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 and 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 have raised questions on how news recommendation systems can be more citizen-oriented so that they facilitate – rather than limit – the normative aims of journalism. In response to this question, this paper presents the conceptual development of such an algorithm by the interdisciplinary research project #NewsDNA. The project aspires to develop, evaluate and test a diversity-promoting news recommendation system. To do so, it combines state-of-the-art research from computer sciences (news recommendation systems); communication sciences (conceptualisations of news diversity); law (right to receive information) and computational linguistics (automated content extraction from text). As such, it hopes to give stakeholders – such as policymakers and news producers – a realistic view on the opportunities and pitfalls of developing a citizen-oriented news recommendation system.

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

1 Introduction

In recent years, news organisations increasingly explore how recommendation systems can be used to personalise their news content for their audiences. For example, The New York Times uses a mix of editorial curation and algorithms

to compose a newsletter tailored to each recipient [41]. In Belgium too, De Standaard and Het Nieuwsblad recently introduced a personal page that collects articles based on the reader's selected topics [51]. In The Netherlands, online newspapers nu.nl and Algemeen Dagblad both invested heavily on personalised news notifications [24]. Hence, news organisations are increasingly exploring implicit and explicit algorithmic news personalisation [48,49], similar to how media companies such as Netflix or Amazon personalise content.

Such personalisations rely primarily on automated recommendation systems. However, in contrast to what people think [14], these systems are not neutral. They are primarily based on a commercial logic: they present recommendations based on calculated relevance (for example based on selected fields of interests or past consumption patterns). This commercial logic contrasts with the normative concept of journalism as 'a marketplace of ideas' in which citizens are confronted with a diverse array of ideas [31]. Although empirical research currently supports a more nuanced view [35], news recommendation systems are argued to be a potential threat to an informed citizenry and the democratic processes between media, politics and audiences [57]. With these concerns in mind, several scholars have raised questions on how recommendation systems can be built in a more citizen-oriented way that maintains the normative aims of journalism [21].

The current paper presents the interdisciplinary research project #News-DNA, which provides a possible answer to this question. More specifically, it outlines the conceptual development of a recommendation system that uses news diversity as a key driver for personalised news recommendations. As this news recommender builds on insights from multiple disciplines, the remainder of this paper is organised per discipline. First, we present a state-of-the-art overview of the most commonly used methodologies to design news recommendation systems within computer sciences. Second, we explain what the conceptual meaning of news diversity is by building on literature in communication sciences. Third, we explore the legal ground of promoting news diversity. To end, we discuss the computational feasibility of news content extraction, provided by computational linguistics, to provide data to the aspired diversity-promoting news recommendation system.

2 News recommendations systems today

In this section, we provide an overview of current news recommendation systems, from a computer science perspective. We present two dominant approaches, collaborative and content-based filtering, and outline the current obstacles related to a citizen-oriented 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 [23]. As such, relevant news articles are predicted based on news articles read by so-called 'neighbours', other users who have historically had similar taste of news [45]. 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 [50].

Collaborative filtering methods have two major drawbacks when recommending news stories. First, news is quick and volatile, which exacerbates the first-rater problem [20,25]: 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 desired for news recommendations. Second, there is the sparsity problem [15], 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 short-comings, and consequently is often used for news recommendations [25,11]. 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, it introduces certain problems of its own. First, over-specialisation: content-based system cannot provide recommendations outside the scope of what the user already has shown interest in. Within journalism, this is the trigger for concerns of filter bubbles and news personalisation [57]. 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 an approach is that users are more likely to be exposed to a narrowing segment of popular items, as the focus lies on the maximisation of 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 [46,22]. 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.

3   Unravelling news diversity

Considering the previous section, it becomes clear that citizen-oriented concepts such as news diversity are currently overlooked in the current field of news recommendation. This contrasts with other academic fields, such as law and 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 [33,56].

In this section, we explore the conceptual meaning of news diversity. However, as the term is difficult to characterise, we start with the conceptual difficulties surrounding its definition. In the field of communication sciences, the first studies on news diversity can be found in the seventies (e.g. [39]) and eighties (e.g. [29]). Most of the research, however, has been published between 1995 and 2005 (e.g. [18,43,52]), examining the effects of audiovisual and digital media such as television and the web. While a significant body of literature now exists around the concept of news diversity, communication scholars are still struggling with the question of what it means, and how it should be measured [42,52]. Consequently, a wide range of diversity dimensions, means of assessments and assumptions are currently used to study news diversity empirically [38].

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 [26]. Second, and linked with the previous, there is a risk of formulating inadequate policy recommendations. For instance, with regards 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' [57].

We argue that a clear description of what news diversity constitutes may be a first stepping stone to solve the current 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.

We forward an approach to unravel the normative and conceptual assumptions underlying this concept. 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?

3.1   Normative assumptions

The first leading question is concerned with the idea of *open* and *reflective diversity* [34,47]. The former evaluates diversity as an equal media representation of all categories. The latter argues that media should reflect the diversity in society [32]. 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.

3.2   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. [36]).

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 measured 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 [28]. 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 [47].

To conclude, 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 to which ideal is steered.

4   A fundamental right to diverse information

This section focuses on the legal ground of news diversity, as opposed to news personalisation. As previously mentioned, the importance of an easily available and diverse news offering has been recognised in the field of communication sciences as well as in the field of law. However, from a legal perspective, the question arises whether, and to what extent, the *right to freedom of expression and information*, laid down in Article 10 of the European Convention on Human Rights ('ECHR') [10], and the positive obligations[5] stemming from the above-mentioned article, include an actual *right to diverse information*. Its existence, including a corresponding responsibility for States to take affirmative action to ensure compliance with the Convention [30], would enable citizens to *force* their respective governments to indeed 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)."

On numerous occasions, the European Court of Human Rights ('ECtHR') has interpreted this 'freedom to receive information and ideas'. 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)" [8]. 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" [4]. 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 following their involvement in the spread of separatist or pro-Kurdish propaganda [53], 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 [3]. 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 [9]. In the Khurshid Mustafa and Tarzibachi v. Sweden case of 2008, which concerned a (prohibition of) reception of information by

---

[5] 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 specifically, to adopt reasonable and suitable measures to protect the rights of individuals" [13]

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" [5]. In the 2009 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)" [6]. 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)" [2].

In conclusion, the European Court of Human Rights has clearly recognised a right of the public to be informed about different viewpoints in relation to a specific topic. The State is ultimately responsible for its effective exercise, and indisputably within the context of the audio-visual media sector. Case-law related to online news personalisation is however not yet available, yet its logic could potentially be transposed to that context as well. While the ECtHR so far has not recognised the right to receive diverse information as a duty toward States in relation to the people's right to receive information and ideas via the Internet, it very well could, once it would consider its influence to be as impactful as that of the broadcast media [1].

5 Automated extraction of content dimensions in written news

This section zooms in on the use of written news content as a diversity dimension. As indicated previously, several possibilities have been investigated in this field, such as actor diversity or the prominence of political parties. These analyses often rely on manual coding of news items. Whereas manual analysis is powerful, is 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, can allow for analysis on a larger scale. Several techniques have been brought to a high level of performance. Named entity recognition, for instance, is a well-established method to extract proper names from a text, which can then be correlated to actors (see [37] for an overview). In this section, we discuss two possible *content dimensions*: news topics and news events.

5.1 Topics

An intuitive analysis of the content of news articles is centered 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 crisis*, 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 [7]. 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 [17].

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 which focus on the discovery of real events in text (see e.g. NewsReader [55], the ACE/ERE programs [19,12], RED [40]) and some which focus on fine-grained text semantics (e.g. the FrameNet project [44]).

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 [12]). 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 [19,12,54]) 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 incoming 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 can not be learned or extracted reliably). Additionally, a fixed taxonomy may not adapt well to a stream of news whose tone changes with time. 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 [16]. 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 attraction 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 [27]. 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.

6   Towards a diversity-promoting news recommender

In this article, we addressed the conceptual development of a citizen-oriented recommendation system, and in particular a recommendation system that promotes news diversity. To do so, we approached the news recommender from four different academic domains: computer sciences, communication sciences, law 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 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 third discussion (i.e. law), we discussed the existence of a so-called 'right to diverse information'. By means of an analysis of relevant case-law of the European Court of Human Rights, 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, and considering its impact on society, especially within the context of the audio-visual media.

In the fourth discussions (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.

To conclude, we argue that citizen-oriented news recommendation systems such as those of #NewsDNA can be developed, but should be critically examined before and during the development phase. On the one hand, conceptual questions stemming from the fields of communication science and law must 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

exist. On the other hand, the fields of computational linguistics and computer science, which enable such a recommender system, present their operational questions and difficulties. Relevant content dimensions must be translated into content extraction algorithms, which is not a solved problem. The design of the recommendation algorithm must also be carefully considered, as the right balance has to be made between relevance and diversity.

References

1. Animal Defenders International v. The United Kingdom (ECtHR). §119 (2013). ECLI:CE:ECHR:2013:0422JUD004887608.
2. Centro europa 7 s.r.l. and di stefano v. italy. §134 (2012). ecli:ce:echr:2012:0607jud003843309.
3. Erdoğdu and İnce v. turkey. §52 (1999). ecli:ce:echr:1999:0708jud002506794.
4. Informationsverein lentia and others v. austria. §138 (1993). ecli:ce:echr:1993:1124jud001391488.
5. Khurshid mustafa and tarzibachi v. sweden. §§31-32 (2008). ecli:ce:echr:2008:1216jud002388306.
6. Manole and others v. moldova. §100 (2009). ecli:ce:echr:2009:0917jud001393602.
7. Media topics, https://iptc.org/standards/media-topics/
8. The sunday times v. the united kingdom (no.1). §68 (1979). ecli:ce:echr:1979:0426jud000653874.
9. Sürek v. turkey (no. 1). §59 (1999). ecli:ce:echr:1999:0708jud002668295.
10. European convention for the protection of human rights and fundamental freedoms, as amended by protocols nos. 11 and 14 (Nov 1950), https://www.refworld.org/docid/3ae6b3b04.html
11. Adnan, M.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 (May 2014). https://doi.org/10.1109/ICIEV.2014.6850800
12. 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 (2014)
13. 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. No. 7 in Human Rights Handbooks, Council of Europe, Strasbourg (Jan 2007), https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent-?documentId=090000168007ff4d accessed 31-03-2019
14. Araujo, T., de Vreese, C., Helberger, N., Kruikemeier, S., van Weert, J., Bol, N., Oberski, D., Pechenizkiy, M., Schaap, G., Taylor, L., et al.: Automated decision-making fairness in an ai-driven world: Public perceptions, hopes and concerns (2018)
15. Chen, Y., Wu, C., Xie, M., Guo, X.: Solving the sparsity problem in recommender systems using association retrieval. Journal of computers **6**(9), 1896–1902 (2011)
16. Colruyt, C., De Clercq, O., Hoste, V.: Comparing event annotations: notes on the EventDNA corpus IAA study (2019)
17. Colruyt, C., De Clercq, O., Hoste, V.: EventDNA: Annotation guidelines for entities and events in dutch news texts (v1.0) (2019)

18. Day, A., 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**(1), 61–71 (2005). https://doi.org/10.1080/141670052000328212

19. 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 LREC (2004)

20. Good, N., Schafer, J.B., Konstan, J.A., Borchers, A., Sarwar, B., Herlocker, J., Riedl, J., et al.: Combining collaborative filtering with personal agents for better recommendations. AAAI/IAAI **439** (1999)

21. Helberger, N.: On the democratic role of news recommenders. Digital Journalism pp. 1–20 (2019)

22. 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 (2010)

23. Lew, D., Sowell, B.: Collaborative filtering, http://www.cs.carleton.edu/cs_comps/-0607/recommend/recommender/collaborativefiltering.html, accessed: 2019-08-08

24. Lieman, R.: Gepersonaliseerd nieuws: matchmaker voor online media of journalistiek-ethisch mijnenveld? (2019), https://www.vn.nl/gepersonaliseerd-nieuws-matchmaker-of-mijnenveld/

25. Liu, J., Dolan, P., Pedersen, E.R.: Personalized news recommendation based on click behavior. In: Proceedings of the 15th international conference on Intelligent user interfaces. pp. 31–40. ACM (2010)

26. Liu, P., Li, Z.: Task complexity: A review and conceptualization framework. International Journal of Industrial Ergonomics **42**(6), 553–568 (2012)

27. 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. International Joint Conferences on Artificial Intelligence Organization (2018). https://doi.org/10.24963/ijcai.2018/773, https://www.ijcai.org/proceedings/2018/773

28. Masini, A., Van Aelst, P.: Actor diversity and viewpoint diversity: Two of a kind? Communications-European Journal of Communication Research **42**(2), 107–126 (2017). https://doi.org/10.1515/commun-2017-0017

29. McCombs, M.: Effect of monopoly in cleveland on diversity of newspaper content. Journalism Quarterly **64**(4), 740–792 (1987). https://doi.org/10.1177/107769908706400408

30. 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)

31. McQuail, D.: Mass communication theory: An introduction. Sage Publications, Inc, Thousand Oaks, CA, US, 2nd edn. (1987)

32. McQuail, D.: Media performance: Mass communication and the public interest. Sage Publications, Inc, Thousand Oaks, CA, US (1992)

33. McQuail, D.: Media performance. The International Encyclopedia of Communication (2008)

34. McQuail, D., Van Cuilenburg, J.J.: Diversity as a media policy goal: A strategy for evaluative research and a netherlands case study. International Communication Gazette **31**(3), 145–162 (1983). https://doi.org/10.1177/001654928303100301

35. Möller, J., Helberger, N., Makhortykh, M., van Dooremalen, S.: Filterbubbels in nederland. Report, Commisariaat voor de media (2019), https://www.cvdm.nl/wp-content/uploads/2019/05/Rapport-Filterbubbels-in-Nederland.pdf

36. 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**(7), 959–977 (2018). https://doi.org/10.1080/1369118X.2018.1444076

37. Nadeau, D., Sekine, S.: A survey of named entity recognition and classification. Lingvisticae Investigationes **30**(1), 3–26 (2007)

38. Napoli, P.M.: Deconstructing the diversity principle. Journal of Communication **49**(4), 7–34 (1999). https://doi.org/10.1111/j.1460-2466.1999.tb02815.x

39. Nestvold, K.J.: Diversity in local television news. Journal of Broadcasting **17**(3), 345–352 (1973). https://doi.org/10.1080/08838157309363697

40. O'Gorman, T., Wright-Bettner, K., Palmer, M.: Richer event description: Integrating event coreference with temporal, causal and bridging annotation pp. 47–56 (2016). https://doi.org/10.18653/v1/W16-5706, http://aclweb.org/anthology/W16-5706

41. Owen, L.: All the news that's fit for you: The new york times' "your weekly edition" is a brand-new newsletter personalized for each recipient (2018), 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/

42. Raeijmaekers, D., Maeseele, P.: Media, pluralism and democracy: what's in a name? Media, Culture Society **37**(7), 1042–1059 (2015). https://doi.org/10.1177/0163443715591670

43. Rodgers, R., Hallock, S., Gennaria, M., Wei, F.: Two papers in joint operating agreement publish meaningful editorial diversity. Newspaper Research Journal **25**(4), 104–109 (2004). https://doi.org/10.1177/073953290402500411,

44. Ruppenhofer, J., Ellsworth, M., Schwarzer-Petruck, M., Johnson, C.R., Scheffczyk, J.: Framenet ii: Extended theory and practice (2006)

45. Sarwar, B.M., Karypis, G., Konstan, J.A., Riedl, J., et al.: Item-based collaborative filtering recommendation algorithms. Www **1**, 285–295 (2001)

46. Smyth, B., McClave, P.: Similarity vs. diversity. In: International conference on case-based reasoning. pp. 347–361. Springer (2001)

47. Takens, J., Ruigrok, N., van Hoof, 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-European Journal of Communication Research **35**(4), 417–438 (2010). https://doi.org/10.1515/comm.2010.022

48. Thurman, N., Moeller, J., Helberger, N., Trilling, D.: My friends, editors, algorithms, and i: Examining audience attitudes to news selection. Digital Journalism pp. 1–23 (2019)

49. Thurman, N., Schifferes, S.: The future of personalization at news websites: lessons from a longitudinal study. Journalism Studies **13**(5-6), 775–790 (2012)

50. Van Damme, K., Marijn, M., Van Leuven, S., Vanden Abeele, M., De Marez, L.: Mapping the mobile dna of news. understanding incidental and serendipitous mobile news consumption (2019)

51. Verhoeven, K., Ghijs, I., Sioen, L.: Nieuw — mis niets over uw favoriete thema's via 'mijn ds' (2019), http://www.standaard.be/cnt/dmf20190304_04229666

52. Voakes, P.S., Kapfer, J., Kurpius, D., Chern, D.S.: Diversity in the news: A conceptual and methodological framework. Journalism Mass Communication Quarterly **73**(3), 582–593 (1996). https://doi.org/10.1177/107769909607300306

53. 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 edn edn. (Dec 2017), http://book.coe.int/eur/en/european-audiovisual-observatory/6691-pdf-iris-themes-freedom-of-expression-the-media-and-journalists.html

54. Vossen, P.: NewsReader at SemEval-2018 task 5: Counting events by reasoning over event-centric-knowledge-graphs p. 7

55. Vossen, P.: Newsreader public summary, vol. 31 (2016)

56. 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**(4), 328–342 (2011). https://doi.org/10.1177/0267323111423377

57. Zuiderveen Borgesius, F., Trilling, D., Möller, J., Bodó, B., de Vreese, C.H., Helberger, N.: Should we worry about filter bubbles? Internet Policy Review **5**(1), 1 – 16 (2016). https://doi.org/10.14763/2016.1.401