News Diversity and Recommendation Systems: Setting the Interdisciplinary Scene

Glen Joris, Camiel Colruyt, Judith Vermeulen, Stefaan Vercoutere, Frederik De Grove, Kristin Van Damme, Orphée De Clercq, Cynthia Van Hee, Lieven De Marez, Veronique Hoste, et al.

▶ To cite this version:

Glen Joris, Camiel Colruyt, Judith Vermeulen, Stefaan Vercoutere, Frederik De Grove, et al.. News Diversity and Recommendation Systems: Setting the Interdisciplinary Scene. 14th IFIP International Summer School on Privacy and Identity Management (Privacy and Identity), Aug 2019, Windisch, Switzerland. pp.90-105, 10.1007/978-3-030-42504-3_7 . hal-03378981

HAL Id: hal-03378981
https://inria.hal.science/hal-03378981

Submitted on 14 Oct 2021

News diversity and recommendation systems: setting the interdisciplinary scene

Glen Joris[1], Camiel Colruyt[2], Judith Vermeulen[3], Stefaan Vercoutere[4], Frederik De Grove[1], Kristin Van Damme[1], Orphée De Clercq[2], Cynthia Van Hee[2], Lieven De Marez[1], Veronique Hoste[2], Eva Lievens[3], Toon De Pessemier[4], and Luc Martens[4]

[1] imec-mict-UGent and Center for Journalism Studies, Department of Communication Sciences, Ghent University, Belgium
[2] Language and Translation Technology Team, Department of Translation, Interpreting and Communication, Ghent University, Belgium
[3] Law & Technology, Department of Interdisciplinary Law, Private Law and Business Law, Ghent University, Belgium
[4] imec-WAVES-UGent, Department of Information Technology, Ghent University, Belgium

Abstract. Concerns about selective exposure and filter bubbles in the digital news environment trigger questions regarding how news recommender systems can become more citizen-oriented and facilitate – rather than limit – normative aims of journalism. Accordingly, this chapter presents building blocks for the construction of such a news algorithm as they are being developed by the Ghent University interdisciplinary research project #NewsDNA, of which the primary aim is to actually build, evaluate and test a diversity-enhancing news recommender. As such, the deployment of artificial intelligence could support the media in providing people with information and stimulating public debate, rather than undermine their role in that respect. To do so, it combines insights from computer sciences (news recommender systems), law (right to receive information), communication sciences (conceptualisations of news diversity), and computational linguistics (automated content extraction from text). To gather feedback from scholars of different backgrounds, this research has been presented and discussed during the 2019 IFIP summer school workshop on 'co-designing a personalised news diversity algorithmic model based on news consumers' agency and fine-grained content modelling'. This contribution also reflects the results of that dialogue.

Keywords: News personalisation · Algorithms · News recommender systems · Right to receive diverse information · News diversity · News content extraction · #NewsDNA

1 Introduction

In recent years, online news organisations – both web-editions of traditional news outlets and digital-only news sites – increasingly explore how recommender systems can be used to provide consumers with a tailor-made news offer. The New York Times, for example, uses a mix of editorial curation and algorithms to compose a newsletter tailored to each recipient [1]. In Belgium too, De Standaard and Het Nieuwsblad recently introduced a personal page that collects articles based on the reader's selected topics [2]. In The Netherlands, online newspapers nu.nl and Algemeen Dagblad both invested heavily in personalised news notifications [3]. Hence, news organisations are increasingly exploring implicit and explicit algorithmic news personalisation [4, 5], similarly to how companies such as Netflix and Amazon individualise content.

Personalisation typically relies on automated decision-making and recommender systems. However, in contrast to what people might think [6], these systems are not neutral as they primarily apply a commercial logic. More specifically, they produce recommendations based on the calculated relevance of news items vis-à-vis individual news consumers (for example taking into account selected fields of interests or past consumption patterns). Such practices contrast with the role of the media as 'a marketplace of ideas' in which citizens are confronted with a diverse array of ideas [7]. Although empirical research currently supports a more nuanced view [8], news recommender systems are argued to be a potential threat to an informed citizenry and the democratic processes between media, politics and audiences [9]. With these concerns in mind, several scholars have raised questions regarding how news recommender systems can be built in a more citizen-oriented way by maintaining the normative aims of journalism [10].

The current chapter presents research conducted by the interdisciplinary research project #NewsDNA at Ghent University (Belgium), which seeks to provide a possible answer in that regard. More specifically, it outlines building blocks for the construction of a recommender system that uses news diversity as a key driver for personalised news recommendations. As such, the deployment of artificial intelligence could support the media in providing people with information and stimulating public debate, rather than undermine their role in that respect. To gather feedback concerning this framework from scholars of different backgrounds, this research has been presented during the 2019 IFIP summer school workshop on 'co-designing a personalised news diversity algorithmic model based on news consumers' agency and fine-grained content modelling'. The results of this exercise are also included in this contribution.

As the current research builds on insights from multiple disciplines, the remainder of this paper is – as was the workshop – organised per discipline. First, we present a state-of-the-art overview of the most commonly used methods to design news recommender systems within computer sciences. Second, we explore the existence of a legal ground for receiving diverse news. Third, we expound on the conceptual meaning of news diversity by building on literature in communication sciences. Finally, we discuss the computational feasibility of news content extraction, provided by computational linguistics, to provide data to the aspired diversity-promoting news recommender system. Each part is followed by related questions we presented the workshop participants with as well as their answers in that regard.

2 News recommendations systems today

In this section, we provide an overview of current news recommendation systems, from a computer science perspective. We present the two dominant approaches, being collaborative and content-based filtering, and outline the obstacles related to the development of a diversity-enhancing news recommender.

One of the most commonly used methods in the field of recommendation systems is *collaborative filtering*. Collaborative filtering assumes that people who had similar interests in the past are likely to have similar interests in the future [11]. As such, relevant news articles are predicted based on news articles read by so-called ‘neighbours’, other users who have historically had similar taste in news [12]. In essence, it is very similar to the concept of ‘word of mouth’: we often consult with our peers when gathering opinions about certain activities or decisions (e.g. interesting movies, tasty drinks). Especially for news, peer recommendations are still perceived as valuable [13].

Collaborative filtering methods have two major drawbacks when recommending news stories. First, news is quick and volatile, which exacerbates the first-rater problem [14, 15]: a new story cannot be recommended to users unless other users have read it before. This becomes problematic when trying to present the latest information in a timely manner, as it is not uncommon for collaborative filtering based methods to take several hours before sufficient clicks have been collected and a new item can be recommended. Generally, as an item gains more clicks, the system becomes more confident in its ability to recommend it. Hence, older and popular items dominate the recommendation process, which is not desirable for news recommendations. Second, there is the sparsity problem [16], which occurs when there is insufficient overlap between the consumption patterns of users. As the relevance of news stories sharply decreases over time, it is not unreasonable to assume little overlap between new and old users.

A second approach is *content-based filtering*, which does not have these shortcomings, and consequently is often used for news recommendations [15, 17]. Content-based systems use the news articles themselves to recommend similar news, both the content and its metadata. For example, the system looks at the topic of the news, the keywords or the broader classification (e.g. sports or domestic affairs), the author, word count, etc. This means that in contrast to collaborative filters, content-based systems treat recommendation as a single user classification problem.

However, this method also suffers from certain drawbacks. First, over-specialisation: content-based systems cannot provide recommendations outside the scope of what the user has already shown interest in. Within journalism, this is the trigger for concerns of filter bubbles and news personalisation [9]. Second, the performance of the system heavily depends on the quality of the content descriptions. In domains where the items consist of music or video, the extraction of a useful representation of the content can be very challenging. In journalism as well, news articles often do not have sufficient metadata, nor are metadata compatible across different news companies. Section 5 illustrates a few content dimensions that can be used in recommendation.

What both these techniques have in common is that they are based on similarity, either between users or items. The risk of such a recommendation strategy is that users are more likely to be exposed to a narrowing segment of popular items, as the focus lies on maximising the overlap between users’ behaviour. As such, recommender systems today strive for news personalisation, which in fact contrasts with the aspired goal of a citizen-aware recommender system. This risk is compounded by the focus on metrics such as accuracy. Often, the performance of a recommender is solely measured in terms of its re-constructional capabilities (i.e. how precise the system is in predicting already consumed articles). All differences between the original and predicted user history are seen as losses in performance. When the lack of diversity is addressed, it is typically done as an adjunct to the standard procedures and through rudimentary means [18, 19].

Increasing diversity and novelty is only considered if it can be done without significantly compromising query similarity, and this application remains limited to aimlessly broadening of the coverage.

Having set the scene, the participants were asked whether they had ever encountered news personalisation in their daily lives and to share their perception of such practices in general. As regards the first question, one attendee indicated to have experienced personalisation at a news site whereby items presumed relevant to a particular consumer were highlighted. However, all news available remained accessible for all users in the same order. Another noted that as social media feeds are being personalised, the news you view on these platforms is so too. In response to the second question, concerns were raised as regards a presumed lack of awareness and transparency in relation to the existence of algorithmic selection processes as well as concerning the logic behind them. It was added that news consumers should be properly informed. One participant furthermore indicated to "dislike the feeling of being steered".

3 A fundamental right to diverse information

The importance of an easily available diverse news offer has been recognised in several recent policy documents. More specifically, it was argued that it has "the potential to make democratic processes more participatory and inclusive" and to foster public debate – which may ultimately secure democracy –, and could even "uncover, counterbalance, and dilute disinformation" [20, 21]. At EU level, therefore, "empower[ing] users with tools enabling a customised and interactive online experience so as to facilitate content discovery and access to different news sources representing alternative viewpoints" was set as a goal [21–23]. Interestingly, the EU Commission's independent High Level Expert Group on Fake news and Online Disinformation, when defining '[a]ctions in support of press freedom and pluralism' in the final report concerning their approach on disinformation, stated, amongst others, that public authorities must ensure the "protection of [a] basic right[…] to […] *diverse* information" (emphasis added) [22]. In that context, the question arises whether, and to what extent, the right to *freedom of expression and information*, laid down in both Article 10 of the European Convention on Human Rights ('ECHR') [24] – under which positive obligations[1] may arise [26] – and Article 11 of the Charter of Fundamental Rights of the European Union ('Charter') [27], indeed includes *a right to receive diverse information*. Its existence, including a corresponding responsibility for authorities to take affirmative action to ensure its effective exercise [28], would enable citizens to *force* policymakers to adopt measures guaranteeing them access, potentially offline as well as online, to a diversity of information.

In its first paragraph, Article 10 ECHR puts forward that:

> "*Everyone* has the right to freedom of expression. This right shall include *freedom* to hold opinions and *to receive* and impart *information and ideas* without interference by public authority and regardless of frontiers […]." (emphasis added)

[1] Negative obligations require States not to interfere in the exercise of rights, while positive obligations entail a duty to take the necessary measures to safeguard a right, or, more specically, to adopt reasonable and suitable measures to protect the rights of individuals" in [25].

On numerous occasions, the European Court of Human Rights ('ECtHR') has interpreted the 'freedom to receive information and ideas'. Indeed, already in the 1979 case of Sunday Times v. the UK, the Court stated that "[n]ot only do the [mass] media have the task of imparting […] information and ideas [concerning matters of public interest]; *the public also has a right to receive them*" (emphasis added) [29]. In the Informationsverein Lentia and Others v. Austria judgment from 1993, it added that "[s]uch an undertaking cannot be successfully accomplished unless it is grounded in *the principle of pluralism, of which the State is the ultimate guarantor*" (emphasis added) and that "[t]his observation is especially valid in relation to audio-visual media, whose programmes are often broadcast very widely" [30]. On 8 July 1999, in the context of its decision in a number of cases against Turkey, all concerning the criminal convictions of the applicants in view of their involvement in the spread of separatist or pro-Kurdish propaganda [31], the ECtHR explicitly referred to "the public's right to be informed of a different perspective" and considered that the domestic authorities failed to sufficiently respect their negative obligation in that regard [32]. It furthermore concretised its by then settled Sunday Times case-law referred to above, by finding that "[i]t is […] incumbent on the press to impart information and ideas on political issues, *including divisive ones*" (emphasis added), whilst the public is entitled to receive them [33]. In the Khurshid Mustafa and Tarzibachi v. Sweden case of 2008, which concerned (a prohibition of) the reception of information by means of a satellite dish, it was held in very clear terms that, "[i]n addition to the primarily negative undertaking of a State to abstain from interferences in Convention guarantees", "the genuine and effective exercise of freedom of expression under *Article 10 may require positive measures of protection, even in the sphere of relations between individuals*" (emphasis added) [34]. In the 2009 Times Newspapers LTD (Nos. 1 and 2) v. the United Kingdom case, it was considered that "the Internet plays an important role in enhancing the public's access to news and facilitating the dissemination of information in general" [35]. Late 2009, in the Manole and Others v. Moldova judgment, the Strasbourg Court ruled that "the State [must] ensure […] that the public has access through television and radio to impartial and accurate information and *a range of opinion and comment, reflecting inter alia the diversity of political outlook within the country*" (emphasis added) [36]. Finally, in 2012, in Centro Europa 7 S.R.L. and Di Stefano v. Italy, it was clarified that, considering the sensitive nature of the audio-visual media sector, member States have a positive obligation to "put in place *an appropriate legislative and administrative framework to guarantee effective pluralism*" (emphasis added) [37].

Article 11 of the Charter provides that:

> "1. *Everyone has the right to* freedom of expression. This right shall include *freedom* to hold opinions and *to receive* and impart *information and ideas* without interference by public authority and regardless of frontiers.
>
> 2. The freedom and *pluralism of the media shall be respected*." (emphasis added)

Article 52(3) of the Charter stipulates that "[i]n so far [the] Charter contains rights which correspond to rights guaranteed by the Convention for the Protection of Human Rights and Fundamental Freedoms, the meaning and scope of those rights shall be the same as those laid down in the European Convention on Human Rights". Clearly, such is the case for Article 11(1) [38, 39]. Moreover, Article 51(1) EU Charter puts forward that the provisions of the Charter are addressed to the institutions, bodies, offices and agencies of the Union and to its member states only when they are implementing Union Law. They should therefore "respect the rights, observe the principles and promote the application thereof in accordance with their respective powers and respecting the limits of the powers of the Union as conferred on it in the Treaties" [27 art 51(1)]. Leaving aside the discussions concerning the exact scope of application of EU fundamental rights in respect of actions of member states on the basis of Article 51(1) of the Charter [40], several scholars have argued that this provision confers on the Union an indirect power to adopt rules or measures protecting fundamental rights in the course of exercising its specific competences under the Treaties [40]. However, such a power would not allow the Union to take action if the protection of fundamental rights were to be the only or primary aim thereof [40]. It appears, for example, that by means of Article 15 as well as recitals 48 and 55 of the Audiovisual Media Directive [41], the EU legislator has sought to safeguard the right to receive information and to promote pluralism of the media by ensuring diversity in the production and programming of news in the EU, and therefore to respect the principles recognised by both paragraphs of Article 11 of the Charter [40]. In its 2013 Opinion in Sky Österreich, the Court of Justice of the European Union ('CJEU'), confirmed that the EU legislature was indeed 'entitled' – though not, on the basis of positive obligations, 'required' – to do so and to take measures to ensure public access to a diversity of information [42].

In conclusion, the ECtHR has clearly recognised a right of the public to be informed about different viewpoints concerning matters of general importance. The State is ultimately responsible for the effective exercise thereof, and this indisputably within the context of the audio-visual media sector. While the ECtHR so far has not recognised such a duty vis-à-vis States in relation to the public's right to receive information and ideas in the online environment, it indeed very well could, given its acknowledgment of the importance of the Internet in enhancing access to news and facilitating the dissemination of information [43]. The CJEU has also stressed the importance of media pluralism and diversity of information available to the public. Whereas the CJEU does not (yet) consider the Union – in view of its nature and competences [40] – to be directly responsible for taking positive action to that end, it found that the latter certainly *may* do so when exercising its attributed competences.

Considering the potential impact a right to diverse information could have in our contemporary society, we presented the participants with the following questions: one, do we need such a right, and two, should the government play a role in ensuring that citizens can access diverse information? The first question was collectively answered in the affirmative. One attendee in particular argued that one should have access to diverse information as it is enables him or her to make informed decisions. Thoughts in relation to the second question were, on the other hand, more varied. One group of participants noted that in countries where confidence in publicly-funded news is high, the government could also be trusted to guarantee diverse exposure to information. Others, however, stated that such involvement could very easily go wrong. They considered, more specifically, that it may lead to a situation in which people would only be shown content the authorities want them to see.

4 Unravelling news diversity

As argued above, it becomes clear that today's news recommender systems do not take into account news diversity, even though the idea of receiving diverse information is an important prerequisite for maintaining a democratic society. This neglect also contrasts with the academic field of communication sciences in which news diversity has a long tradition in helping to understand and evaluate the role of news media in the public sphere [44, 45]. In fact, most research on news diversity date back to the arrival of audio-visual and digital media such as television and the web (i.e., 1995-2005, see e.g. [46–48]). Despite the existence of a significant body of literature around the concept of news diversity, however, communication scholars are still struggling with the question of what it means, and how it should be measured [48, 49]. Consequently, a wide range of diversity dimensions, assessments and assumptions are currently used to study news diversity [50].

The broad and ambiguous use of the concept is argued to have several academic and political implications. First, it endangers the broader validity and reliability of existing and future research, which is, in turn, essential for the organisation and application of scientific findings related to news diversity [51]. Second, and linked with the previous, there is a risk of formulating inadequate policy recommendations. For instance, with regard to the discussions on the existence of selective exposure or filter bubbles in the digital environment, the current literature is not able to present a clear overview on the state and outcome of diversity research in the digital environment. As a consequence, policy recommendations are rather limited to 'more research should be done' or 'insight into filter bubbles are indispensable' [9].

We argue that a clear description of what news diversity constitutes may be a first stepping stone to solve the above-mentioned issues. First, it may help scholars to map the current field and identify areas of ambiguity or neglect. Second, it enables news diversity scholars to make informed decisions when studying news diversity. This might be of particular importance for future diversity research, but also for the development of news recommenders.

In this section, we forward an approach to unravel the normative and conceptual assumptions underlying this concept ([52] for an extensive overview). These assumptions range from explicitly formulating the normative position to deciding on what kind of dimensions to measure (see Table 1). We will further elaborate on these assumptions by presenting three leading questions that enable the discussion on the meaning of news diversity.

Table 1. Distinguishing normative and conceptual assumptions of news diversity

Assumption	**Leading question**
Normative assumption: normative stance	Should news media reflect the diversity in society or should it treat all categories under study equally?
Conceptual assumption: sample selection	What or whom is studied: production, consumption or distribution?
Conceptual assumption: diversity dimensions	Which dimensions in news media content (e.g., gender, sentiment) or structure (e.g., ownership) are studied?

4.1 Normative assumptions

The first leading question is concerned with the idea of *open and reflective diversity* [53, 54]. The former evaluates diversity as an equal media representation of all categories. The latter argues that media should reflect the diversity in society [55]. Take, for instance, research on the diversity of political opinions in the news. From an open point of view, diversity would be evaluated as an equal representation of all voices in the political spectrum. From a reflective viewpoint, evaluation of diversity would be based on the question to what extent these voices coincide with the current distribution of political opinions in society.

4.2 Conceptual assumptions

A second question is related to *what or whom is studied*. Traditionally, this means a choice between the production side, in which news is made available, and the consumption side, in which people engage with news. However, in the current news environment, distribution actors such as search engines, recommendation systems, and aggregators could also be considered (e.g. [56]).

The third question deals with the most fundamental part of what constitutes news diversity: *the studied dimension(s) of diversity*. It concerns the focus of analysis, what researchers actually measure to make conclusions about news diversity. This might be centered on dimensions in the content or structure of news media. To name a few examples, we explain the content dimensions 'actor diversity' and 'party diversity'. The former refers to the affiliation or occupation of the actors who are quoted or paraphrased in the news [57]. The latter is concerned with the number of political parties across which a medium distributes its attention, either implicitly in terms of topics or explicitly in terms of party name [54].

To conclude, we want to emphasize that news diversity is a very broad concept, covering several aspects related to news, media and democracy. As such, news diversity remains an ambiguous concept when it is not accompanied by explications of the assumptions underlying this concept. Especially in the context of news recommendations systems, informed decisions on each of these assumptions as well as explicit statements should be made. Audiences, in the first place, but also other stakeholders such as policymakers should be aware of what kind of diversity is tweaked and which ideal is pursued.

During the workshop, the participants were asked to give their opinions in relation to the normative assumption related to news diversity. In particular, they were asked whether news media should reflect the diversity in society (i.e., reflective diversity) or should news media treat all categories under study equally (i.e., open diversity). Interestingly, a majority leaned towards 'open diversity'. Participants pointed out that while reflective diversity indeed mirrors society, it could, when taken too far, limit the forming of opinions. Instead, there could be a 'free market of diversity', in which an increase in reports on right-wing opinions would for example trigger a rise in the distribution of left-wing points of view (and vice versa). As such, a kind of equilibrium could be achieved. Other participants, moreover, pointed out that open diversity would allow niche opinions to grow and even become the majority. Therefore, it could also encourage change. For example, participants pointed to the idea that more coverage of female football, female scientists etc. could have a positive effect on the emancipation of women in society. Nonetheless, the participants discussed whether also the most extreme opinions should be allowed to circulate. Where one attendee stated to prefer to know of their existence, because this helps to assess one's own position on the spectrum, another considered that the right to access diverse information should be restricted in the same way as the freedom to express opinions. Accordingly, a diversity-enhancing news recommender should not promote content involving hate speech.

Concluding this discussion, we want to stress the importance of reflecting on the normative assumptions related to news diversity. As this workshop has shown, several (counter-)arguments may be used in favour or against open/reflective diversity. Future diversity research should focus on these (counter-)arguments to explore the consequences of each assumption.

5 Automated extraction of content dimensions in written news

This section zooms in on a number of diversity dimensions that can be detected in written news content. As indicated previously, several possibilities, such as actor diversity or the prominence of political parties, have been investigated in this field. These analyses often rely on manual coding of news items. Whereas manual analysis is powerful, it also practically restricts the number of media items that can be parsed, and easily leads to methodological differences between individual researchers. Automating the extraction of relevant dimensions, using techniques from the fields of computational linguistics and artificial intelligence, on the other hand, would allow for a standardised and fine-grained feeding of news algorithms on a large scale. The #NewsDNA research project focusses on the automated extraction of two possible content dimensions that can serve as building blocks for a diversity-driven news recommender: news topics and news events.

5.1 Topics

An intuitive analysis of the content of news articles is centred around news topics. In this context, topics are the general areas on which an article touches, such as politics, international news, or entertainment. News publishers use topics tags to organise their own news output. While some features of a topic taxonomy tend to recur, there is considerable variation between news outlets, making it hard to establish a mapping between them. Additionally, depending on the outlet, articles may belong exclusively to a single topic or to multiple topics. Variable tags based on current events, like *Brexit* or *immigration*, may be used alongside general tags. This lack of uniformity makes it generally impractical for the researcher to use outlet-provided topics tags for automatic analysis across publishers.

Some efforts exist to encourage consistent use of media topics in the news industry. An example of a topics framework promoted as a global standard is the IPTC Media Topics taxonomy [58]. It defines 17 top-level codes which hierarchically subdivide into subtopics up to five levels down. For example, the bottom-level code “housing and urban planning” can be traced back through “interior policy” and “government policy” to the top-level code “politics”. The deeper into the tree, the more granular the topic definitions become.

5.2 Events

Topics provide a general idea of the content of an article by describing which aspects of society it touches on, but they do not say anything about its specific contents. More semantics-driven algorithms can shed light on the events described in the text. We briefly discuss one such technique applied in the #NewsDNA project and illustrate the research effort involved.

An attractive and little-explored dimension of analysis is that of news events, i.e. the real-world events which provide the material and context for news articles. For example, in a fictional example entitles “Russian spies arrested in England”, the arrest of the Russian spies is the event that leads to the article being written. The goal of event extraction is to identify the real-world events referred to in news texts, as well as information on the actors, time, place, etc. involved in the event. In the example, the “Russian spies” are entities involved in the event and “England” is its stated location. Note that upstream technologies such as named entity extraction can play a role in discovering these participants [59].

An event extraction system, then, is an algorithm which takes as input a text and returns a number of event descriptions it has found in the text. Such a model is obtained through machine learning. First, a set of articles is prepared in which event descriptions have been manually annotated. Second, a machine-learning algorithm goes over this set and, through trial and error, learns to identify event descriptions matching the human-made gold standard. The system can then be run on previously unseen articles to produce new event descriptions.

Inevitably, to extract news events, we need to define what we consider to be a news event. Many different conceptions of “events” have been examined, some of which focus on the discovery of real events in text (see e.g. NewsReader [60], the ACE/ERE programs [61, 62], RED [63]) and some of which focus on fine-grained text semantics (e.g. the FrameNet project [64]).

Typically, a taxonomy of event types is used, such that each event mention found in the text can be classified in a semantic category such as “Conflict- Attack” or “Transaction-TransferOwnership” (from ERE [62]). The advantage of a fixed taxonomy is that it naturally defines the scope of news events: events that cannot be classified are not recognised.

A sizeable body of work (around the previously cited research programs [61, 62, 65]) focuses on event extraction in a closed data context, where the corpus of articles is given and the event type taxonomy is fixed. This leads to systems that perform well at extracting those specific categories of events, but fail at handling unrestricted news text discussing a wider variety of events. In an open data context, an automatic system must capture all relevant events from in- coming news texts. Designing a taxonomy for this is a difficult balancing act: a small taxonomy will exclude many relevant events, while a taxonomy with many different types will suffer from data sparsity (i.e. some event types are so rare that they cannot be learned or extracted reliably). Additionally, a fixed taxonomy may not adapt well to a stream of news whose tone changes with time. This imposes a process of constant retraining of the algorithm, which is feasible provided that data are available. For instance, suppose that incoming news focuses on a certain terrorist attack one month, whereas the big story of the next month is centred around the question of immigration. A system trained on data from one period in time may be disadvantaged when dealing with news from another. A natural way to sidestep this limitation is to allow for events of type 'unknown' to be extracted, but even in that case care needs to be taken so that 'unknown' events remain a minority within the training data [66]. The prediction of events without type has not been fully explored, as the theoretical applications of this technology tend to presume event type prediction is a desirable feature, or, at least, useful for other downstream applications.

For the purposes of news recommendation, the appeal of event extraction lies in linking event descriptions across articles. Given two event descriptions, specialized systems can establish identity links between them; two mentions that refer to the same event are called co-referent. Co-reference links can be established within but also across articles. It has been thoroughly researched for nominal entities, but not for events, and even less across documents [67]. It allows us to link together articles based on a deep semantic interpretation. For instance, using a topic-based system, we are able to cluster articles based on tags such as politics or business, or if our system is capable of fine-grained topic analysis, more current tags such as Brexit or economic crisis. If we know the specific events that occur in the articles, and if we know how to establish co-reference links between events across articles, we can create clusters based on single events. For example, we could gather all articles discussing Theresa May's resignation in June, with far greater precision than using topic-based methods. In terms of addressing diversity, we could also use these clusters and links to broaden the scope of recommendations in a more organic way by, for example, recommending articles located at the edge of a cluster or from closely neighbouring ones.

Cross-document event recognition and co-reference is key to moving the state of the art in natural language understanding and personalised recommendation. While solutions based on dimensions such as topics and actors work well with recommender systems, we propose that a more granular semantic analysis based on events can further enhance the precision of news recommenders.

After having explained the difference between 'topics' and 'events, we asked our audience to think of other content dimensions which could be of relevance in the context of automated text analysis. A first participant considered it would be interesting to categorise new articles according to their 'level of argumentation'. Well-argued opinions could consequently be singled out and further discussed, which in turn may facilitate public debate. Another attendee suggested that content extraction techniques could be used to verify whether the title – rather than functioning as a 'clickbait' –matches the information contained in an article. The detection of 'viewpoints' also put forward as an option. In that regard, others put forward that structural elements, such as an author's affiliation or background, or his or her country of origin, could also serve as indicators of 'bias' in a news items.

6 Towards a diversity-promoting news recommender

In this article, we addressed the conceptual development of a news diversity-enhancing recommendation system. To do so, we approached news recommendations from four different academic domains: computer sciences, law, communication sciences, and computational linguistics.

In the first section (i.e. computer sciences), we reviewed the state of the art of current news recommendation systems. In particular, we described two dominant methodologies – collaborative and content-based filtering – and unravelled their assumptions and drawbacks. We ended this section with a critique in that citizen – oriented concepts such as news diversity are currently underrepresented in these methodologies. Other concepts, such as accuracy or maximisation of the overlap between users' behavior, currently dominate the discourses in this field.

In the second discussion (i.e. law), we discussed the existence of a so-called 'right to diverse information'. By means of an analysis of relevant fundamental rights documents and case-law of the ECtHR and CJEU, we were able to support this statement. As a result, we concluded that governments carry the ultimate responsibility for the effective exercise of this right, in the past predominantly with respect to audio-visual media, but in the future potentially also with regard to the accessibility of online news.

In the third section (i.e. communication sciences), we explored the meaning of the mere notion of news diversity. As argued, diversity may function as an alternative, more citizen-oriented strategy to design news recommendation systems, yet the concept itself is characterized by ambiguity. As such, we started our discussion with the conceptual difficulties of this concept and their implications. Then, we presented an approach to unravel the normative and conceptual assumptions underlying this concept.

In the fourth discussion (i.e. computational linguistics), we explored how computational methods may enrich manual analysis in order to extract news content dimensions such as topics and events. We illustrated the usage of topic tags, and introduced automatic event extraction, citing applications, drawbacks and obstacles that emerge when these methods are set into practice.

Based on the feedback we received from the participants of the 2019 IFIP workshop on 'co-designing a personalised news diversity algorithmic model based on news consumers' agency and fine-grained content modelling', we consider that ensuring and promoting diversity in information exposure is an important public policy goal. As future developments in the fields of recommender systems and automated content extraction may contribute to its achievement, further research into the conceptualization of a news-diversity enhancing algorithm will continue to be undertaken in the #NewsDNA project across the four disciplines. On the one hand, conceptual questions stemming from the fields of communication science and law will be considered on a fundamental level. This concerns questions such as 'which dimensions should be selected to conceptualise news diversity?' or 'what is the optimal outcome of diversity to which audiences are steered?' to which no unequivocal answers yet exist. On the other hand, the fields of computational linguistics and computer sciences, which enable such a recommender system, still carry operational questions and difficulties. Relevant content dimensions must be translated into content extraction algorithms, which is not a solved issue. The design of the recommendation algorithm must also be carefully considered, as the right balance has to be made between relevance and diversity.

7 References

1. All the news that’s fit for you: The New York Times’ “Your Weekly Edition” is a brand-new newsletter personalized for each recipient, https://www.niemanlab.org/2018/06/all-the-news-thats-fit-for-you-the-new-york-times-your-weekly-edition-is-a-brand-new-newsletter-personalized-for-each-recipient/, last accessed 2019/12/05.
2. Mis niets over uw favoriete thema’s via “Mijn dS,” https://www.standaard.be/cnt/dmf20190304_04229666, last accessed 2019/08/09.
3. Gepersonaliseerd nieuws: matchmaker voor online media of journalistiek-ethisch mijnenveld?, https://www.vn.nl/gepersonaliseerd-nieuws-matchmaker-of-mijnenveld/, last accessed 2019/12/05.
4. Thurman, N., Moeller, J., Helberger, N., Trilling, D.: My Friends, Editors, Algorithms, and I. Digital Journalism. 7, 447–469 (2019). https://doi.org/10.1080/21670811.2018.1493936.
5. Thurman, N., Schifferes, S.: The Future of Personalization at News Websites. Journalism Studies. 13, 775–790 (2012). https://doi.org/10.1080/1461670X.2012.664341.
6. Araujo, T.B., Vreese, C.H. de, Helberger, N., Kruikemeier, S., Weert, J.C.M. van, Bol, N., Oberski, D., Pechenizkiy, M., Schaap, G., Taylor, L.: Automated Decision-Making Fairness in an AI-driven World: Public Perceptions, Hopes and Concerns. (2018).
7. Mcquail, D.: McQuail’s Mass Communication Theory. SAGE Publications Ltd, London ; Thousand Oaks, Calif (2010).
8. Möller, J., Helberger, N., Makhortkh, M., van Dooremalen, S.: Filterbubbels in Nederland (2019). Commissariaat voor de media (2019).
9. Borgesius, F.J.Z., Trilling, D., Möller, J., Bodó, B., Vreese, C.H. de, Helberger, N.: Should we worry about filter bubbles? Internet Policy Review. (2016).
10. Helberger, N.: On the Democratic Role of News Recommenders. Digital Journalism. 7, 993–1012 (2019). https://doi.org/10.1080/21670811.2019.1623700.
11. Recommendation Systems: General Collaborative Filtering Algorithm Ideas, http://www.cs.carleton.edu/cs_comps/0607/recommend/recommender/collaborativefiltering.html, last accessed 2019/12/05.
12. Sarwar, B., Karypis, G., Konstan, J., Riedl, J.: Item-based Collaborative Filtering Recommendation Algorithms. In: Proceedings of the 10th International Conference on World Wide Web. pp. 285–295. ACM, New York, NY, USA (2001). https://doi.org/10.1145/371920.372071.
13. Damme, K.V., Martens, M., Leuven, S.V., Abeele, M.V., Marez, L.D.: Mapping the Mobile DNA of News. Understanding Incidental and Serendipitous Mobile News Consumption. Digital Journalism. 1–20 (2019). https://doi.org/10.1080/21670811.2019.1655461.
14. Good, N., Schafer, J.B., Konstan, J.A., Borchers, A., Sarwar, B., Herlocker, J., Riedl, J.: Combining Collaborative Filtering with Personal Agents for Better Recommendations. In: Proceedings of the Sixteenth National Conference on Artificial Intelligence and the Eleventh Innovative Applications of Artificial Intelligence Conference Innovative Applications of Artificial Intelligence. pp. 439–446. American Association for Artificial Intelligence, Menlo Park, CA, USA (1999).

15. Liu, J., Pedersen, E., Dolan, P.: Personalized News Recommendation Based on Click Behavior. In: 2010 International Conference on Intelligent User Interfaces (2010).
16. Chen, Y., Wu, C., Xie, M., Guo, X.: Solving the Sparsity Problem in Recommender Systems Using Association Retrieval. JCP. 6, 1896–1902 (2011). https://doi.org/10.4304/jcp.6.9.1896-1902.
17. Adnan, Md.N.M., Chowdury, M.R., Taz, I., Ahmed, T., Rahman, R.M.: Content based news recommendation system based on fuzzy logic. In: 2014 International Conference on Informatics, Electronics Vision (ICIEV). pp. 1–6 (2014). https://doi.org/10.1109/ICIEV.2014.6850800.
18. Smyth, B., McClave, P.: Similarity vs. Diversity. In: Aha, D.W. and Watson, I. (eds.) Case-Based Reasoning Research and Development. pp. 347–361. Springer, Berlin, Heidelberg (2001). https://doi.org/10.1007/3-540-44593-5_25.
19. Lathia, N., Hailes, S., Capra, L., Amatriain, X.: Temporal Diversity in Recommender Systems. In: Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 210–217. ACM, New York, NY, USA (2010). https://doi.org/10.1145/1835449.1835486.
20. United Nations (UN) Special Rapporteur on Freedom of Opinion and Expression, Organization for Security and Co-operation in Europe (OSCE) Representative on Freedom of the Media, Organization of American States (OAS) Special Rapporteur on Freedom of Expression, African Commission on Human and Peoples' Rights (ACHPR) Special Rapporteur on Freedom of Expression and Access to Information: Joint Declaration on Freedom of Expression and "Fake News", Disinformation and Propaganda. (2017).
21. Commission: Tackling online disinformation: a European Approach. Communication (2018).
22. High level Group on fake news and disinformation: A multi-dimensional approach to disinformation. Report of the independent High level Group on fake news and online disinformation. European Commission Directorate-General for Communication, Networks, Content and Technology (2018).
23. EU Code of Practice on Disinformation. (2018).
24. European Convention for the Protection of Human Rights and Fundamental Freedoms. ETS No. 005 (1950).
25. Akandji-Kombe, J.-F.: Positive obligations under the European Convention on Human Rights. A guide to the implementation of the European Convention on Human Rights. Council of Europe, Strasbourg (2007).
26. Council of Europe/European Court of Human Rights: Positive obligations on member States under Article 10 to protect journalists and prevent impunity. (2011).
27. Charter of Fundamental Rights of the European Union. OJ C 326/391 (2000).
28. McGonagle, T.: Positive obligations concerning freedom of expression: mere potential or real power? In: Journalism at risk: Threats, challenges and perspectives. pp. 9–35. Strasbourg: Council of Europe (2015).
29. The Sunday Times v. the United Kingdom (No.1). (1979).
30. Informationsverein Lentia and others v. Austria. (1993).
31. Voorhoof, D., van Loon, A., Vier, C.: IRIS Themes – Vol. III – Freedom of Expression, the Media and Journalists. Case-law of the European Court of Human Rights. European Audiovisual Observatory, Strasbourg (2017).

32. Erdoğdu and İnce v. Turkey. (1999).
33. Sürek v. Turkey (No. 1). (1999).
34. Khursid Mustafa and Tarzibachi v. Sweden. (2008).
35. Times Newspapers LTD (Nos. 1 and 2) v. the United Kingdom. (2009).
36. Manole and Others v. Moldova. (2009).
37. Centro Europa 7 S.R.L. and Di Stefano v. Italy. (2012).
38. Explanations relating to the Charter of Fundamental Rights. OJ C 303/17 (2007).
39. Advocate General Jääskinen: Google Spain. (2013).
40. Beijer, M.: Limits of Fundamental Rights Protection by the EU: The Scope for the Development of Positive Obligations. Intersentia, Cambridge Antwerp Portland (2017).
41. Directive 2010/13/EU of the European Parliament and of the Council of 10 March 2010 on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services (Audiovisual Media Services Directive) (Text with EEA relevance). (2010).
42. Sky Österreich. (2013).
43. Animal Defenders International v. the United Kingdom. (2013).
44. McQuail, D.: Media Performance. In: The International Encyclopedia of Communication. American Cancer Society (2008). https://doi.org/10.1002/9781405186407.wbiecm045.
45. van der Wurff, R.: Do audiences receive diverse ideas from news media? Exposure to a variety of news media and personal characteristics as determinants of diversity as received. European Journal of Communication. 26, 328–342 (2011). https://doi.org/10.1177/0267323111423377.
46. Day, A.G., Golan, G.: Source and content diversity in Op-Ed Pages: assessing editorial strategies in The New York Times and the Washington Post. Journalism Studies. 6, 61–71 (2005). https://doi.org/10.1080/1461670052000328212.
47. Rodgers, R., Hallock, S., Gennaria, M., Wei, F.: Two Papers in Joint Operating Agreement Publish Meaningful Editorial Diversity. Newspaper Research Journal. 25, 104–109 (2004).
48. Voakes, P.S., Kapfer, J., Kurpius, D., Chern, D.S.-Y.: Diversity in the News: A Conceptual and Methodological Framework. Journalism & Mass Communication Quarterly. 73, 582–593 (1996). https://doi.org/10.1177/107769909607300306.
49. Raeijmaekers, D., Maeseele, P.: Media, pluralism and democracy: what's in a name? Media, Culture & Society. 37, 1042–1059 (2015). https://doi.org/10.1177/0163443715591670.
50. Napoli, P.M.: Deconstructing the diversity principle. Journal of Communication. 49, 7–34 (1999). https://doi.org/10.1111/j.1460-2466.1999.tb02815.x.
51. Liu, P., Li, Z.: Task complexity: A review and conceptualization framework. International Journal of Industrial Ergonomics. 42, 553–568 (2012). https://doi.org/10.1016/j.ergon.2012.09.001.
52. Joris, G., De Grove, F., Van Damme, K., De Marez, L.: News diversity reconsidered: A systematic literature review unravelling the diversity in conceptualizations (submitted).
53. McQuail, D., Van Cuilenburg, J.J.: Diversity as a Media Policy Goal: a Strategy for Evaluative Research and a Netherlands Case Study. Gazette (Leiden, Netherlands).

31, 145–162 (1983). https://doi.org/10.1177/001654928303100301.

54. Takens, J., Ruigrok, N., van, H.A.M.J., Scholten, O.: Old ties from a new(s) perspective: Diversity in the Dutch press coverage of the 2006 general election campaign. Communications. 35, 417–438 (2010). https://doi.org/10.1515/comm.2010.022.
55. McQuail, D.: Media Performance: Mass Communication and the Public Interest. Sage Publications (CA) (1992).
56. Möller, J., Trilling, D., Helberger, N., 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, 959–977 (2018). https://doi.org/10.1080/1369118X.2018.1444076.
57. Masini, A., Van, A.P.: Actor diversity and viewpoint diversity: Two of a kind? Communications. 42, 107–126 (2017). https://doi.org/10.1515/commun-2017-0017.
58. Media Topics, https://iptc.org/standards/media-topics/.
59. Colruyt, C., De Clercq, O., Hoste, V.: EventDNA : guidelines for entities and events in dutch news texts (v1.0).
60. Vossen, P.: Newsreader Public Summary. (2016).
61. Doddington, G., Mitchell, A., Przybocki, M., Ramshaw, L., Strassel, S., Weischedel, R.: The Automatic Content Extraction (ACE) Program – Tasks, Data, and Evaluation. In: Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04). European Language Resources Association (ELRA), Lisbon, Portugal (2004).
62. Aguilar, J., Beller, C., McNamee, P., Van Durme, B., Strassel, S., Song, Z., Ellis, J.: A Comparison of the Events and Relations Across ACE, ERE, TAC-KBP, and FrameNet Annotation Standards. In: Proceedings of the Second Workshop on EVENTS: Definition, Detection, Coreference, and Representation. pp. 45–53. Association for Computational Linguistics, Baltimore, Maryland, USA (2014). https://doi.org/10.3115/v1/W14-2907.
63. O'Gorman, T., Wright-Bettner, K., Palmer, M.: Richer Event Description: Integrating event coreference with temporal, causal and bridging annotation. In: Proceedings of the 2nd Workshop on Computing News Storylines (CNS 2016). pp. 47–56. Association for Computational Linguistics, Austin, Texas (2016). https://doi.org/10.18653/v1/W16-5706.
64. Ruppenhofer, J., Ellsworth, M., Schwarzer-Petruck, M., Johnson, C.R., Scheffczyk, J.: FrameNet II: Extended theory and practice. International Computer Science Institute (2006).
65. Vossen, P.: NewsReader at SemEval-2018 Task 5: Counting events by reasoning over event-centric-knowledge-graphs. In: Proceedings of The 12th International Workshop on Semantic Evaluation. pp. 660–666. Association for Computational Linguistics, New Orleans, Louisiana (2018). https://doi.org/10.18653/v1/S18-1108.
66. Colruyt, C., De Clercq, O., Hoste, V.: Comparing event annotations: notes on the EventDNA corpus IAA study. (2019).
67. Lu, J., Ng, V.: Event Coreference Resolution: A Survey of Two Decades of Research. In: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence. pp. 5479–5486 (2018).