
TIME	time value
WORK_OF_ART	name of work of art

Legend:

Mentions with this classes were used for linking

Mentions with this classes were excluded, because they can not be linked to the knowledge graph

Figure 17: Entity classes used in the process of Named Entity Linking for articles in the EUMIG article corpus.

Entities are linked to the WikiData [54] knowledge graph, a public, community-maintained knowledge graph, in which each entity is assigned a unique Q-Identifier. The Q-Identifier for the European Union is Q458[15], for example. Unique identification is necessary as a name can be ambiguous and can therefore refer to multiple different entities. Entity disambiguation predicts the entity most likely to be referenced in the given context. For entity disambiguation REL considers latent relations between entities mentioned in the news article [30]. For each prediction, a certain linking confidence is assigned. To replicate the setting of the experiments in the KRED paper [34, Section 3.1], the EUMIG article corpus only contains entities with a linking confidence of at least 0.9.

[14] `www.huggingface.co/flair/ner-english-ontonotes-large` (Last accessed: 2021-08-20)
[15] `https://www.wikidata.org/wiki/Q458` (Last accessed: 2021-08-20)

An example of an ambiguous entity in the context of migration in the EU is the Moria refugee camp. Even though mentions of Moria in the EUMIG article corpus get tagged as a geo-political entity by the flair NER-model, in some rare cases, REL disambiguates them to Moria (middle-earth), a fictional location in the works of J. R. R. Tolkien.

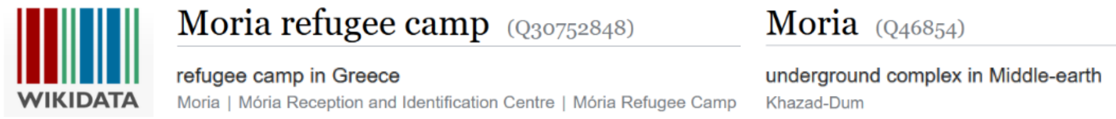


Figure 18: Two ambiguous entities[16,17] in the context of migration in the EU.

This is a remarkable but rare edge case that was observed during entity linking in the EUMIG article corpus. Since entity disambiguation is a research problem in its own right, it would go beyond the scope of this bachelor thesis to solve this problem. We leave it at this point with a hint to the problem for further research.

4.3.2 Linking statistics

Figure 19 shows that geographic entities were linked most frequently, accompanied by important politicians, political parties, and political institutions.

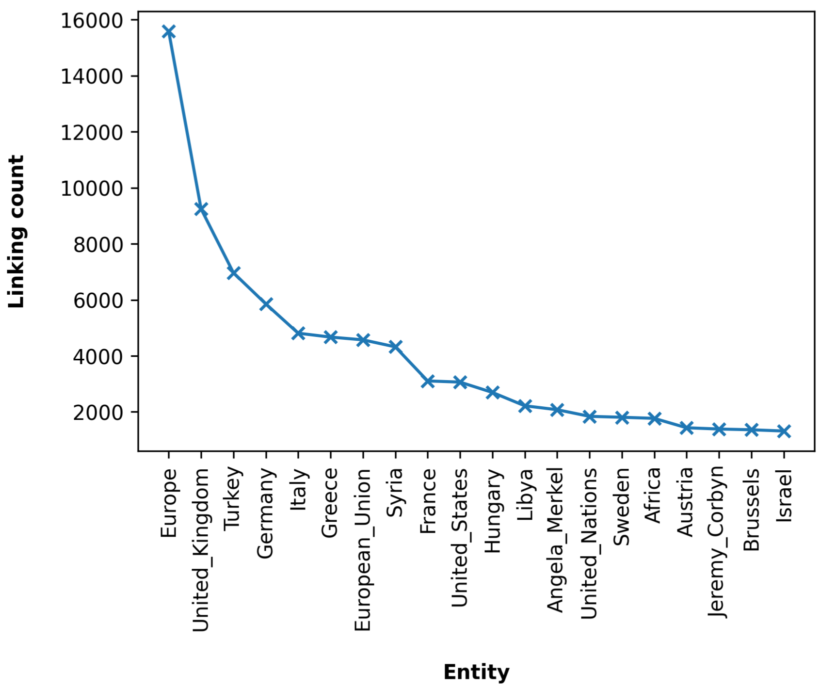


Figure 19: Entities most frequently linked in the EUMIG article corpus.

Statistical evaluations of the most frequently linked entities for the articles of each of the news outlets in the EUMIG article corpus are appended in Appendix A.

[16] `https://www.wikidata.org/wiki/Q30752848` (Last accessed: 2021-09-27)
[17] `https://www.wikidata.org/wiki/Q46854` (Last accessed: 2021-09-27)

4.3.3 EUMIG article corpus format

News article information in the EUMIG article corpus consists of seven tab-separated columns: News-ID, Category, Subcategory, Title, Abstract, URL, Title Entities (entities mentioned in the title of this news article), Abstract Entities (entities mentioned in the abstract of this news article). The News-ID suffix identifies the outlet, which denotes the article's political bias. Table 2 illustrates the structure of an exemplary news article.

Column	**Content**
News ID	N-0020-B
Category	news
Subcategory	eu migration
Title	Nearly 570 Boat Migrants Land in UK in Four Days, Highest Number of 336 Land in One Day
Abstract	The Home Office has confirmed that 568 illegals crossed the English Channel in small boats to reach Britain in just four days.
URL	`www.breitbart.com/europe/2021/06/01/nearly-570-boat-migrants-land-uk-four-days/`
Title Entities	[{"Label": "United_Kingdom", "Type": "GPE", "WikidataId": "Q145", "Confidence": 0.9992049336433411, "OccurrenceOffsets": [33], "SurfaceForms": "UK"}]
Abstract Entities	[{"Label": "Home_Office", "Type": "ORG", "WikidataId": "Q763388", "Confidence": 0.9230267206827799, "OccurrenceOffsets": [0], "SurfaceForms": "The Home Office"}, {"Label": "English_Channel", "Type": "LOC", "WikidataId": "Q34640", "Confidence": 0.926857570807139, "OccurrenceOffsets": [56], "SurfaceForms": "the English Channel"}, {"Label": "United_Kingdom", "Type": "GPE", "WikidataId": "Q145", "Confidence": 0.9999415874481201, "OccurrenceOffsets": [100], "SurfaceForms": "Britain"}]

Table 2: Format of news articles in the EUMIG article corpus illustrated by an exemplary news article from Breitbart[18].

The original EUMIG article corpus assembled as described in sections 4.2 and 4.3.1 contains columns for article body and body entity information. Due to licensing and as that information is not available in the MIND dataset, the model was trained on, we exclude body and body entity information from the bias analysis. We release the EUMIG article

[18] `www.breitbart.com/europe/2021/06/01/nearly-570-boat-migrants-land-uk-four-days/` (Last accessed: 2021-08-24)

corpus along with the code for this thesis at `https://github.com/hannahgreven/kred-bias-analysis`.

4.3.4 Knowledge graph embeddings

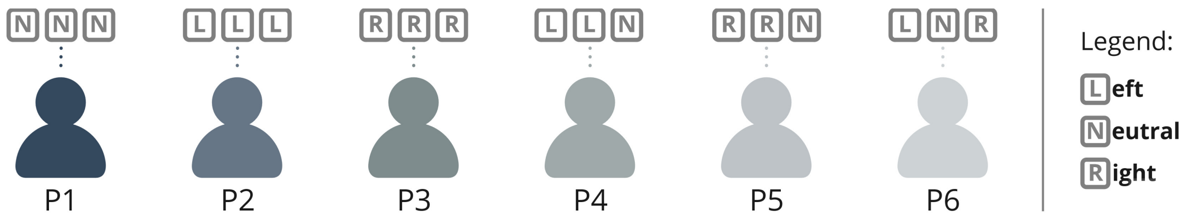
We use the 100-dimensional TransE embeddings of the WikiData knowledge graph (entity_embedding.vec and relation_embedding.vec) supplied with the MIND dataset. With embeddings of 3,275,149 entities in the English WikiData the MIND knowledge graph embeddings cover 91,82 % of the entities linked in titles and abstracts of the news articles in the EUMIG article corpus.

4.4 Bias analysis

4.4.1 Recommendation setup

To study the behavior of the KRED NRS regarding political bias introduced from the articles in a user's read history, user data is necessary. User data in the news domain is sensitive personal data according to the GDPR [43]. Thus we do not have access to real-world user data for the collected articles in the EUMIG article corpus. Instead, we simulate user behavior synthetically based on distinct news reception profiles.

News reception is a field of research in media science that deals with the behavioral patterns of news consumption. Related to this field of study, we defined six exemplary news reception profiles listed in Figure 20.



P1: Least biased	Only reads news from Reuters
P2: Left	Only reads news from The Canary
P3: Right	Only reads news from Breitbart
P4: Left preference	Reads news from Reuters and The Canary (balance 1:2)
P5: Right preference	Reads news from Reuters and Breitbart (balance 1:2)
P6: Diverse	Reads news from all outlets in equal (balance 1:1:1)

Figure 20: News reception profiles for the bias analysis covering selective exposure (P1, P2, P3), mixed exposure (P4, P5), and diverse exposure (P6).

With those profiles, we model different kinds of exposure to political bias from the articles read. As explained in section 4.2 an article's bias is assumed to be the same as the bias of the outlet in which it is published. Correspondingly for our experiment, the outlet-balance in the user's news diet defines the political bias the user is exposed to. Selective exposure profiles model the user's exclusive exposure to news that covers one side of the political spectrum (P2, P3) or deliberately avoid any political opinion (P1). Mixed exposure profiles model the user's exposure towards one side of the political spectrum and additional least-biased news (P4, P5). The diverse exposure profile (P6) models the user's equal exposure to content politically least-biased, left, and right. Based on these profiles, we generated 100 user-item sets to perform recommendation experiments. Figure 21 shows a schematic overview of the experiment setup.

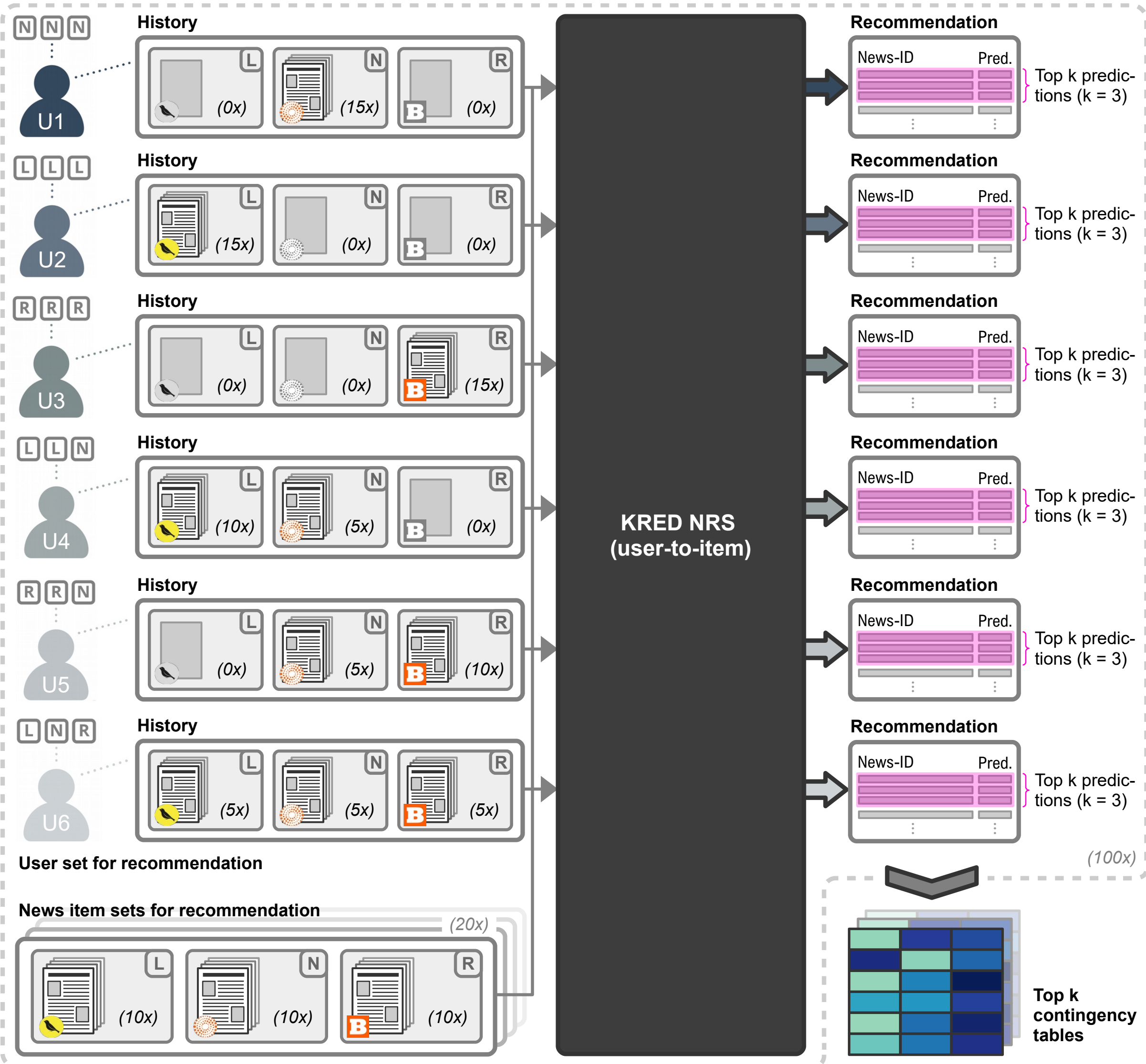


Figure 21: Schematic overview of the recommendation experiment and analysis setup.

Each user set consists of six synthetically generated users (figure 21, U1-U6), one for each news reception profile (figure 20, P1-P6). We generate users synthetically by randomly inserting 15 articles from the EUMIG article corpus into the user behavior file that match the outlet balance in the associated reading profile. Articles in the user histories were chosen so that they do not overlap. Furthermore, articles added to the user histories are excluded from the articles to be recommended, as users are usually not interested in reading articles twice. The remaining articles are split into item sets of thirty news articles, containing ten articles from each outlet to equally balance bias. The resulting user-item test sets are then passed to the KRED model as input for user-to-item recommendation.

4.4.2 Recommendation analysis of exposure to political bias

In the following, we use the notation for deep learning NRS introduced in section 2.1. We use our trained KRED working model to generate recommendations for a user u_i from the user-set $U = \{u_1, ..., u_6\}$ on a news item set $V = \{v_1, ..., v_{30}\}$ with bias-balanced news items v_j. The resulting preference prediction values $\hat{y}_{ij} = f(u_i, v_j) \in [0, 1] \subset \mathbb{R}$ are sorted decreasingly. With this method, we generate recommendations for 100 user sets on 20 item sets each. Each item set contains 30 bias-balanced news items. In total, we thus generate 360,000 user-to-item recommendations. To answer the questions on whether there is a significant relationship between political bias in the user reading profile and outlet exposure, we analyze the top-k recommendations for $k = 3$, $k = 5$, and $k = 10$. This means we count the frequency of news outlets in the news items with the k highest preference predictions for a user on a given item set. These values are summed up by news outlets and user reading profiles, resulting in a contingency table. We analyze the outlet frequencies in the top k recommendations using Pearson's Chi-Square test. The Chi-Square test determines whether there is a stochastic correlation between outlets with associated political bias in the news reception profiles and outlets with associated political bias in the recommendations of the KRED NRS. Additionally, we analyze the contingency tables using heatmap plots.

5 Results

This chapter reports the results of the analysis and their interpretation. The subsequent discussion section explores the three key components of the KRED NRS with respect to internal biases that could contribute to bias within the KRED NRS.

5.1 Analysis results and interpretation

A Chi Square test of independence was performed to assess the relation between news reception profiles and outlet frequency in the top-k recommendations. In the following, we exemplary refer to the results of the analysis for k = 3. The relation between news reception profiles and outlet frequency is significant, $\chi^2(10,\ N = 36,000) = 2438.131$, $p < .001$. Further insights derive from the frequency distribution in the contingency table.

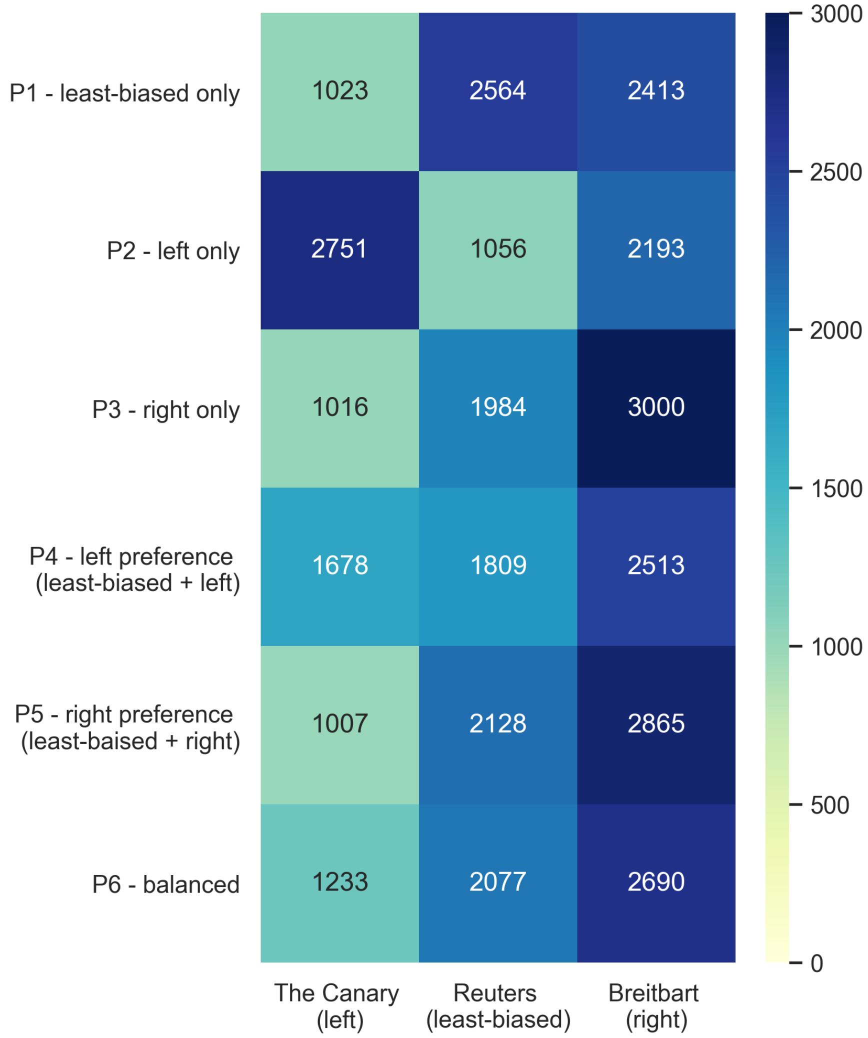


Figure 22: Absolute frequencies of news outlets in the top-3 recommendations according to news reception profiles. Yellow denotes low frequency, blue denotes high frequency.

The heatmap plot of the contingency table for the selective exposure profiles P1, P2, and P3 (see Figure 23) shows that the news outlet from the news reception profile is recommended most frequently.

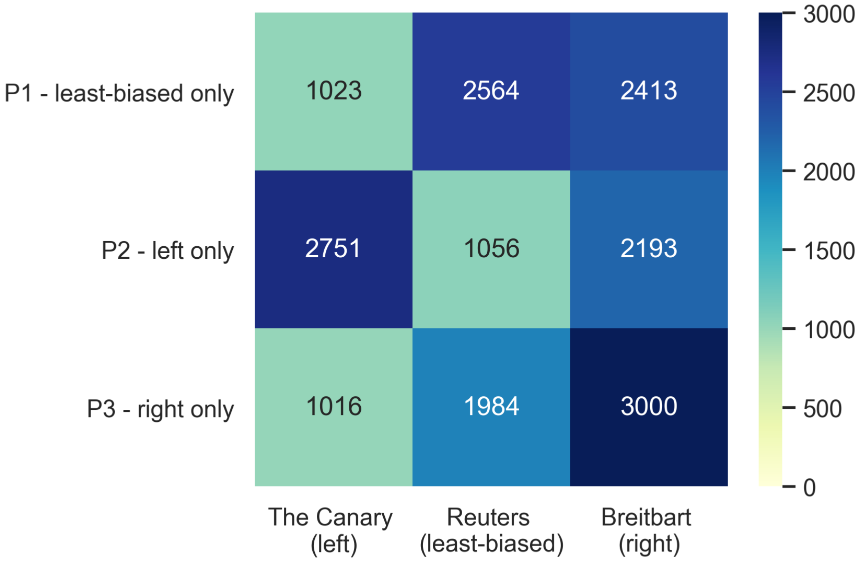


Figure 23: Absolute frequencies of news outlets in the top-3 recommendations for selective exposure news reception profiles.

Consequently, the KRED NRS recommendations captures user preference for news outlets and therefore amplifies political bias contained in the articles in the user history. However, this only happens to a limited extent, as items from other outlets are still recommended in the top-k of the recommendation list. For the mixed exposure profiles P4 and P5, articles from the preferred news source are recommended more often for the respective profile when compared by outlets among the two profiles. Articles from The Canary are more frequently recommended to user with news reception profile P4 than to users with news reception profile P5. The same applies to P5 and Breitbart. For P4, however, articles from Breitbart are recommended most frequently overall.

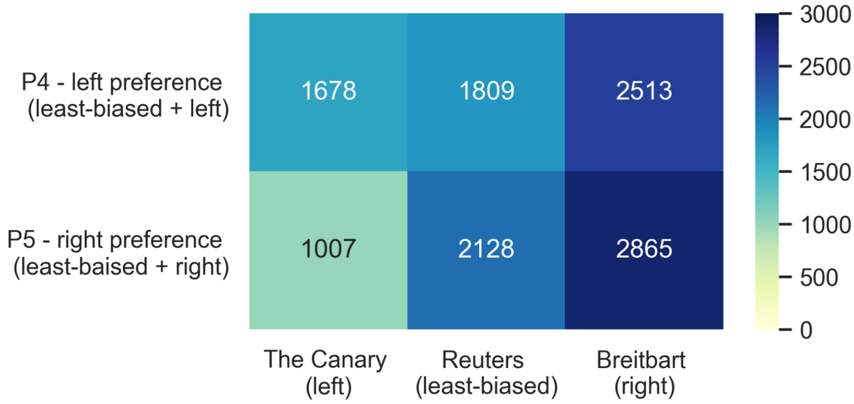


Figure 24: Absolute frequencies of news outlets in the top-3 recommendations for mixed exposure news reception profiles.

Comparing the columns of the contingency table in Figure 22, we find that articles from Breitbart, the news outlet with right political bais in the EUMIG article corpus, are recommended considerably more often across all news reception profiles. The frequencies for Reuters articles and Breitbart articles are similar as long as one of the two news outlets is present in the news reception profile (see Figure 22, P2). The lowest frequencies in recommendations across all reading profiles have articles from The Canary, the outlet with left political bias in the EUMIG article corpus. This means the KRED NRS unfairly favors the news outlet with right political bias from the EUMIG article corpus in its recommendations. This is most evident from P6 (see Figure 25). Although all outlets are read equally frequent, Breitbart is represented in the top-3 recommendations with above-average frequency, whereas The Canary is represented with below-average frequency. The expected average frequency in this case would be 2,000. The frequency of least-biased articles in the recommendations for P6 is 2,077, which is close to the expected frequency.

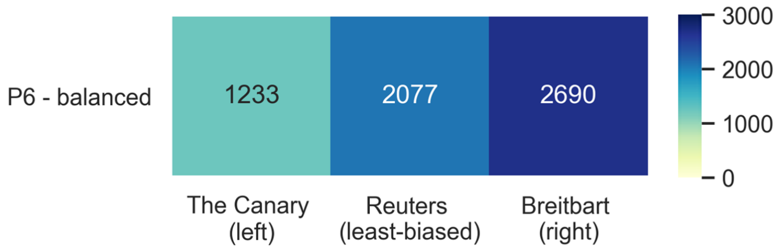


Figure 25: Absolute frequencies of news outlets in the top-3 recommendations for the diverse exposure news reception profile.

The same results are evident for the news outlet frequencies in the top-5 and top-10 recommendations. Heatmap plots of the corresponding contingency tables are enclosed in Appendix B.

Summary of findings from the analysis of political bias in exposure:

F1 The relationship between the news reception profiles and the outlet frequency in the top-k recommendations of the KRED NRS is stochastically significant (result of Pearson Chi Square Test).

F2 In case of selective exposure (P1-P3), when the user's history exclusively contains news from one outlet with politically left or right or least-biased content, the KRED NRS recommends articles from the user's preferred news outlet most frequently.

F3 Articles from the news outlet Breitbart are preferenced by the KRED NRS regardless of the user's associated news reception profile.

5.2 Discussion

Findings F1 and F2 are to a certain extend desirable for an accurate news recommendation system. The users receive recommendations for content and outlets that previously interested them and are therefore most likely to be of interest to them in the future. Articles from other news outlets on the same topic are also recommended, although less frequently. With regard to the beyond accuracy metrics (see Figure 11) and the different stakeholders (see Figure 1), more fairness for providers in terms of exposure and visibility could be achieved by subsequently rebalancing the recommendations. Rebalancing is also applicable to establish diverse political opinion in the news content the user is exposed to. Furthermore, recommendations could be subsequently expanded using methods for topic clustering and political opinion detection so that the spectrum of political stances on a topic is covered equally.

In contrast to findings F1 and F2, observation F3 reveals unexpected behavior that requires further investigation. A systemic bias towards a specific political stance present in a NRS is generally undesirable, since news recommendation systems with a broad outreach contribute to the formation of social opinion. In this section we want to explore where the bias towards the right news outlet in recommendations of the KRED NRS could stem from. Finding the cause within the time constraints of this thesis is not feasible due to the complexity of the KRED NRS. Instead, we intend to highlight existing research related to bias in the three key components the KRED NRS builds on. Methods found to debias those components point out directions for future research.

Each key component has its own internal biases which impact the recommendation quality of the KRED NRS. The three key components that can introduce bias to the KRED NRS are the initial BERT news article embeddings, the WikiData knowledge graph along with its embeddings used to enhance the initial news article embeddings, and the MIND dataset used for training.

5.2.1 Bias in BERT

In our experimental KRED NRS setup, a pre-trained BERT model calculates the initial semantic-level news article vector embeddings. BERT is a bidirectional transformer that is available pre-trained. BERT has been trained on the BookCorpus dataset [63] and the Wikipedia using masked language modeling and next sentence prediction.

Specific forms of bias in BERT can be assessed by querying BERT for the likelihood of specific tokens. Tokens in this context denote building blocks of natural language in general. This means tokens can also be subwords or words. Masking a token with a specific mask token allows to measure the likelihood of other tokens (e.g. specific words) to replace the mask tokens in a sentence.

Racial and gender bias Certain tokens, e.g., ethnicity-specific names or gender-specific pronouns, are unequally distributed in the Wikipedia and Toronto BookCorpus training data. By conducting experiments with European American and African American names, [10] et al. show that BERT suffers from a racial bias, more precisely from bias in the representation of African American names due to the underrepresentation of African American names in the BookCorpus dataset and Wikipedia. Similar findings were reported for gender pronouns by Kurita et al. [29].

Both are biases to be aware of when using BERT in news recommendation systems, as BERT encodes this bias into the news articles' vector representations. Racial bias might be particularly relevant in the context of migration in the EU. Since BERT was trained on books and Wikipedia in English, it is likely that non-Western names in general are underrepresented. Non-Western names are not atypical in the context of migration in the EU. As a result, the predictions of the KRED NRS might vary depending on cultural associations with news articles. Future research might explore, whether negative sentiments toward ethnic groups introduced through political bias in the news article, could have a reinforcing effect.

Debias options It would be interesting to see, whether this kind of racial bias is also observed for the multilingual BERT model (mBERT)[19], which supports 104 languages and was released along with BERT [14]. Research on bias in BERT mostly focuses on the English BERT model. Zhao et al. research gender bias in multilingual embeddings and cross lingual transfer including mBERT [62]. To the best of our knowledge research on racial bias in BERT has been limited to EnglishBERT so far, which is used in this thesis.

Racial bias is only one manifestation of many societal biases that can be reflected in BERT. To eliminate bias from the BERT news article vector representations, several methods have been proposed. Retraining BERT is usually not feasible as the amount of training data is massive and training therefore is of a computational intensity that requires training on hundreds of machines for several weeks. Therefore, research on debiasing BERT focuses on post-hoc debiasing techniques, which add a post-training debiasing step that mitigates bias before the model is used [33, 28, 19]. Using such a post-hoc debaising technique before using BERT in the KRED NRS should help mitigate bias. However, a trade-off must be made between most accurate and bias-free article representation, as noted by Kaneko et al. [28].

[19] `https://github.com/google-research/bert/blob/master/multilingual.md` (Last accessed: 2021-08-26)

5.2.2 Bias in WikiData

As a knowledge-aware NRS the KRED NRS builds on a knowledge graph. In our experimentel setup, this knowledge graph is a subgraph of WikiData. WikiData is a crowd-sourced, collaboratively created kowledge graph, which means that in theory everyone with internet access can add information to the WikiData knowledge graph. This allows for a fast and high-quality acquisition of complexly structured content.

Contributor bias Information stored in a knowledge graph can be highly influenced by the population of contributors. Each contributor has an implicit personal bias due to socialization and previous experience. Demartini [13] investigates implicit bias in crowd-sourced knowledge graphs such as WikiData. To understand which factors significantly influence attitudes toward controversial facts and thus can indicate implicit contributor bias, he examines the relationships of contributor gender, age, task duration, and search result rank of supporting evidence in three crowdsourcing experiments. Demartini deems it most feasible to capture indicators for implicit contributor bias as additional triples in the knowledge graph to make bias transparent.

Completely eliminating bias from a knowledge graph is difficult since it models "real-world" data, which contains social biases by nature. However, certain biases arise from groups of people over- or under-represented in WikiData relative to real-world data. Shaik et al. [48] identify differences in representation to characterize the biases present in WikiData by comparing WikiData queries to real-world datasets. They find that WikiData is skewed overall towards the white race, while all other races are underrepresented, and that WikiData overrepresents European and North American individuals by a factor of 3-6 compared to the real world population. To improve the representation of underrepresented groups, they propose injecting catalogs of individuals belonging to these categories with a minority bot.

Bias in the embeddings Besides bias in the information in the knowledge graph itself, bias can also be encoded in knowledge graph embeddings. Same as for sentence embeddings like BERT, knowledge graph embeddings can encode potentially harmful social stereotypes from unequal distributions of people of different genders, ethnicities, religions and nationalities in the knowledge graph. Fisher et al. propose a binary contrastive measure to analyze bias in knowledge graph embeddings [17]. In their work they conduct experiments on TransE embeddings of WikiData, which show that WikiData also suffers from gender and racial bias related to professions. Of particular interest is the finding that the bias is not entirely attributable to the frequency of certain groups in the knowledge graph. For example, in WikiData, the number of entries for African American publishers is greater than the number of entries for Jewish publishers. Yet with TransE [5] "publisher" is sterotypically embedded as a Jewish profession. The same is true for the professions "entrepreneur" and "economist" [17].

Different from Fisher at al. Bourli et al. [6] focus on whether bias in the knowledge graph is amplified by the knowledge graph embeddings. They find that gender bias regarding professions exists in WikiData and is amplified by TransE embeddings. They propose the first and, to the best of our knowledge, only debias method proposed so far for knowledge graph embeddings. Their debias approach is based on projections on the gender subspace. In general the method can be applied to any type of sensitive binary attribute. It is novel that the approach is tunable in the amount of bias it removes, which allows to adjust the penalty on accuracy.

In the KRED NRS, the initial BERT article embeddings are extended with additional background knowledge using mentioned entities and the WikiData knowledge graph to obtain knowledge-level embeddings. Bias in the knowledge graph as well as in the knowledge graph embeddings thus propagates to the KRED NRS. Potentially, different components may even amplify each others bias. This might be the case for BERT and WikiData, since both exhibit racial bias.

5.2.3 Bias in MIND

The MIND Dataset was collected from the user behavior logs of Microsoft News. From October 12 to November 22, 2019 Wu et al. randomly sampled 1 million users who at least clicked 5 news articles [59]. As a result, MIND contains user click behaviors on more than 160k English news articles. As our KRED NRS was trained on MIND, imbalances and bias in MIND can introduce bias to the recommendation system.

To the best of our knowledge, there is no literature on bias in MIND. To better understand the training data and thus possible induced biases in the KRED NRS, this section examines the source website `https://microsoftnews.msn.com/` and the composition of the news articles in the training dataset regarding biasing factors.

Positional bias at the source website Figure 26 shows an exemplary news feed at the Microsoft News (MSN) website today. The content of the MSN website is composed of several such thematic feeds, which are placed below each other in scrolling direction. It is noticeable that, displayed on PC, the news article in the first position in the top left of each feed clearly stands out from the other smaller articles in the feed due to its size and the full image background. For tablets a similar structure applies. On smartphones, articles in a feed are listed as equally sized items. Likewise, in archived records of the MSN website from the collection period of the MIND dataset, the news article in the first position of each feed clearly stands out from the other smaller articles in the feed due to its size and header image. Figure 27 shows an exerpt of the news feeds from Internet Archive records of the MSN website dated November 19, 2019.

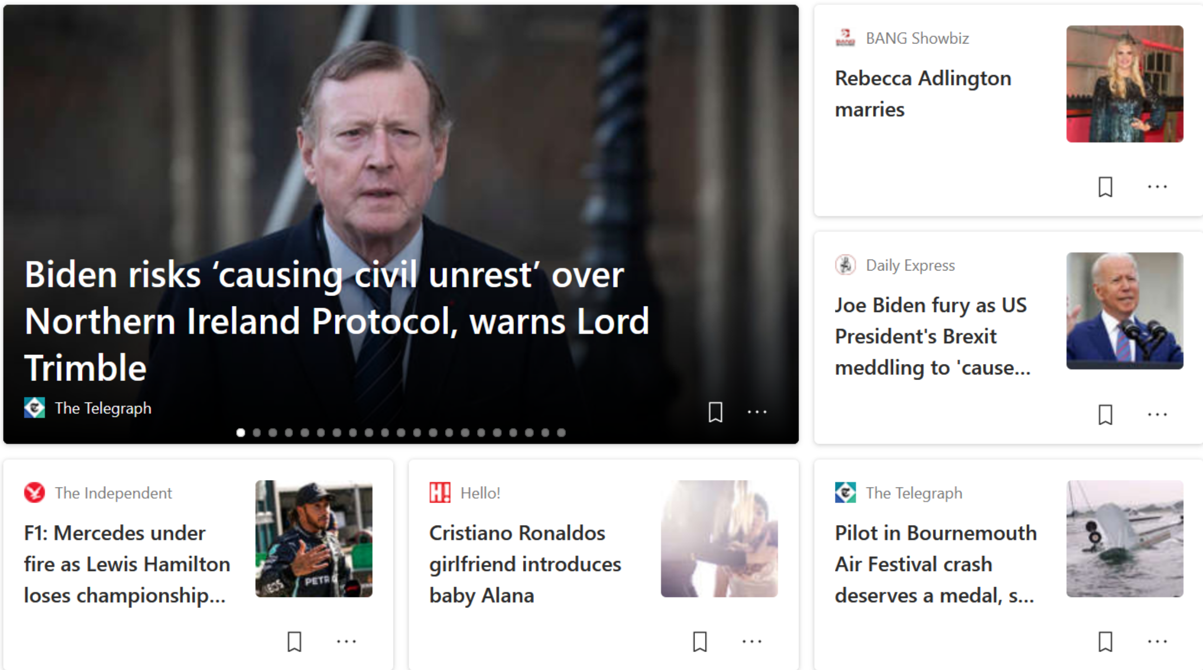


Figure 26: Example MSN home feed[20] accessed on PC September 15, 2021.

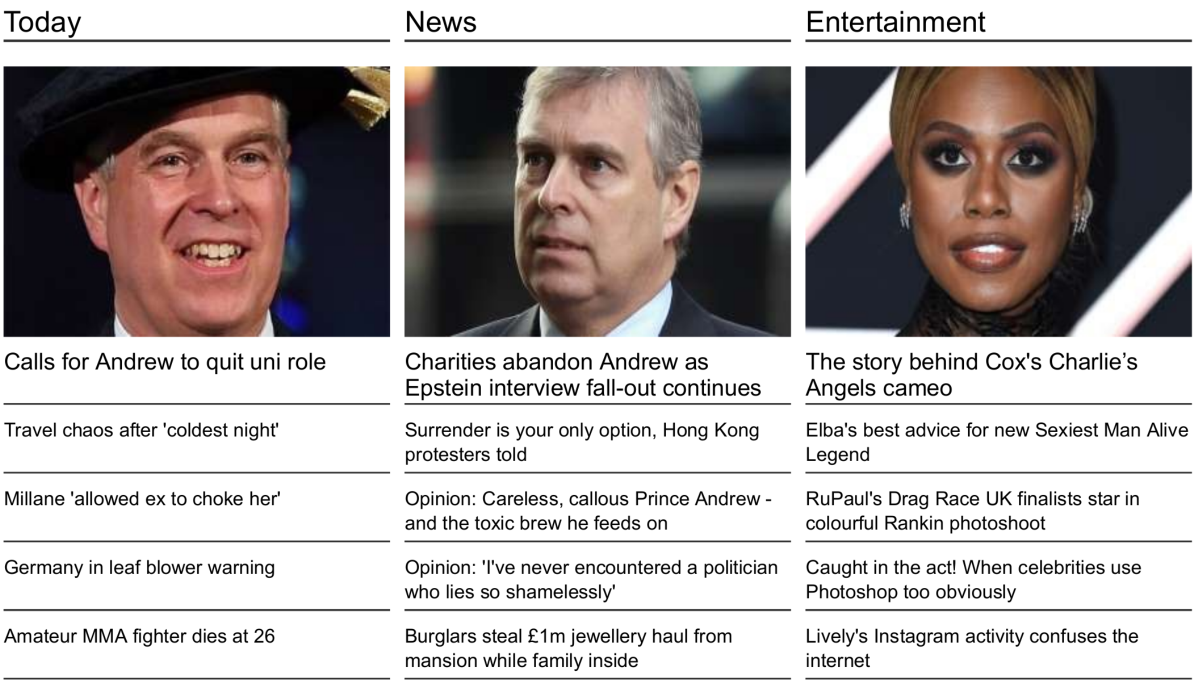


Figure 27: Excerpt from a MSN feed[21] archived November 19, 2019

[20] https://microsoftnews.msn.com/ (Last accessed: 2021-09-06)
[21] https://web.archive.org/web/20191119092639/http://www.msn.com/ (Last accessed: 2021-09-15)

In examining factors influencing news user decisions regarding news diversity, Loecherbach et al. [35] measure the impact of news article position on the user's likelihood of interacting with a news article. They find that the position effect is particularly pronounced in Western cultures and on devices with smaller screens. In their experiment, the news articles in the first position were clicked in 31 % of all cases, and in 50 % of the cases for smartphones. The click probability of an article decreases towards the last position (lower right corner), where the click probability in their experiment was 8 %. The highlighting of the first news article by size and image on the MSN website is likely to further increase this effect. Thus, it is possible that the position effect both within articles of a feed and between feeds for specific news categories introduces bias to the MIND user click data.

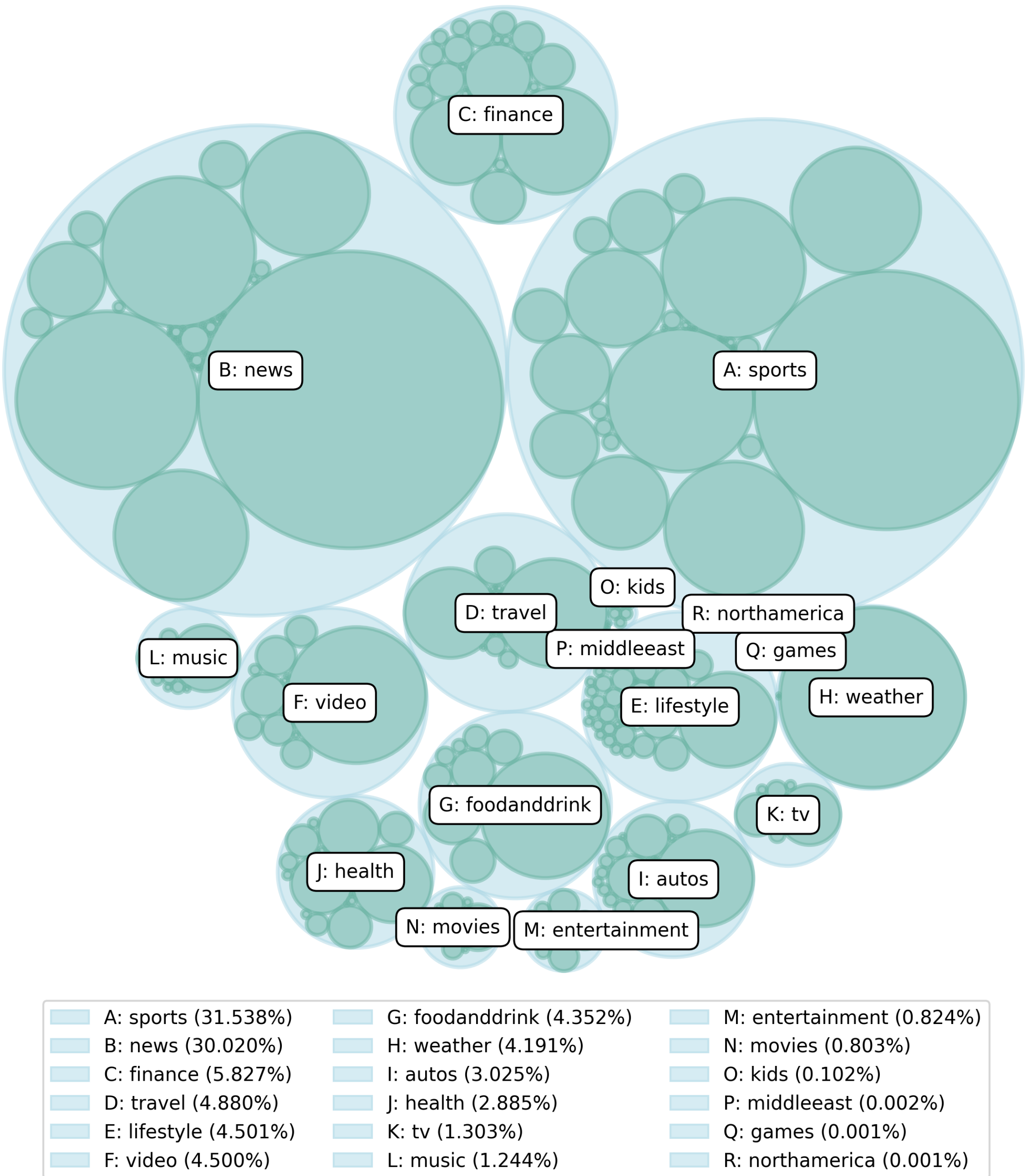


Figure 28: Partition of the MIND-large training data by categories and subcategories. Blue circles denote categories. Green circles depicted inside indicate subcategories.

Composition of the training data The news article data in the MIND dataset includes category and subcategory for each news article. Categories and Subcategories provide metadata that the KRED NRS does not use in user-to-item recommendation. But frequencies of categories and subcategories in the training data may indicate bias. Figure 28 shows the proportions of categories and subcategories in the MIND training dataset. A list of categories and subcategories with total frequencies is appended in Appendix C.

It is noticeable that the categories are very unevenly distributed. The most frequent category is “sports“, directly succeeded by “news“. All other categories are significantly less prevalent in the news data. The decisive factor with regards to bias is whether this imbalance is also reflected in the user data logs.

Accordingly, we examine the user behavior logs in the training data in terms of viewed and clicked items for each category. Table 3 shows the frequencies of displayed and clicked articles in the MIND-large training data by category.

Category	Displayed	Clicked	Percentage clicked
news	22,779,114	996,645	4.38 %
lifestyle	9,350,922	382,275	4.09 %
sports	8,439,316	396,445	4.70 %
finance	8,066,103	292,574	3.63 %
foodanddrink	5,276,950	158,336	3.00 %
entertainment	5,017,376	152,425	3.04 %
travel	4,512,849	118,724	2.63 %
health	4,364,023	158,779	3.64 %
autos	3,844,996	105,667	2.75 %
music	3,839,268	225,633	5.88 %
tv	3,509,976	207,879	5.92 %
movies	1,868,052	60,152	3.22 %
video	1,355,799	62,475	4.61 %
weather	1,279,127	65,545	5.12 %
kids	2,598	74	2.85 %
northamerica	394	12	3.05 %
middleeast	0	0	0.00 %
games	0	0	0.00 %

Table 3: Frequency of displayed and clicked articles in the MIND-large training data by category.

“News“ is the category most frequently displayed and clicked. Since the articles in the EUMIG article corpus belong to the category “news“, this does not indicate a cause for the observed bias in the KRED NRS.

The number of articles clicked on as a percentage of articles displayed is highest for “tv“, “music“, and “wheather“ (see Table 3, Percentage clicked). For all categories with displayed items, the percentage clicked is between 2% and 6%. Thus, there are no large disparities that would suggest a bias from the unequal distribution of the categories seen in Figure 28. The categories “middleeast“ and “games“, are poorly represented in the training data and not among the displayed articles in the user data logs. The KRED NRS thus learns nothing about articles in these categories.

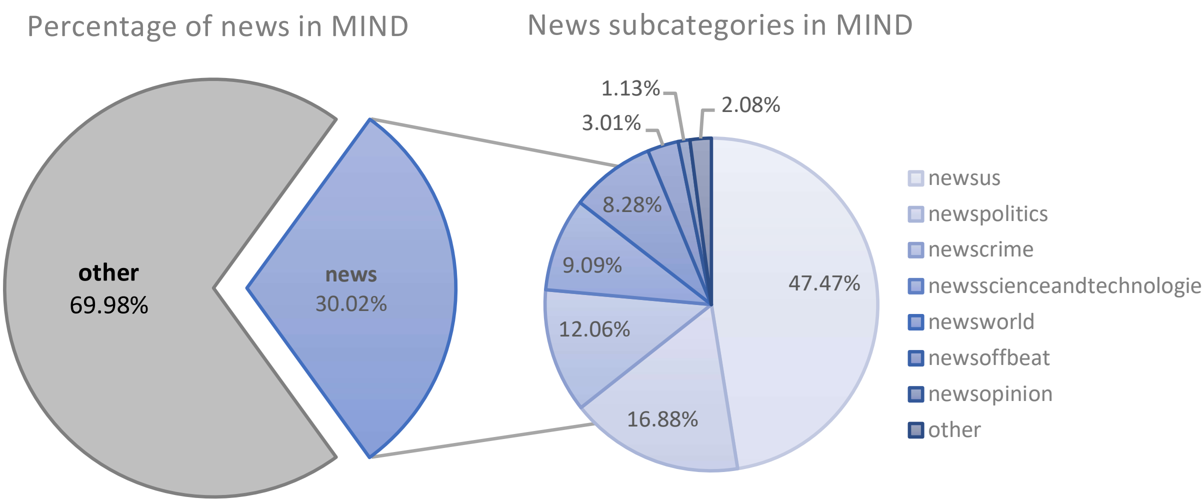


Figure 29: Percentages of category “news“ and news subcategories in the MIND-large training data.

Figure 29 shows the proportions of the “news“ category in the MIND dataset and its most frequent subcategories. It is noticeable that the category “newsus“ accounts for 47.47 % of the news articles in the MIND-large training dataset. The category “newsus“ covers news related to the United States. In contrast, articles in the EUMIG article corpus are topic related to the European Union. In this case, essentially the entities mentioned in the articles differ. In terms of bias, this is otherwise unobtrusive.

The previous considerations addressed bias at the abstraction level of categories and subcategories. For finer conclusions regarding bias in the MIND dataset, future research might examine bias indicators at the article level.

6 Limitations, prospects and conclusion

This chapter outlines limitations, resulting prospects for future research, and concludes the achievement of this thesis.

6.1 Limitations and prospects

Number of news outlets in the EUMIG article corpus The EUMIG article corpus collected for the bias analysis includes articles from three news outlets. Based on the Media Bias/Fact Check classification of the outlets, we assume that each news outlet represents a particular political bias. Expanding the EUMIG article corpus to include articles from other outlets with similar political bias is necessary to resolve the distinct association of a political bias with a single outlet. The bias analysis in section 5.1 showed that the KRED NRS unfairly favors articles from Breitbart. Given the one-to-one association of outlet and political bias, this does not equate to a systemic right bias. The bias analysis of the KRED NRS with expanded article corpus may provide further insight into the KRED NRS's systemic effects regarding political bias.

Level of political bias classification In addition to expanding the dataset, it would be useful for future research to label political bias in the EUMIG article corpus at the article level. In this thesis, a news article's political bias was assumed to be identical with its outlet's political bias. News outlets in the EUMIG article corpus were selected to be unambiguous in their political bias. However, this assumption does not apply to news outlets in general. Labeling items by political bias at the article level could solve the challenge of ambiguous political outlet bias when expanding the article corpus.

News article scope used for recommendation We analyzed the KRED NRS trained on news titles and abstracts since the MIND dataset does not include the text bodies of the news articles. Recrawling the text bodies and linking the entities contained therein was too time-consuming with over 160,000 news articles in the MIND dataset, the given hardware, and limited time. For further research, it might be beneficial to train the KRED NRS on the text bodies of news articles since they contain more semantics and additional knowledge in the form of entities.

Topic-specific knowledge graph In this thesis, we use the WikiData subgraph supplied with the MIND-large dataset. Creating a subgraph and TransE embeddings on the hardware available to us would have exceeded the time frame for the thesis. Using the knowledge graph embeddings from the MIND dataset, we achieve 91.82 % coverage of the entities in the EUMIG article corpus. For further research, building a subgraph on the topic of migration in the EU and training TransE embeddings for this subgraph is worth considering.

6.2 Conclusion

Previous research has shown that evaluating recommendation systems exclusively on accuracy metrics neglects bias effects. Undesirable social bias reflected in training data can manifest itself in the recommendation systems.

In this thesis, we investigated bias in the KRED NRS, a content-based, knowledge-aware NRS trained for personalized news recommendations on the MIND-large dataset. To investigate exposure diversity to political bias in the NRS recommendations, we created the EUMIG article corpus, a corpus of news articles on the politically diverse topic of migration in the EU. We analyze the diversity of the recommended articles to news outlets and associated political bias using different news reception profiles and user behaviors synthesized from them. The bias analysis shows that the KRED NRS learns the user preference for certain news outlets. For users who are associated with a single exposure news reception profile, articles from other news outlets are still recommended. This is positive in terms of diversity, as users are not exclusively recommended articles from their preferred outlet. On a critical note regarding exposure, recommendations are not fairly distributed across all outlets. The KRED NRS prefers articles from Breitbart regardless of the user's associated news reception profile.

Possible factors leading to bias in the KRED NRS are diverse due to the complex nature of content-based, knowledge-aware news recommendation systems. This thesis discussed three key components of the KRED NRS that potentially introduce bias; news article embeddings, creation and embeddings of the knowledge graph, and acquisition and balance of the training data.

As bias is a relatively new area of research for news recommendation systems, literature on the behavior of proposed news recommendation systems concerning bias is sparse. To our knowledge, this thesis provides the first assessment of bias for the KRED NRS. As such, it has limitations that provide opportunities for future research. This thesis is a first step for KRED as a research model towards responsible application in a NRS.

A Entity statistics by news outlets

A.1 The Canary

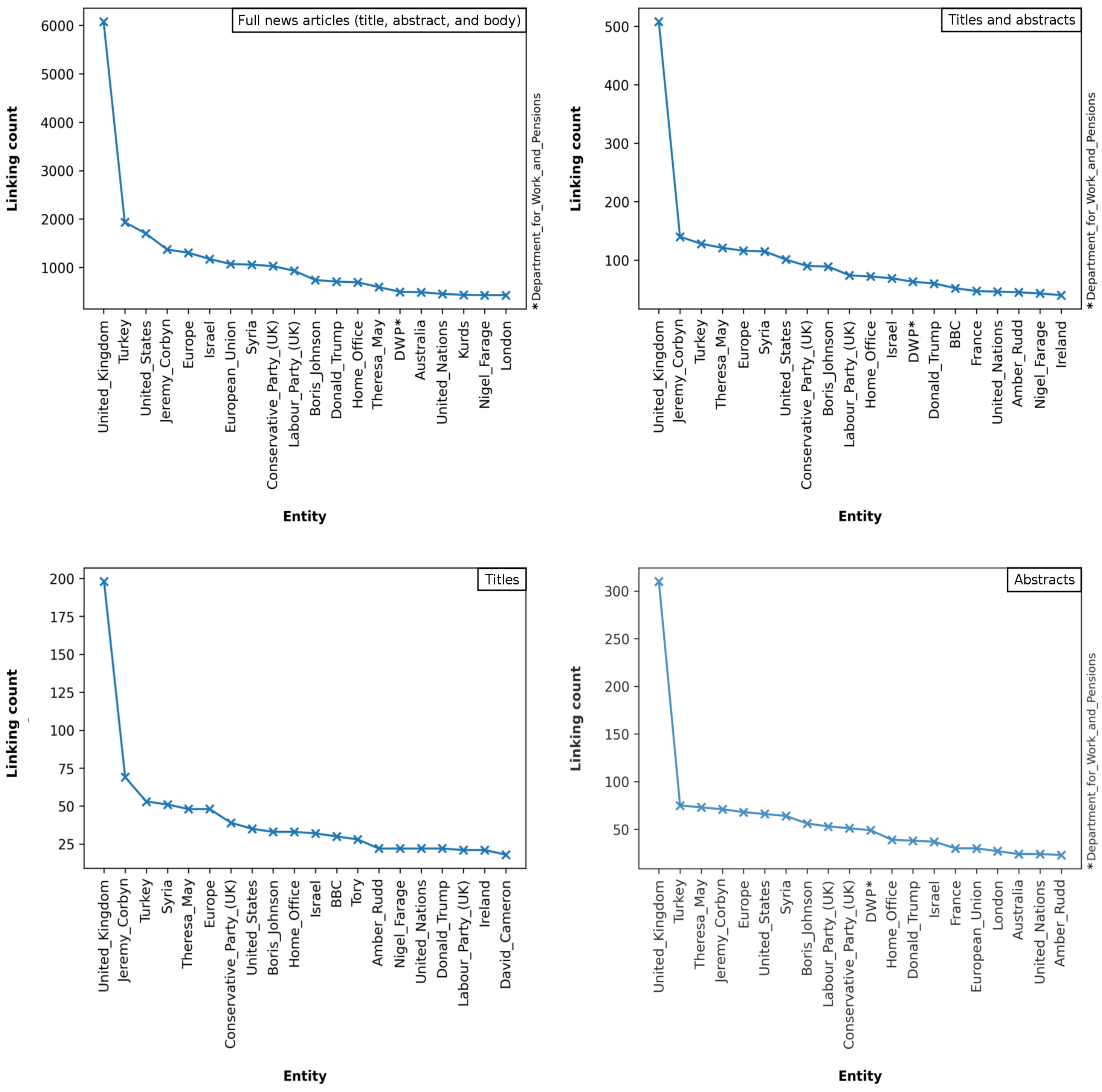


Figure 30: Most frequently linked entities in The Canary news articles in the EUMIG article corpus.

A.2 Reuters

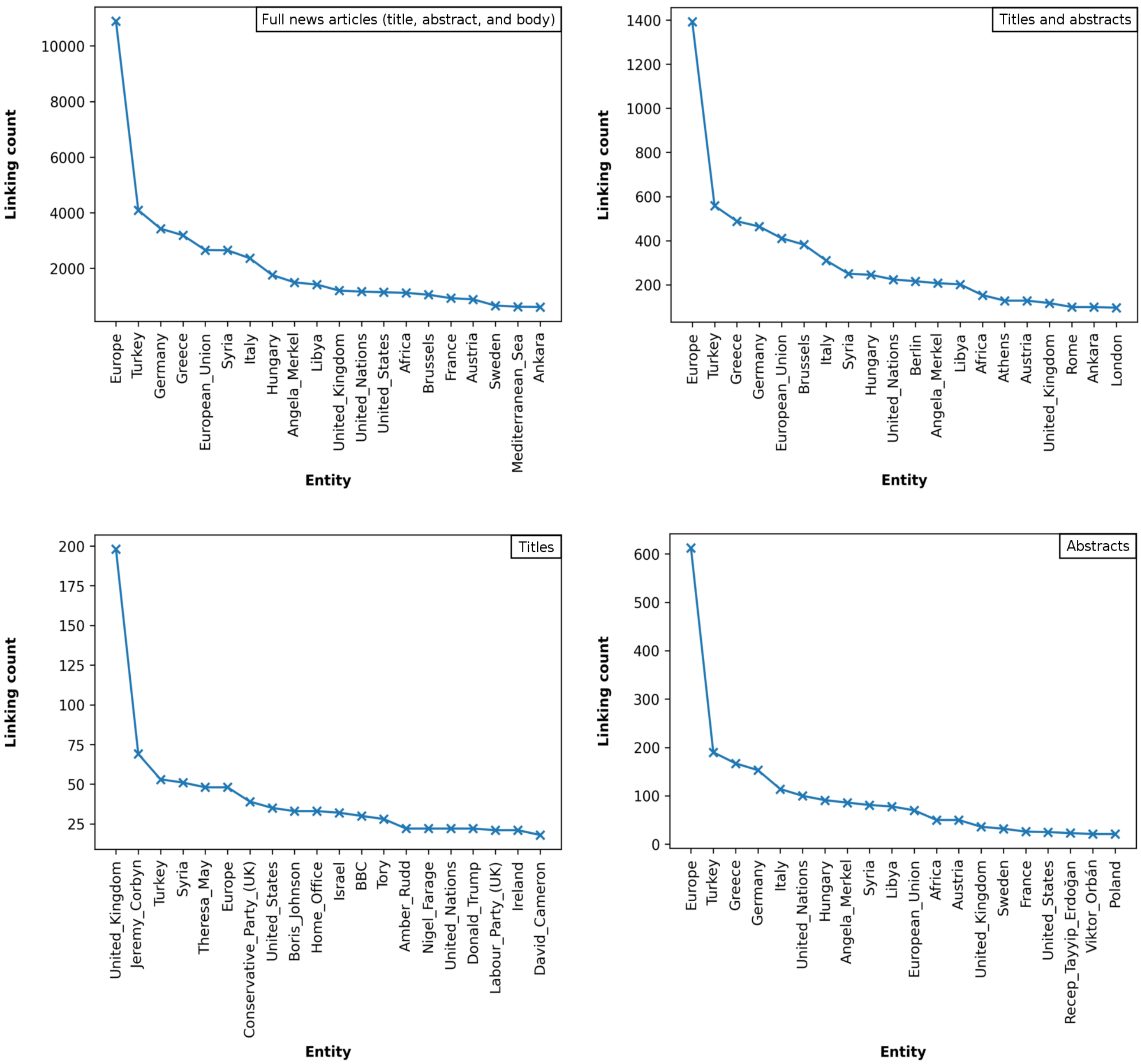


Figure 31: Most frequently linked entities in Reuters news articles in the EUMIG article corpus.

A.3 Breitbart

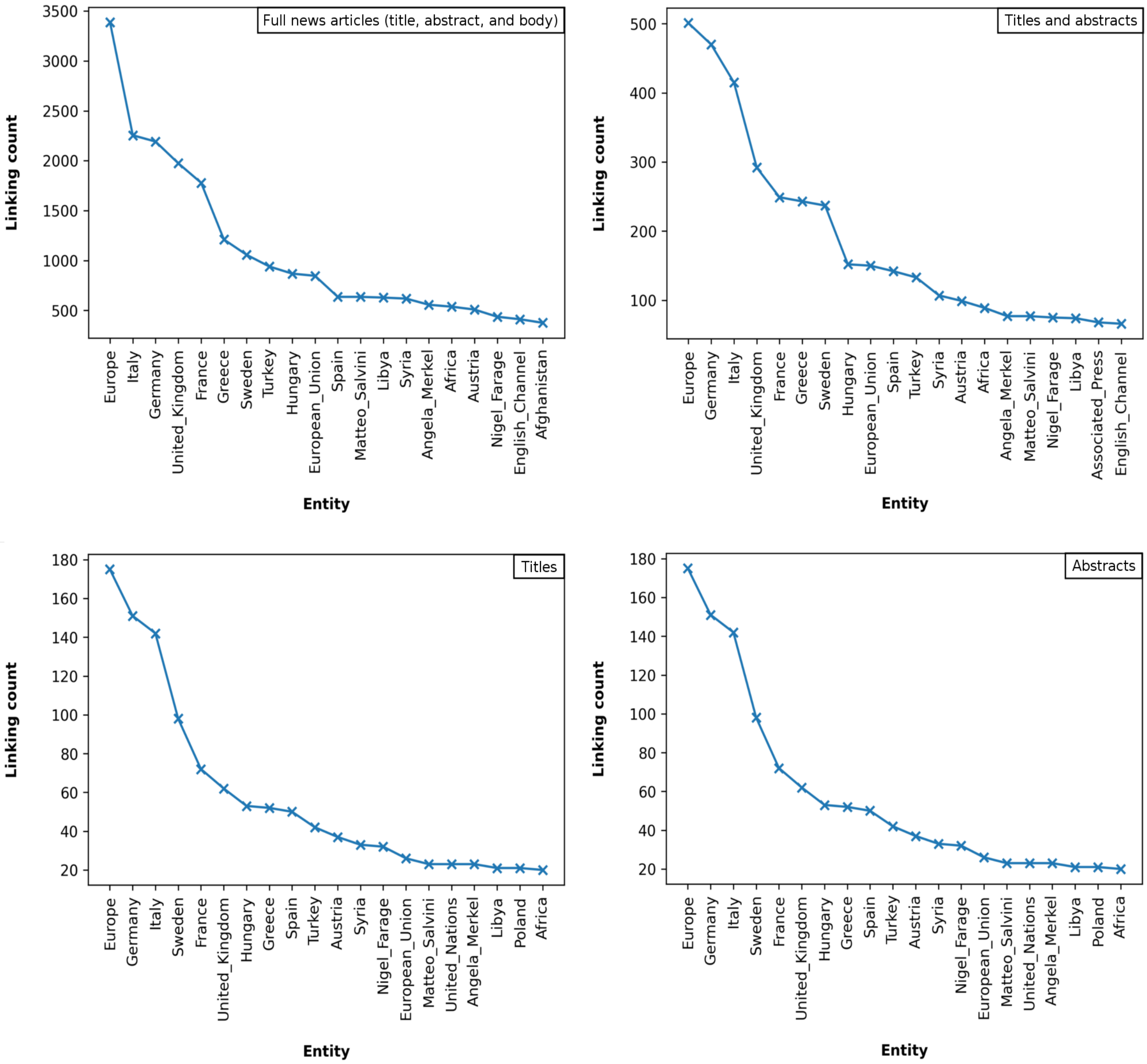


Figure 32: Most frequently linked entities in Breitbart news articles in the EUMIG article corpus.

B Heatmap plots of contingency tables

B.1 Top-5 recommendation results

The relation between news reception profiles and outlet frequency in the top-5 recommendations is significant, $\chi^2(10,\ N = 60,000) = 3228.709$, $p < .001$.

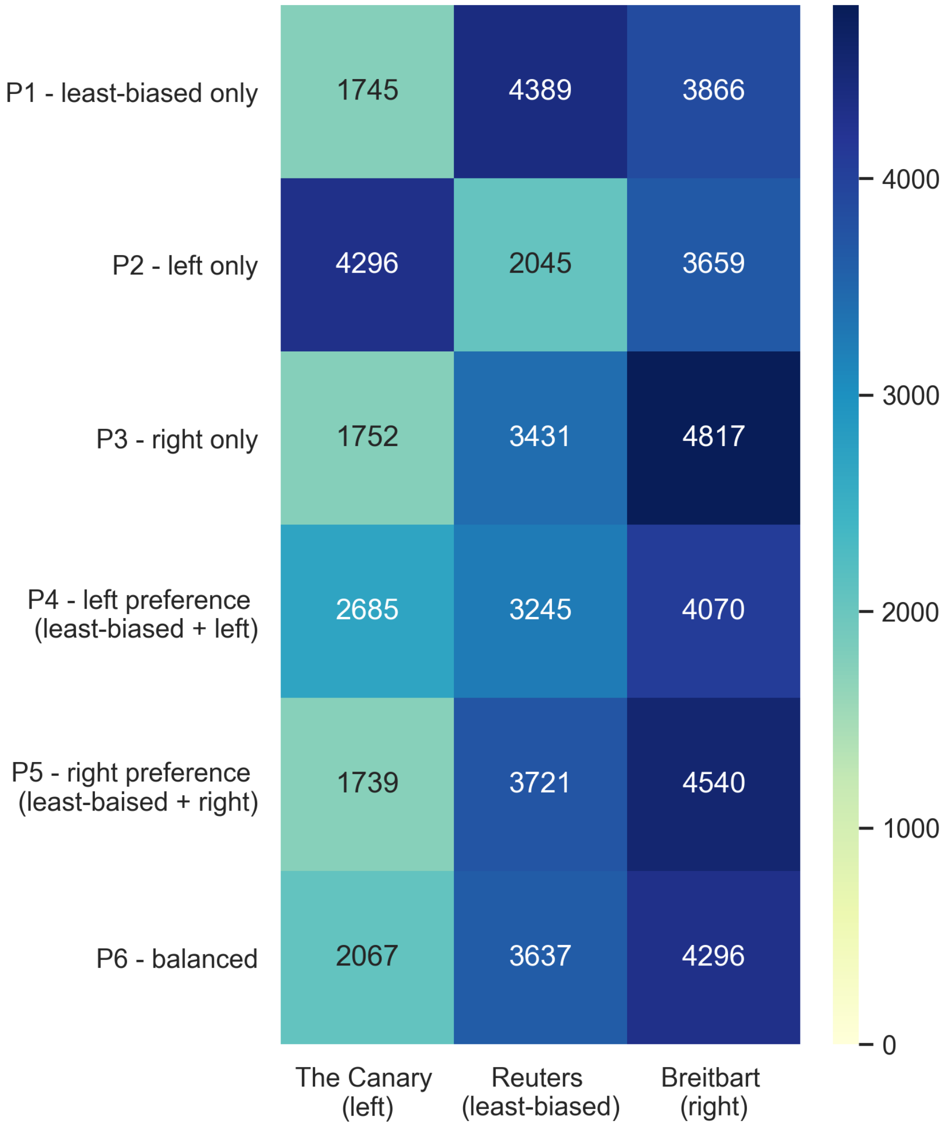


Figure 33: Absolute frequencies of news outlets in the top-5 recommendations according to news reception profile. Yellow denotes low frequency. Blue denotes high frequency.

B.2 Top-10 recommendation results

The relation between news reception profiles and outlet frequency in the top-10 recommendations is significant, $\chi^2(10,\ N = 120,000) = 3425.225$, $p < .001$.

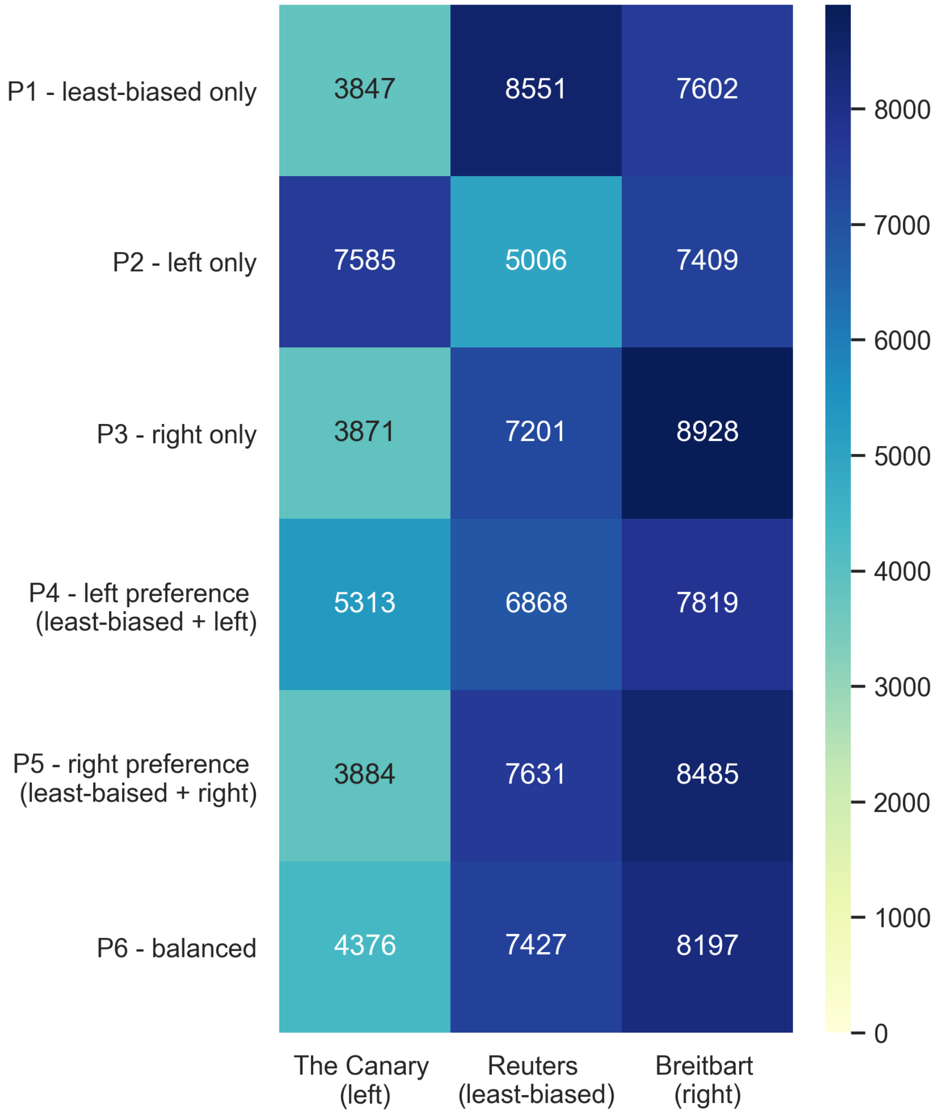


Figure 34: Absolute frequencies of news outlets in the top-10 recommendations according to news reception profile. Yellow denotes low frequency. Blue denotes high frequency.

C MIND-large train data categories and subcategories

- A: sports (32020)
- A.1: football_nfl (11813)
- A.2: baseball_mlb (3617)
- A.3: football_ncaa (3450)
- A.4: basketball_nba (3226)
- A.5: more_sports (2801)
- A.6: basketball_ncaa (1674)
- A.7: icehockey_nhl (1524)
- A.8: mma (1053)
- A.9: golf (759)
- A.10: racing (698)
- A.11: soccer (488)
- A.12: tennis (255)
- A.13: football_nfl_videos (219)
- A.14: outdoors (94)
- A.15: soccer_mls (88)
- A.16: soccer_epl (57)
- A.17: baseball_mlb_videos (49)
- A.18: football_ncaa_videos (48)
- A.19: boxing (47)
- A.20: basketball_ncaa_videos (17)
- A.21: basketball_nba_videos (12)
- A.22: boxing-mma (9)
- A.23: basketball_wnba (8)
- A.24: sports_news (4)
- A.25: soccer_videos (1)
- A.26: olympics-videos (1)
- A.27: golfvideos (1)
- A.28: mmaufc (1)
- A.29: othersports (1)
- A.30: baseball (1)
- A.31: basketball (1)
- A.32: soccer_bund (1)
- A.33: tennis_intl (1)
- A.34: soccer_fifa_wwc (1)
- B: news (30478)
- B.1: newsus (14467)
- B.2: newspolitics (5145)
- B.3: newscrime (3676)
- B.4: newsscienceandtechnology (2771)
- B.5: newsworld (2523)
- B.6: newsoffbeat (918)
- B.7: newsopinion (343)
- B.8: newsgoodnews (190)
- B.9: newsbusiness (138)
- B.10: elections-2020-us (135)
- B.11: factcheck (42)
- B.12: newsphotos (24)
- B.13: newsfactcheck (21)
- B.14: newsscience (18)
- B.15: newsweather (17)
- B.16: newselection2020 (10)
- B.17: newsworldpolitics (5)
- B.18: causes (3)
- B.19: causes-environment (3)
- B.20: newstechnology (3)
- B.21: newsvideo (3)
- B.22: personalfinance (2)
- B.23: empowering-the-planet (2)
- B.24: yearinoffbeatgoodnews (2)
- B.25: photos (2)
- B.26: causes-disaster-relief (2)
- B.27: indepth (2)
- B.28: causes-poverty (1)
- B.29: newslocalpolitics (1)
- B.30: newstvmedia (1)
- B.31: other (1)
- B.32: narendramodi_opinion (1)
- B.33: newsnational (1)
- B.34: newsrealestate (1)
- B.35: technology (1)
- B.36: newsother (1)
- B.37: causes-military-appreciation (1)
- B.38: newsvideos (1)
- C: finance (5916)
- C.1: financenews (1932)
- C.2: finance-real-estate (1303)
- C.3: finance-companies (718)
- C.4: markets (460)
- C.5: finance-career-education (314)
- C.6: finance-top-stocks (214)
- C.7: finance-video (166)
- C.8: personalfinance (139)
- C.9: finance-saving-investing (113)
- C.10: finance-retirement (103)
- C.11: finance-savemoney (93)
- C.12: finance-small-business (89)
- C.13: finance-taxes (60)
- C.14: finance-technology (45)
- C.15: finance-insurance (38)
- C.16: finance-credit (28)
- C.17: finance-career (23)
- C.18: finance-healthcare (14)
- C.19: finance-billstopay (13)
- C.20: finance-education (13)
- C.21: retirement (10)
- C.22: technologyinvesting (6)
- C.23: causes (6)
- C.24: finance-mutual-funds (5)
- C.25: career-news (3)
- C.26: finance-home-loans (1)
- C.27: finance-savingsrates (1)
- C.28: finance-insidetheticker (1)
- C.29: finance-homesandpropertysection (1)
- C.30: company-news (1)
- C.31: finance-auto-insurance (1)
- C.32: finance-startinvesting (1)
- C.33: spendingandborrowing (1)
- D: travel (4955)
- D.1: travelarticle (2624)
- D.2: travelnews (1812)
- D.3: traveltripideas (318)
- D.4: traveltips (86)
- D.5: video (66)
- D.6: travel-points-rewards (11)
- D.7: travel-adventure-travel (10)
- D.8: voices (8)
- D.9: ustravel (6)
- D.10: internationaltravel (5)
- D.11: travel-videos (3)
- D.12: holidays (2)
- D.13: causes (1)
- D.14: travel-accessible (1)
- D.15: newstrends (1)
- D.16: traveltrivia (1)
- E: lifestyle (4570)
- E.1: lifestylebuzz (1944)
- E.2: lifestylefamily (314)
- E.3: lifestyleroyals (273)
- E.4: lifestylehomeandgarden (239)
- E.5: lifestylepets (182)
- E.6: lifestyledidyouknow (128)
- E.7: lifestylehoroscope (126)
- E.8: lifestylesmartliving (125)
- E.9: lifestylepetsanimals (122)
- E.10: shop-holidays (111)
- E.11: voices (110)
- E.12: lifestylerelationships (91)
- E.13: lifestylemindandsoul (74)
- E.14: lifestylecareer (71)
- E.15: lifestylecelebstyle (71)
- E.16: lifestyleparenting (64)
- E.17: lifestylebeauty (62)
- E.18: shop-all (54)
- E.19: lifestylecleaningandorganizing (41)
- E.20: lifestylefashion (40)
- E.21: lifestylefamilyandrelationships (33)
- E.22: advice (30)
- E.23: lifestyleweddings (29)
- E.24: shop-apparel (26)
- E.25: lifestylelovesex (25)
- E.26: lifestylevideo (25)
- E.27: lifestyleshopping (22)
- E.28: lifestyle-news-feature (17)
- E.29: lifestylediy (17)
- E.30: causes-animals (17)
- E.31: lifestyledecor (16)
- E.32: causes (16)
- E.33: awardstyle (12)
- E.34: shop-home-goods (9)
- E.35: causes-green-living (8)
- E.36: lifestylemarriage (4)
- E.37: lifestylehoroscopefish (2)
- E.38: holidays (2)

(a)

Figure 35: Totals of articles by category and subcategory in the MIND-large training data

E.39: lifestyletravel (2)
E.40: shop-computers-electronics (2)
E.41: shop-books-movies-tv (2)
E.42: shop-toys (1)
E.43: relationships (1)
E.44: lifestyleanimals (1)
E.45: lifestyle-wedding (1)
E.46: pregnancyparenting (1)
E.47: lifestyle (1)
E.48: halloween (1)
E.49: lifestyleshoppinghomegarden (1)
E.50: travel (1)
E.51: lifestylewhatshot (1)
E.52: lifestylestyle (1)
E.53: lifestyledesign (1)
F: video (4569)
F.1: news (3316)
F.2: animals (328)
F.3: viral (269)
F.4: lifestyle (170)
F.5: science (163)
F.6: peopleandplaces (150)
F.7: popculture (122)
F.8: tunedin (21)
F.9: wonder (10)
F.10: sports (7)
F.11: downtime (6)
F.12: watch (3)
F.13: comedy (2)
F.14: healthandfitness (1)
F.15: foodanddrink (1)
G: foodanddrink (4418)
G.1: newstrends (2714)
G.2: recipes (541)
G.3: foodnews (318)
G.4: tipsandtricks (316)
G.5: restaurantsandnews (200)
G.6: videos (122)
G.7: beverages (100)
G.8: quickandeasy (56)
G.9: wines (30)
G.10: causes-food-insecurity (5)
G.11: cocktails (5)
G.12: foodrecipes (4)
G.13: cooking (3)
G.14: seasonal (2)
G.15: cookingschool (1)
G.16: foodtips (1)
H: weather (4255)
H.1: weathertopstories (4253)
H.2: photos (1)
H.3: weatherfullscreenmaps (1)
I: autos (3071)
I.1: autosnews (1703)
I.2: autosmotorcycles (394)
I.3: autosenthusiasts (334)
I.4: autosclassics (211)
I.5: autossports (80)
I.6: autossuvs (79)
I.7: autosbuying (56)
I.8: autostrucks (38)
I.9: autosluxury (31)
I.10: autosownership (30)
I.11: autosresearchguides (24)
I.12: autossema (23)
I.13: autosresearch (18)
I.14: autosvideonew (14)
I.15: autoshybrids (8)
I.16: autostokyo (7)
I.17: autospassenger (6)
I.18: autosvideos (5)
I.19: autosreview (3)
I.20: autoscartech (2)
I.21: autosconvertibles (1)
I.22: autoscompact (1)
I.23: autosvans (1)
I.24: autosmidsize (1)
I.25: autoslosangeles (1)
J: health (2929)
J.1: medical (954)
J.2: health-news (576)
J.3: wellness (573)
J.4: nutrition (281)
J.5: fitness (171)
J.6: weightloss (168)
J.7: voices (73)
J.8: healthnews (46)
J.9: video (16)
J.10: causes (15)
J.11: pregnancyparenting (13)
J.12: mentalhealth (11)
J.13: ads-lung-health (11)
J.14: weight-loss (7)
J.15: cardio (2)
J.16: strength (2)
J.17: healthagingwell (2)
J.18: familyhealth (2)
J.19: recipes (2)
J.20: smartliving (1)
J.21: mindandbody (1)
J.22: healthyliving (1)
J.23: health-cancer (1)
K: tv (1323)
K.1: tvnews (777)
K.2: tv-celebrity (375)
K.3: tv-gallery (83)
K.4: tvvideos (44)
K.5: humor (18)
K.6: tv-recaps (18)
K.7: tv-golden-globes (4)
K.8: tv-reviews (2)
K.9: topnews (1)
K.10: tv-golden-globes-video (1)
L: music (1263)
L.1: musicnews (874)
L.2: music-celebrity (225)
L.3: music-gallery (53)
L.4: cma-awards (38)
L.5: music-awards (31)
L.6: musicvideos (20)
L.7: ads-latingrammys (10)
L.8: music-reviews (6)
L.9: music-grammys (4)
L.10: topnews (1)
L.11: humor (1)
M: entertainment (837)
M.1: gaming (256)
M.2: entertainment-celebrity (168)
M.3: celebrity (151)
M.4: entertainment-books (147)
M.5: news (35)
M.6: video (31)
M.7: awards (31)
M.8: humor (10)
M.9: celebritynews (2)
M.10: hollywood (1)
M.11: games (1)
M.12: tv (1)
M.13: entertainmentmusic (1)
M.14: entertainmenttv (1)
M.15: celebhub (1)
N: movies (815)
N.1: movienews (406)
N.2: movies-celebrity (224)
N.3: movievideo (89)
N.4: movies-gallery (72)
N.5: movies-awards (12)
N.6: movies-oscars (7)
N.7: reviews (5)
O: kids (104)
O.1: video (24)
O.2: people-places (24)
O.3: fun (22)
O.4: science (19)
O.5: animals (10)
O.6: sports (5)
P: middleeast (2)
P.1: middleeast-top-stories (2)
Q: games (1)
Q.1: games-news (1)
R: northamerica (1)
R.1: northamerica-video (1)

(b)

Figure 35: Totals of articles by category and subcategory in the MIND-large training data

Glossary

Bidirectional Encoder Representations from Transformers A pretrained transformer-based machine learning model for natural language processing used for text embedding in this thesis. 11, 50

Breitbart A right-wing news website in the USA with a subbranch for news in Europe/London. 19, 20, 29, 30, 39, 40, 43

Convolutional Neural Network A NN whose name comes from the mathematical operation of convolution performed by some of its layers. CNNs provide excellent results in various fields of machine-learning such as NLP. 10, 50

EUMIG article corpus A corpus of 5127 news articles collected on the devisive topic of "Migration in the EU" for analyzing purposes in the scope of this thesis. v, 16, 20, 21, 22, 23, 24, 25, 27, 30, 37, 38, 39, 40, 41, 42, 43

HyperText Markup Language The standard language for content that is intended to be displayed in a web browser. 19, 50

Media Bias/Fact Check A website that rates the bias, factuality, and credibility of media sources through an approach that combines objective measures and subjective analysis[22]. 19, 39

Named Entity Linking A subtask in NLP, which aims to find the corresponding entity in a knowledge graph for a named entity mentioned in a text. 7, 50

Named Entity Recognition A subtask in NLP, which aims to locate and classify tokens in text that represent named entities, e.g. persons or locations. 7, 50

Natural Language Processing A subfield in computer science referring to automatic computational processing of the human language. 5, 50

Neural Network A data processing system consisting of a large number of simple, highly interconnected processing elements in an architecture inspired by the structure of the cerebral cortex portion of the brain [53]. Neural networks attempt to learn patterns in large amounts of data as a target function. To that end, the data is processed through an input layer, possibly several hidden layers, and an output layer. A layer consists of neurons, operational nodes that perform a mathematical operation on the neuron's inputs and pass the result on to connected neurons.. 7, 50

[22]For their methodology see: `https://mediabiasfactcheck.com/methodology/` (Last accessed: 2021-07-26)

News reception News reception is a research area that deals with the behavioral patterns of news consumption. The term reception stems from latin *recipere*, in english to acquire or to receive. 16, 25, 26, 27, 28, 29, 30, 40, 44, 45

Reuters An international news agency. 19, 20, 30, 42

The Canary A left-wing news website in the United Kingdom. 19, 20, 29, 30, 41

TransE A machine learning model that produces state-of-the-art embeddings for multi-relational knowledge graphs. 18, 25, 33, 34, 39, 40

WikiData A public, community-maintained knowledge graph, that contains references to facts instead of facts itself to provide diverse knowledge about a given entity. 10, 18, 22, 25, 31, 33, 34, 39

Acronyms

AI Artificial Intelligence. 7

AUC Area Under the ROC Curve. 13, 18

BERT Bidirectional Encoder Representations from Transformers. 11, 18, 31, 32, 33, 34

CBF content-based filtering. 4

CF collaborative filtering. 4

CNN Convolutional Neural Network. 10

DKN deep knowledge-aware network. 10

DL Deep Learning. 7

DNN Deep Neural Network. 7, 8

DSA Digital Service Act. 2

EU European Union. 3, 16, 18, 20, 23, 32, 40, 48

GDPR General Data Protection Regulation. 2

HTML HyperText Markup Language. 19, 20

KRED Knowledge aware Representation Enhancement for Documents. 10, 11, 14, 16, 17, 18, 21, 22, 25, 27, 28, 30, 31, 32, 33, 34, 37, 38, 39, 40

MIND Microsoft News Dataset. 16, 18, 21, 24, 25, 31, 34, 36, 37, 38, 39, 40, 46, 47

ML Machine Learning. 7

MSN Microsoft News. 34, 35, 36

NDCG Normalized Discounted Cumulative Gain. 13, 14, 18

NEL Named Entity Linking. 7

NER Named Entity Recognition. 7, 22, 23

NLP Natural Language Processing. 5, 6, 7, 11, 48

NN Neural Network. 7, 48

NRS News Recommendation System. 1, 2, 3, 4, 5, 10, 11, 12, 14, 15, 16, 17, 18, 19, 25, 27, 28, 30, 31, 32, 33, 34, 37, 38, 39, 40

REL Radboud Entity Linker. 21, 22, 23

ROC Receiver Operating Characteristic. 13, 50

URL Uniform Resource Locator. 19, 20, 24

References

[1] Baeza-Yates, R. Bias on the web. *Communications of the ACM 61*, 6 (May 2018), 54–61.

[2] Baly, R., Karadzhov, G., An, J., Kwak, H., Dinkov, Y., Ali, A., Glass, J., and Nakov, P. What Was Written vs. Who Read It: News Media Profiling Using Text Analysis and Social Media Context. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (Online, 2020), Association for Computational Linguistics, pp. 3364–3374.

[3] Berry, M., Garcia-Blanco, I., Moore, K., Morani, M., Gross, B., Askanius, T., and Linné, T. Press Coverage of the Refugee and Migrant Crisis in the EU: A Content Analysis of Five European Countries. Tech. rep., UNHCR, 10 2015.

[4] Boratto, L., and Marras, M. Advances in Bias-Aware Recommendation on the Web. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining* (New York, NY, USA, 2021), WSDM '21, Association for Computing Machinery, p. 1147–1149. tutorial slides: `https://www.slideshare.net/MirkoMarras/tutorial-on-advances-in-biasaware-recommendation-on-the-web-wsdm-2021` last accessed: 2021-08-07.

[5] Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., and Yakhnenko, O. Translating embeddings for modeling multi-relational data. *Advances in neural information processing systems 26* (2013).

[6] Bourli, S., and Pitoura, E. Bias in Knowledge Graph Embeddings. In *2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (2020), IEEE, pp. 6–10.

[7] Bradley, A. P. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition 30*, 7 (1997), 1145–1159.

[8] Carlyle, T. *Heroes and hero-worship*, vol. 12. Chapman and Hall, 1869.

[9] Chen, J., Dong, H., Wang, X., Feng, F., Wang, M., and He, X. Bias and debias in recommender system: A survey and future directions. *arXiv preprint arXiv:2010.03240* (2020).

[10] Chern Tan, Y., and Celis, L. E. Assessing Social and Intersectional Biases in Contextualized Word Representations. *arXiv e-prints* (2019), arXiv–1911.

[11] Commission, E. Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC. COM/2020/825 final, Dec 2020.

[12] Community, R. SIGIR 2020 Main Conference Program. `www.sigir.org/sigir2020/schedule/`, 2020.

[13] Demartini, G. Implicit bias in crowdsourced knowledge graphs. In *Companion Proceedings of The 2019 World Wide Web Conference* (2019), pp. 624–630.

[14] Devlin, J., Chang, M., Lee, K., and Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *CoRR abs/1810.04805* (2018).

[15] Elejalde, E., Ferres, L., and Herder, E. On the nature of real and perceived bias in the mainstream media. *PLOS ONE 13* (Mar. 2018).

[16] Eskens, S. The personal information sphere: An integral approach to privacy and related information and communication rights. *Journal of the Association for Information Science and Technology 71*, 9 (2020), 1116–1128.

[17] Fisher, J., Palfrey, D., Christodoulopoulos, C., and Mittal, A. Measuring social bias in knowledge graph embeddings. *arXiv preprint arXiv:1912.02761* (2019).

[18] Gao, Y., Liang, J., Han, B., Yakout, M., and Mohamed, A. Building a large-scale, accurate and fresh knowledge graph. *KDD-2018, Tutorial 39* (2018), 1939–1374.

[19] Garimella, A., Amarnath, A., and Rathlavath, K. K. He is very intelligent, she is very beautiful? On Mitigating Social Biases in Language Modelling and Generation. *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021* (2021).

[20] Ge, M., Delgado-Battenfeld, C., and Jannach, D. Beyond Accuracy: Evaluating Recommender Systems by Coverage and Serendipity. In *Proceedings of the Fourth ACM Conference on Recommender Systems* (New York, NY, USA, 2010), RecSys '10, Association for Computing Machinery, p. 257–260.

[21] Guo, Q., Zhuang, F., Qin, C., Zhu, H., Xie, X., Xiong, H., and He, Q. A Survey on Knowledge Graph-Based Recommender Systems. *IEEE Transactions on Knowledge and Data Engineering* (2020).

[22] Helberger, N. The political power of platforms: How current attempts to regulate misinformation amplify opinion power. *Digital Journalism 8*, 6 (2020), 842–854.

[23] Hopmans, T. A recommendation system for blogs: Setting up the prerequisites (part 1). https://www.themarketingtechnologist.co/building-a-recommendation-engine-for-geek-setting-up-the-prerequisites-13/, Apr 2016. Image accessed via researchgate: https://www.researchgate.net/figure/Content-based-filtering-vs-Collaborative-filtering-Source_fig5_323726564, 2021-07-28.

[24] Huang, J., and Ling, C. Using AUC and accuracy in evaluating learning algorithms. *IEEE Transactions on Knowledge and Data Engineering 17*, 3 (2005), 299–310.

[25] Hulst, J., Hasibi, F., Dercksen, K., Balog, K., and de Vries, A. REL: An Entity Linker Standing on the Shoulders of Giants. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval* (07 2020), pp. 2197–2200.

[26] Järvelin, K., and Kekäläinen, J. Cumulated Gain-Based Evaluation of IR Techniques. *ACM Trans. Inf. Syst. 20*, 4 (Oct. 2002), 422–446.

[27] Kaminskas, M., and Bridge, D. Diversity, Serendipity, Novelty, and Coverage: A Survey and Empirical Analysis of Beyond-Accuracy Objectives in Recommender Systems. *ACM Trans. Interact. Intell. Syst. 7*, 1 (Dec. 2016).

[28] Kaneko, M., and Bollegala, D. Debiasing pre-trained contextualised embeddings. *arXiv preprint arXiv:2101.09523* (2021).

[29] Kurita, K., Vyas, N., Pareek, A., Black, A. W., and Tsvetkov, Y. Measuring bias in contextualized word representations. *arXiv preprint arXiv:1906.07337* (2019).

[30] Le, P., and Titov, I. Improving entity linking by modeling latent relations between mentions. *arXiv preprint arXiv:1804.10637* (2018).

[31] Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P. N., Hellmann, S., Morsey, M., Van Kleef, P., Auer, S., et al. Dbpedia–a large-scale, multilingual knowledge base extracted from wikipedia. *Semantic web 6*, 2 (2015), 167–195.

[32] Li, F.-F., Krishna, R., and Xu, D. Lecture notes in Convolutional Neural Networks for Visual Recognition, April 2021. Lecture 4, p. 45 (Last accessed: 2021-10-06).

[33] Liang, P. P., Li, I. M., Zheng, E., Lim, Y. C., Salakhutdinov, R., and Morency, L.-P. Towards debiasing sentence representations. *arXiv preprint arXiv:2007.08100* (2020).

[34] Liu, D., Lian, J., Wang, S., Qiao, Y., Chen, J.-H., Sun, G., and Xie, X. KRED: Knowledge-Aware Document Representation for News Recommendations. *Fourteenth ACM Conference on Recommender Systems* (Sept. 2020), 200–209. arXiv: 1910.11494.

[35] Loecherbach, F., Welbers, K., Moeller, J., Trilling, D., and Van Atteveldt, W. Is this a click towards diversity? Explaining when and why news users make diverse choices. In *13th ACM Web Science Conference 2021* (2021), pp. 282–290.

[36] Mansoury, M., Abdollahpouri, H., Pechenizkiy, M., Mobasher, B., and Burke, R. Feedback loop and bias amplification in recommender systems. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management* (2020), pp. 2145–2148.

[37] Milano, S., Taddeo, M., and Floridi, L. Recommender systems and their ethical challenges. *AI & SOCIETY* (2020), 957–967.

[38] Möller, J., Trilling, D., Helberger, N., and van Es, B. Do not blame it on the algorithm: an empirical assessment of multiple recommender systems and their impact on content diversity. *Information, Communication & Society 21*, 7 (2018), 959–977.

[39] Newman, N., Fletcher, R., Schulz, A., Andi, S., and Kleis Nielsen, R. Reuters Institute Digital News Report 2020. Tech. rep., Reuters Institution, 2020.

[40] Newman, N., Fletcher, R., Schulz, A., Andi, S., Robertson, C. T., and Kleis Nielsen, R. Reuters Institute Digital News Report 2021. Tech. rep., Reuters Institution, 2021.

[41] on Information Retrieval, S. I. G. RecSys 2020 Conference Program. www.recsys.acm.org/recsys20/program/, 2020.

[42] Pariser, E. *The filter bubble: What the Internet is hiding from you.* Penguin UK, 2011.

[43] Parliament, E. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance). OJ L 119, 4.5.2016, p. 1–88, May 2016.

[44] PATANKAR, A., BOSE, J., AND KHANNA, H. A Bias Aware News Recommendation System. In *2019 IEEE 13th International Conference on Semantic Computing (ICSC)* (2019), pp. 232–238.

[45] RAZA, S., AND DING, C. News Recommender System: A review of recent progress, challenges, and opportunities. *arXiv:2009.04964 [cs]* (Feb. 2021). arXiv: 2009.04964.

[46] RIBEIRO, M. H., OTTONI, R., WEST, R., ALMEIDA, V. A. F., AND MEIRA, W. Auditing Radicalization Pathways on YouTube. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (New York, NY, USA, 2020), FAT* '20, Association for Computing Machinery, p. 131–141.

[47] SCHWETER, S., AND AKBIK, A. FLERT: Document-Level Features for Named Entity Recognition, 2020.

[48] SHAIK, Z., ILIEVSKI, F., AND MORSTATTER, F. Analyzing Race and Country of Citizenship Bias in Wikidata. *arXiv preprint arXiv:2108.05412* (2021).

[49] SILVEIRA, T., ZHANG, M., LIN, X., LIU, Y., AND MA, S. How good your recommender system is? A survey on evaluations in recommendation. *International Journal of Machine Learning and Cybernetics 10*, 5 (2019), 813–831.

[50] SINGHAL, A. Introducing the Knowledge Graph: things not strings. `https://blog.google/products/search/introducing-knowledge-graph-things-not/`, 05 2012.

[51] SUCHANEK, F. M., KASNECI, G., AND WEIKUM, G. Yago: a core of semantic knowledge. In *Proceedings of the 16th international conference on World Wide Web* (2007), pp. 697–706.

[52] TWINGLY. Social Data from Millions of Sources. `https://www.twingly.com/news-data`. Accessed: 2021-10-05.

[53] UHRIG, R. Introduction to artificial neural networks. In *Proceedings of IECON '95 - 21st Annual Conference on IEEE Industrial Electronics* (1995), vol. 1, pp. 33–37 vol.1.

[54] VRANDEČIĆ, D., AND KRÖTZSCH, M. Wikidata: a free collaborative knowledgebase. *Communications of the ACM 57*, 10 (2014), 78–85.

[55] WANG, H., ZHANG, F., WANG, J., ZHAO, M., LI, W., XIE, X., AND GUO, M. Exploring High-Order User Preference on the Knowledge Graph for Recommender Systems. *ACM Transactions on Information Systems 37*, 3 (July 2019), 1–26.

[56] Wang, H., Zhang, F., Xie, X., and Guo, M. DKN: Deep Knowledge-Aware Network for News Recommendation. In *Proceedings of the 2018 World Wide Web Conference* (Republic and Canton of Geneva, CHE, 2018), WWW '18, International World Wide Web Conferences Steering Committee, p. 1835–1844.

[57] Wang, Y., Wang, L., Li, Y., He, D., Chen, W., and Liu, T.-Y. A theoretical analysis of NDCG ranking measures. In *Proceedings of the 26th annual conference on learning theory (COLT 2013)* (2013), vol. 8, Citeseer, p. 6.

[58] Wu, C., Wu, F., Wang, X., Huang, Y., and Xie, X. Fairness-aware News Recommendation with Decomposed Adversarial Learning. In *Proceedings of the AAAI Conference on Artificial Intelligence* (2021), pp. 4462–4469.

[59] Wu, F., Qiao, Y., Chen, J.-H., Wu, C., Qi, T., Lian, J., Liu, D., Xie, X., Gao, J., Wu, W., and Zhou, M. MIND: A Large-scale Dataset for News Recommendation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (2020), Association for Computational Linguistics, pp. 3597–3606.

[60] Wu, L., Petroni, F., Josifoski, M., Riedel, S., and Zettlemoyer, L. Scalable zero-shot entity linking with dense entity retrieval. *arXiv preprint arXiv:1911.03814* (2019).

[61] Yi, J., Wu, F., Wu, C., Li, Q., Sun, G., and Xie, X. DebiasedRec: Bias-aware User Modeling and Click Prediction for Personalized News Recommendation. *arXiv preprint arXiv:2104.07360* (2021).

[62] Zhao, J., Mukherjee, S., Hosseini, S., Chang, K.-W., and Awadallah, A. H. Gender bias in multilingual embeddings and cross-lingual transfer. *arXiv preprint arXiv:2005.00699* (2020).

[63] Zhu, Y., Kiros, R., Zemel, R., Salakhutdinov, R., Urtasun, R., Torralba, A., and Fidler, S. Aligning books and movies: Towards story-like visual explanations by watching movies and reading books. In *Proceedings of the IEEE international conference on computer vision* (2015), pp. 19–27.

[64] Zuiderveen Borgesius, F., Trilling, D., Möller, J., Bodó, B., De Vreese, C. H., and Helberger, N. Should we worry about filter bubbles? *Internet Policy Review. Journal on Internet Regulation 5*, 1 (2016).

Assertion

I truthfully declare that I have written this thesis independently, that I have completely and accurately indicated all aids used, and that I have marked everything that has been taken from the work of others, either unchanged or with modifications, and that I have complied with the KIT Statutes for Safeguarding Good Scientific Practice in the currently valid version.

Ich versichere wahrheitsgemäß, die Arbeit selbstständig verfasst, alle benutzten Hilfsmittel vollständig und genau angegeben und alles kenntlich gemacht zu haben, was aus Arbeiten anderer unverändert oder mit Abänderungen entnommen wurde sowie die Satzung des KIT zur Sicherung guter wissenschaftlicher Praxis in der jeweils gültigen Fassung beachtet zu haben.

Karlsruhe, October 15, 2021 Hannah Greven

[redacted]