

Recommender Systems in the Media Industry: *Relevance, Recency, Popularity, and Diversity*

 **David Graus**

 david.graus@fdmediagroep.nl

 [@dvdgrs](https://twitter.com/dvdgrs)

whoami



-  **Academia**

- BA Media Studies @ UvA (2008)
- MSc Media Technology @ Universiteit Leiden (2012)
- PhD Information Retrieval @ UvA (2017)

-  **Industry**

- Editor radio/online public broadcaster NTR (between BA & MSc)
- Research Intern @ Microsoft Research, US
- Data Scientist @ FD Mediagroep
- **Lead Data Scientist @ FD SMART Journalism / BNR SMART Radio**

In what is to follow...

- An introduction of FD Mediagroep
- Personalization & RecSys at FD Mediagroep:
 - SMART Radio
 - SMART Journalism
 - Lessons learnt building/productizing a RecSys;
 - Relevance
 - Usefulness
 - Trust

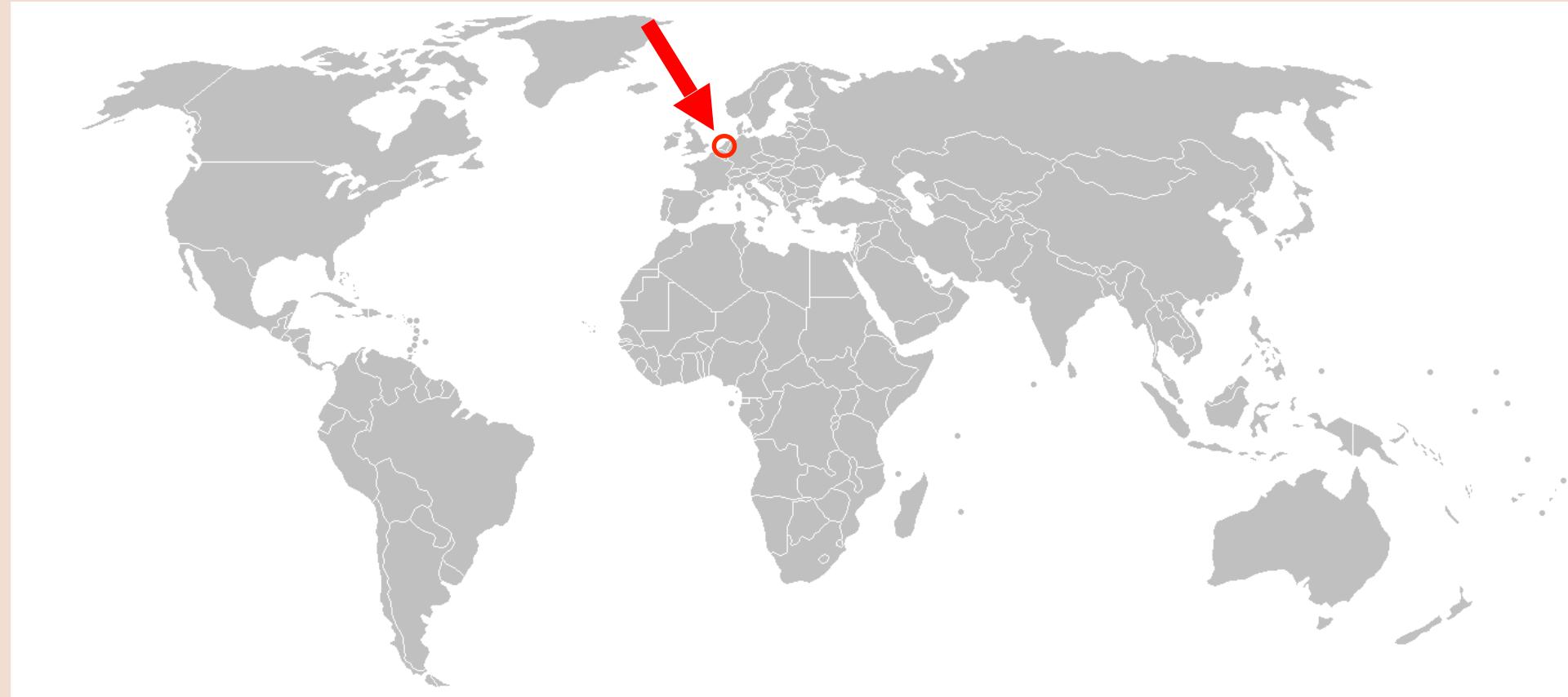
Part 1: Introduction

het financieele dagblad

BNR NIEUWSRADIO 100 FM

100 FM





**The leading information provider in the financial economic domain
in the Netherlands**





AGENDA

WEEKMAGAZINE FD

Beleid en beleidsvoering
Bijdragen van de politiek aan de ontwikkeling van
de economie en de marktvoering



10 maart 2019

BEURZEN

Mutatie t.o.v. vorige handelsdag



Bartjens

Na Uwek komt zonneschijn

fd.

het financieele dagblad

Woensdag 8 mei 2019 • Nummer 109 • 32 pagina's
'Al het goede tot het financiële eenige betrekking heeft' - sinds 1796

FINANCIËLE SECTOR

Platforms voor crowdfunding

Handelshuis
Teamspeler
gaat alleen
verder in top
Flow Traders

PAGINA 16

Beloofde
versus werkelijke
rendementen

De vooruitstrevende
rendementen bereiken gemiddeld
8,9%

AI @ FD Mediagroup

Google News Initiative

Digital News Innovation Fund Overview Insights Projects Report Participate **Funding Closed**

Projects > **BNR Smart Radio (Round 3)**

BNR Smart Radio (Round 3)

SHARE:    ...



NETHERLANDS | PROJECT TYPE: MEDIUM | BNR NIEUWSRADIO AND FD MEDIAGROEP B.V.

Summary

BNR Smart Radio aims to provide listeners with a personalized playlist, filled with relevant and automatically generated audio fragments based on BNR's radio shows and podcasts. Users can tell us their interests and we will match them with a database of automatically segmented audio files with a rich set of metadata, generated by a model based on the latest AI/ML techniques.

This creates a non-linear and personalized listening experience for spoken word audio, catering to a new audience of younger listeners, while at the same time better serving our current listeners who are increasingly more interested in shorter bits of information instead of longer radio shows.

The solution

Up till now, radio in The Netherlands is mostly linear product. Finding a certain item after an episode has aired is almost impossible. By segmenting radio shows, we will be able to open our archive of radio shows to the general public. It will be easy to find short items, stay up to date with the latest items matching your interests, and to share these items with your friends and colleagues. Analyzing and segmenting all of our audio by hand, is a monumental task. We believe this time should be spent on journalism and creating more high-quality content. This project aims to completely automate the analysis and segmentation of audio, by using the latest AI/ML techniques.

AI @ FD Mediagroup

 Google News Initiative

Digital News Innovation Fund Overview Insights **Projects** Report Participate Funding Closed

Projects > TL;DR - A personalized summary of your news (Round 4)

TL;DR - A personalized summary of your news (Round 4)

SHARE:    ...

 NETHERLANDS | PROJECT TYPE: LARGE | HET FINANCIËLE DAGBLAD B.V. (PART OF FD MEDIAGROEP B.V.)

Summary

The TL;DR project will deliver completely personalised summaries of FD Mediagroep content, generated by an innovative machine learning/AI algorithm, helping users stay up to date with the latest news.

The solution

Analysing content and creating accurate summaries is still a manual task. Providing a single summary, let alone multiple summaries, for every FD article is a monumental task and would require a serious investment in time or money, which should be spent on creating more high-quality journalism. TL;DR will automate this process by building a model using the latest AI/ML techniques that analyses content to provide multiple summaries, which can be combined with our recommendation engine or help increase the availability of content. The result will be a landing page that not only contains the information you're interested in, but also presents it in a way that matches your reading style. This will be a valuable addition for current subscribers, and a new product to attract new customers looking for a quick and user-friendly update of the latest

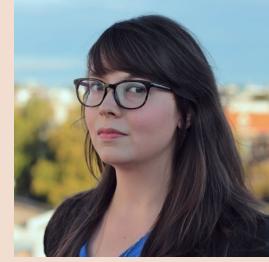
Team



Dung



Bahadir



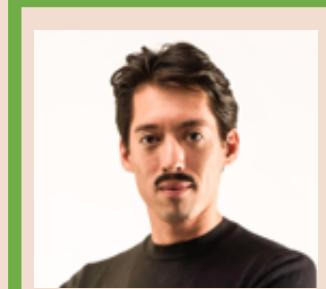
Anca



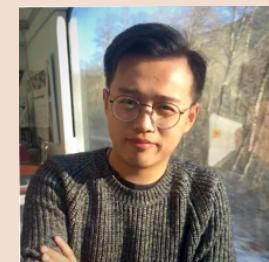
Philippe



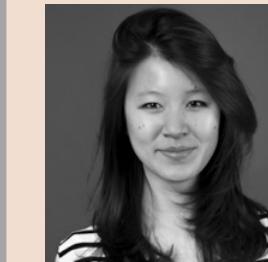
Maya



David



Feng



Li'ao



Klaus



Oberon



Manon



Azamat

AI @ FDMG: Academia/Industry

SMART Jou Recor

Maya Sappelli, Dung Chu, Bahadir Cambraus, David Graus

Abstract

FD Mediagroep (FDMG¹) is the leading informant in the financial economic domain in the Netherlands. FDMG operates "Het Financieele Dagblad," (FD) a daily financial newspaper, similar to the Financial Times. In addition, FDMG operates the daily all-news radio station "Business News Radio" (BNR). As we have various backgrounds and interests in media and radio (both news articles and radio snippets), we match the interests of a particular user by personalizing the app's content based on their interaction with the app.

- Personalized news: Recommending summaries of news articles based on user preferences and interests of our users.
- Personalized radio: A non-radio snippets that match the user's interests.

CCS CONCEPTS

- Information systems → Personalization;

KEYWORDS

personalization, news, radio

ACM Reference Format:
 Maya Sappelli, Dung Manh Chu, Bahadir Cambraus, David Graus. 2018. SMART Radio: Personalized News. In *Dutch-Belgian Information Retrieval Workshop (IWBIR)*, NY, USA, Article 4, 2 pages.

1 INTRODUCTION

FD Mediagroep (FDMG²) is the leading informant in the financial economic domain in the Netherlands. FDMG operates "Het Financieele Dagblad," (FD) a daily financial newspaper, similar to the Financial Times. In addition, FDMG operates the daily all-news radio station "Business News Radio" (BNR). FDMG personalizes the news experience, both in news and radio [3], since we believe this will help to serve our users better by helping them to unlock long tail content from FDMG's news sources.

ABSTRACT

In this demonstration paper we present a new feature for BNR Nieuwsradio.¹ The SMART Radio app, which offers a personalized news and radio experience. It does so by attempting to offer our users more targeted and relevant news and radio shows. We employ audio tagging techniques to achieve this. In its present form, users can subscribe to suggestions of relevant radio fragments and like to improve the app's personalization of the user's interaction with the app.

ABSTRACT

Personalization of media services e.g., through the rise of personalization systems across media outlets. The rise of distrust and skepticism of personal data, spurred by the introduction of the GDPR, data breaches and concerns in general. We feel transparent about the information used in this position paper we provide transparency through explanations. Specifically, we aim to explain how our recommender system works. Profile explanations can help users understand and trust our recommender system. To allow users to provide feedback on our recommended user profiles, we believe that user profile explanations as an essential part of the current online news reading experience. However, news recommender systems are being criticized for their 'black-box' approach to data collection and processing, and for their lack of explainability and transparency. This paper focuses on explaining user profiles constructed from aggregated reading behavior data, used to provide content-based recommendations. By doing so, the paper makes a first step toward consolidating epistemic values of news providers and news readers. We present an evaluation of an explanation interface reflecting these values, and find that providing users with different goals for self-actualization (i.e., *Broaden Horizons* vs. *Discover the Unexplored*) influences their reading intentions for news recommendations.

KEYWORDS

explainability, user profile, user control, self-actualization, news recommender systems

ACM Reference Format:
 Emily Sullivan, Dimitrios Bountouridis, Jaron Harambam, Shabnam Najafian, Felicia Loecherbach, Mykola Makhortykh, Domokos Kelen, Daricia Wilkinson, David Graus, & Nava Tintarev. 2019. Reading News with a Purpose: Explaining User Profiles for Self-Actualization. In *Proc. of 27th Conference on User Modeling, Adaptation & Personalization (UMAP 2019)*, June 9–12, 2019, Larnaca, Cyprus, pp. 1–10. ACM, New York, NY, USA.

1 INTRODUCTION

Personalized experiences powered by recommender systems permeate our lives. However, the rise of distrust and skepticism around the collection and use of personal data generates an increased interest in *explainability* and *transparency* of black-box recommender systems [12, 13, 31]. Many works [14, 19, 35] show that explanations are integral for users to understand their consumption preferences and achieve their *epistemic goals* i.e., goals for knowledge development. Moreover, researchers have begun exploring methods that support the exploration and understanding of users' current, aspirational, and self-actualized goals to provide transparency in recommender systems [11, 22]. Yet, the task of opening the black box of recommender systems remains notoriously hard to achieve [8, 28]. For the domain of *online news*, this might limit the readers' control over information diets [15]. Limited explainability can also diminish users' engagement with recommended content [21], as well as raising concerns about the societal role of news media [9] and consequences of increased algorithmization [26, 33].

To address these worries, this paper focuses on the goal of explaining recommendations in the context of *online news*. We explore methods of explaining one aspect of how a content-based recommender system works: the *user profile*. We aim to automatically summarize and visualize the recommender's high-dimensional internal representations of users. These profiles represent their personal interests and

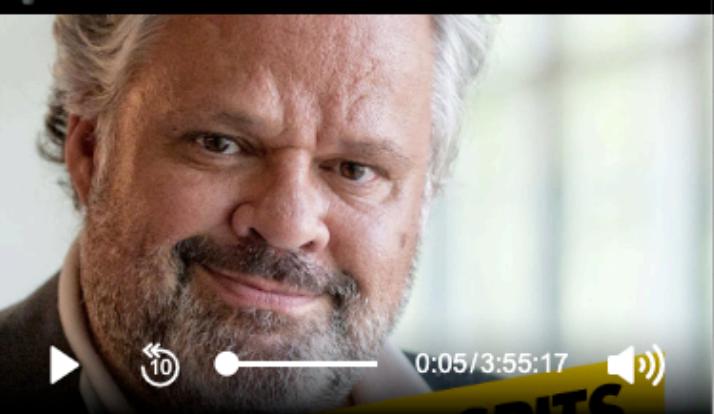
¹<http://www.bnrr.nl>

**Reading News with a Purpose:
Explaining User Profiles for Self-Actualization**

Emily Sullivan*, Dimitrios Bountouridis*, Jaron Harambam[†], Shabnam Najafian*, Felicia Loecherbach^{††}, Mykola Makhortykh[‡], Domokos Kelen^{‡‡}, Daricia Wilkinson[¶], David Graus^{**}, Nava Tintarev^{*}

^{*}Delft University of Technology, The Netherlands
[†]Amsterdam School of Communication Research, University of Amsterdam, The Netherlands
^{††}Vrije University, Amsterdam, The Netherlands
[‡]Institute for Computer Science and Control, Hungarian Academy of Sciences, Hungary
[¶]FD Mediagroep
^{**}Clemson University, Clemson, USA
^{*}e.e.sullivan-mumm@tudelft.nl

Part 2: SMART Radio



DE OCHTENDSPITS | 9 SEPTEMBER

DELEN <

DE OCHTENDSPITS



DE OCHTENDSPITS | 9 SEPTEMBER

BNR Webredactie / Vandaag, 11:40

Om ook in de toekomst handel te kunnen drijven met China moet er een stevig handels- en investeringsakkoord met de Chinezen worden gesloten. Daarvoor pleit VNO-NCW vanmiddag bij de China-hoorzitting in de Tweede Kamer.

[De Ochtendspits | 9 september](#)

Meer nieuws uit De Ochtendspits

Overheid betaalt dubbel aan paleizen koningshuis

D66 wil veestapel halveren

FNV: subsidie ontmoedigt bedrijven om lonen te verhogen



ALLES OVER DE OCHTENDSPITS

vind je hier



POPULARSTE PODCASTS



Kloof tussen economische en financiële realiteit groeit



'Nederland niet klaar voor cyberaanval'



Negatieve spaarrente lijkt onvermijdelijk



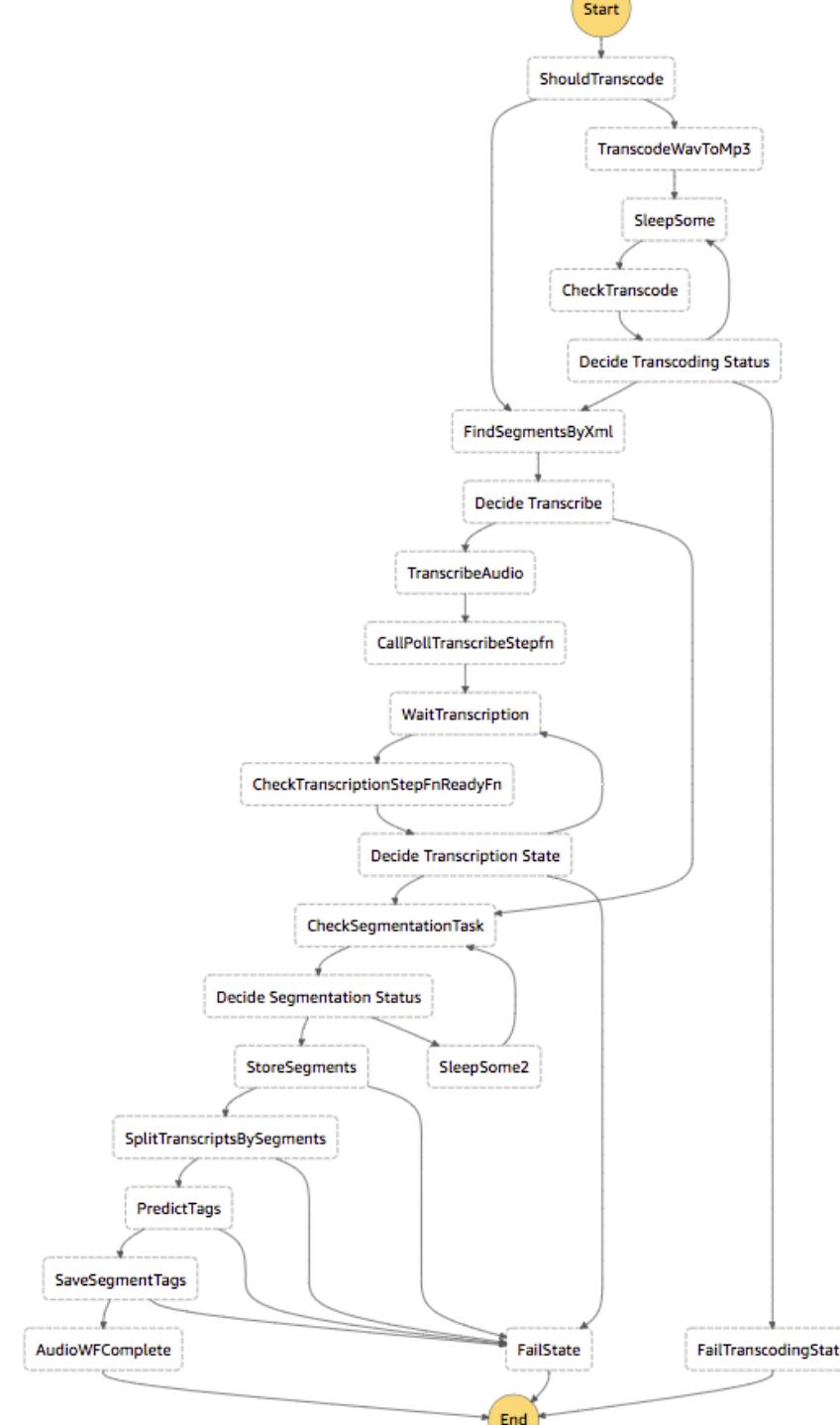
De Ochtendspits | 9 september



Speciale aanbieding

SMART Radio

- (Transcribe)
- Segment
- Tag
- Serve



Transcribe

Segment

- Based on metadata, text, and audio.

Improving automated segmentation of radio shows with audio embeddings

Oberon Berlage
University of Amsterdam
post@oberonberlage.nl

Mahsasat Shahshahani
University of Amsterdam

David Graus
FD Mediagroep

ABSTRACT

Audio features have been proven useful for increasing the performance of automated topic segmentation systems. This study explores the use of audio embeddings as a way of representing audio to improve automated, topically coherent segmentation of radio shows, as this has not been done before. Three different audio embedding generators were created with a multi-class classification task on three datasets from different domains. We generate audio embeddings by taking the output of the pre-final layer of a Convolutional Neural Network with spectrograms as input. The created audio embeddings are evaluated on their performance in a topic segmentation task and we compare them against a text-only baseline. We find that using audio embeddings based on a non-speech sound event classification task significantly outperform the baseline by 32.3% in F1-measure and that different classification tasks yield audio embeddings that show different segmentation performances.

1 INTRODUCTION

With the rise of on-demand media services and podcasts as a medium, people have gained the ability to compose their personalised playlists. To bring this on-demand approach to radio, BNR Nieuwsradio created the SMART Radio application [20]. SMART Radio provides listeners with a playlist of fragments from radio shows that are matched to the listeners preferred topics. To convert hour-long radio shows into topically coherent fragments, BNR created a segmentation system based on Automatic Speech Recognition (ASR) transcripts. We are interested in extending the topic segmentation system by including information from the audio signal, because it contains information that is complementary to the text-only features.

Previous research has shown that incorporating information from the audio in a topic segmentation system is beneficial for the topic segmentation performance [18, 26, 27]. When using audio as an input for analysis tasks, it is common practice to use hand-crafted features [6] or spectrogram representations of the single-

sound event classification task, a part of radio fragment classification task and a word classification task.

We evaluate the audio embedding generators on their performance in an existing topic segmentation system that was proposed by Sheikh [24], and compare the differences between the three embedding generators.

We find that using audio embeddings generated from a sound event classification task significantly outperform the text-only baseline by 32.3%. The combination of text features and audio embeddings in the segmentation task always leads to an increase in recall and overall performance. We show that segmentation based on audio embeddings alone can have a comparable performance to a text-only system. We see a clear difference in the segmentation performance of different embedding generators that were studied.

In this study the following research questions are addressed:

RQ1: Can audio embeddings be used to improve automatic, topically coherent segmentation of radio shows?

RQ2: What is the effect of different training tasks for embedding generators on the topic segmentation performance?

2 RELATED WORK

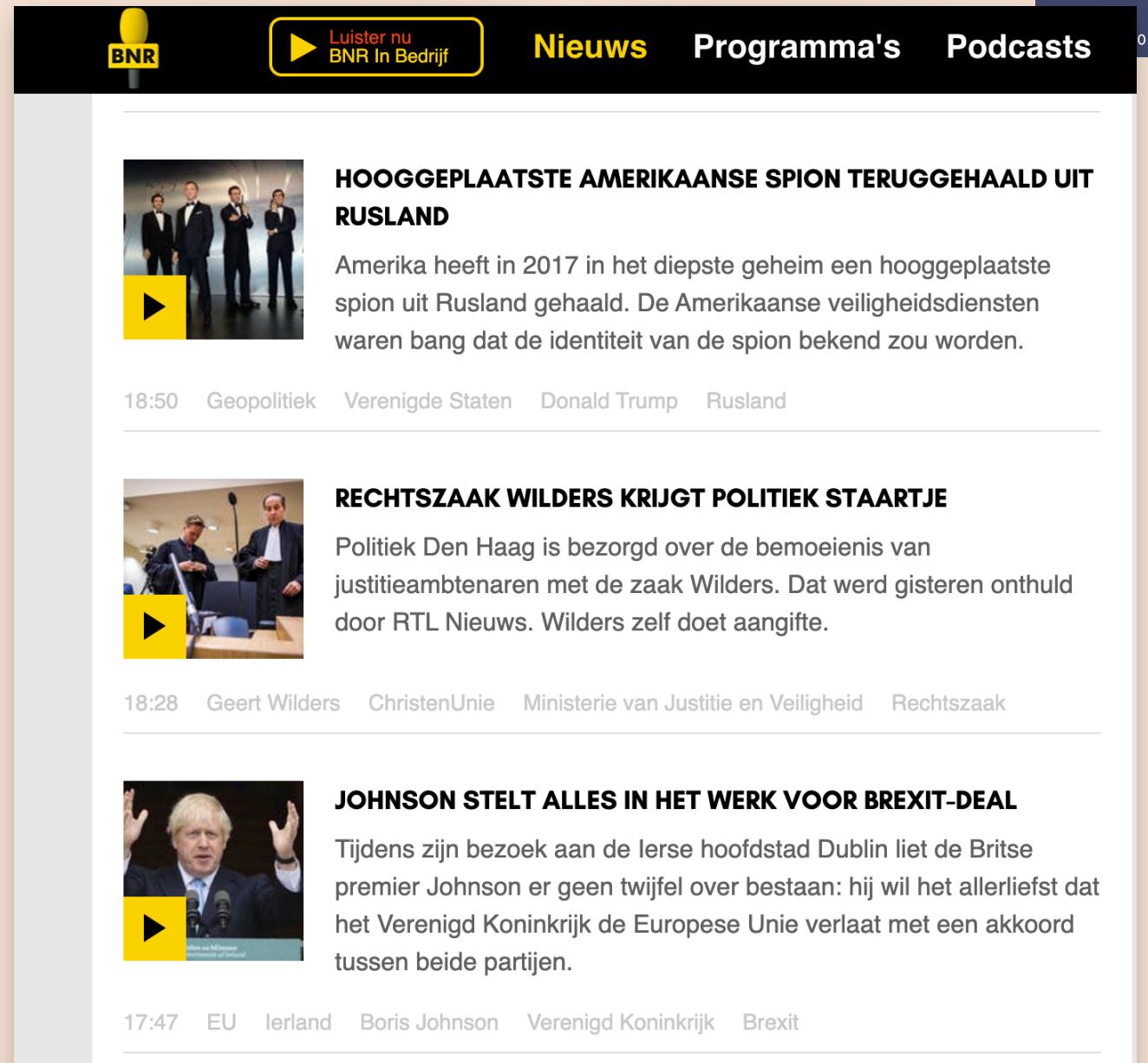
The following sections provide an overview of earlier work that is related to topic segmentation and the use of audio in machine learning tasks. We first introduce the field of topic segmentation, after which we discuss the use of audio in topic segmentation tasks. Furthermore, we discuss the use of sound in machine learning problems in general, and conclude with an introduction to the concept of audio embeddings.

2.1 Topic Segmentation

Research has been performed in the field of topic segmentation for decades, on a variety of modalities such as text or audio. Previous work defines topic segmentation as the process of dividing a document into

Tag

- Simple multilabel text classifier
- Trained on transcripts of segments + associated tags from website

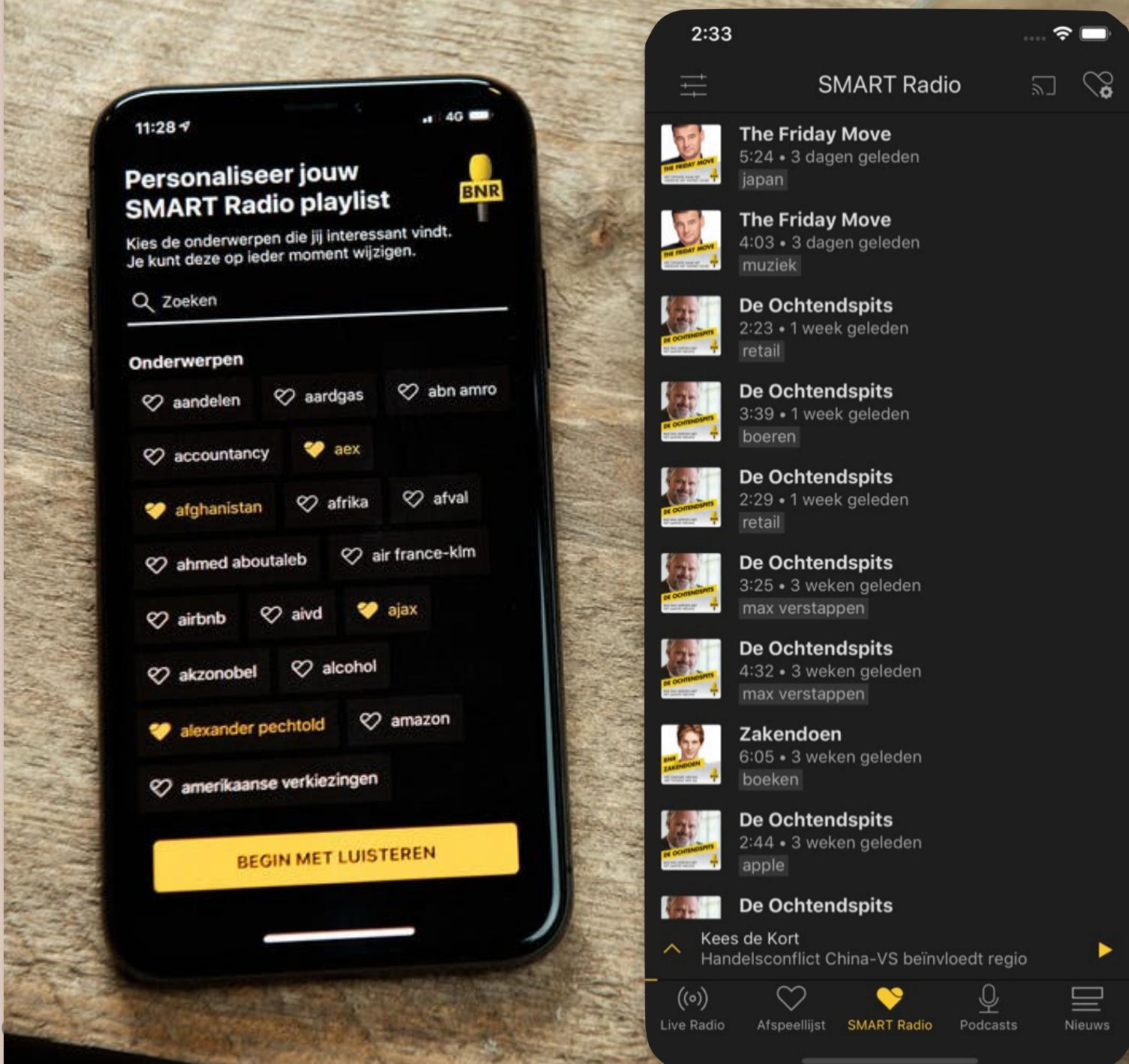


The screenshot shows a news website layout with a black header bar. The header includes the BNR logo (a yellow microphone icon), a button for "Luister nu BNR In Bedrijf", and navigation links for "Nieuws", "Programma's", and "Podcasts". Below the header, there are three news items, each with a thumbnail image, a title, and a timestamp.

- HOOGGEPLAATSTE AMERIKAANSE SPION TERUGGEHAALD UIT RUSLAND**
Amerika heeft in 2017 in het diepste geheim een hooggeplaatste spion uit Rusland gehaald. De Amerikaanse veiligheidsdiensten waren bang dat de identiteit van de spion bekend zou worden.
18:50 Geopolitiek Verenigde Staten Donald Trump Rusland
- RECHTSZAAK WILDERS KRIJGT POLITIEK STAARTJE**
Politiek Den Haag is bezorgd over de bemoeienis van justitieambtenaren met de zaak Wilders. Dat werd gisteren onthuld door RTL Nieuws. Wilders zelf doet aangifte.
18:28 Geert Wilders ChristenUnie Ministerie van Justitie en Veiligheid Rechtszaak
- JOHNSON STELT ALLES IN HET WERK VOOR BREXIT-DEAL**
Tijdens zijn bezoek aan de Ierse hoofdstad Dublin liet de Britse premier Johnson er geen twijfel over bestaan: hij wil het allerliefst dat het Verenigd Koninkrijk de Europese Unie verlaat met een akkoord tussen beide partijen.
17:47 EU Ierland Boris Johnson Verenigd Koninkrijk Brexit

Serve

- iOS/Android app



Part 3: SMART Journalism

TL;DR - A personalized summary of your news (Round 4)

SHARE: [f](#) [t](#) [G+](#) ...

NETHERLANDS | PROJECT TYPE: LARGE | HET FINANCIËLE DAGBLAD B.V. (PART OF FD MEDIAGROEP B.V.)

Summary

The TL;DR project will deliver completely personalised summaries of FD Mediagroep content, generated by an innovative machine learning/AI algorithm, helping users stay up to date with the latest news.

The solution

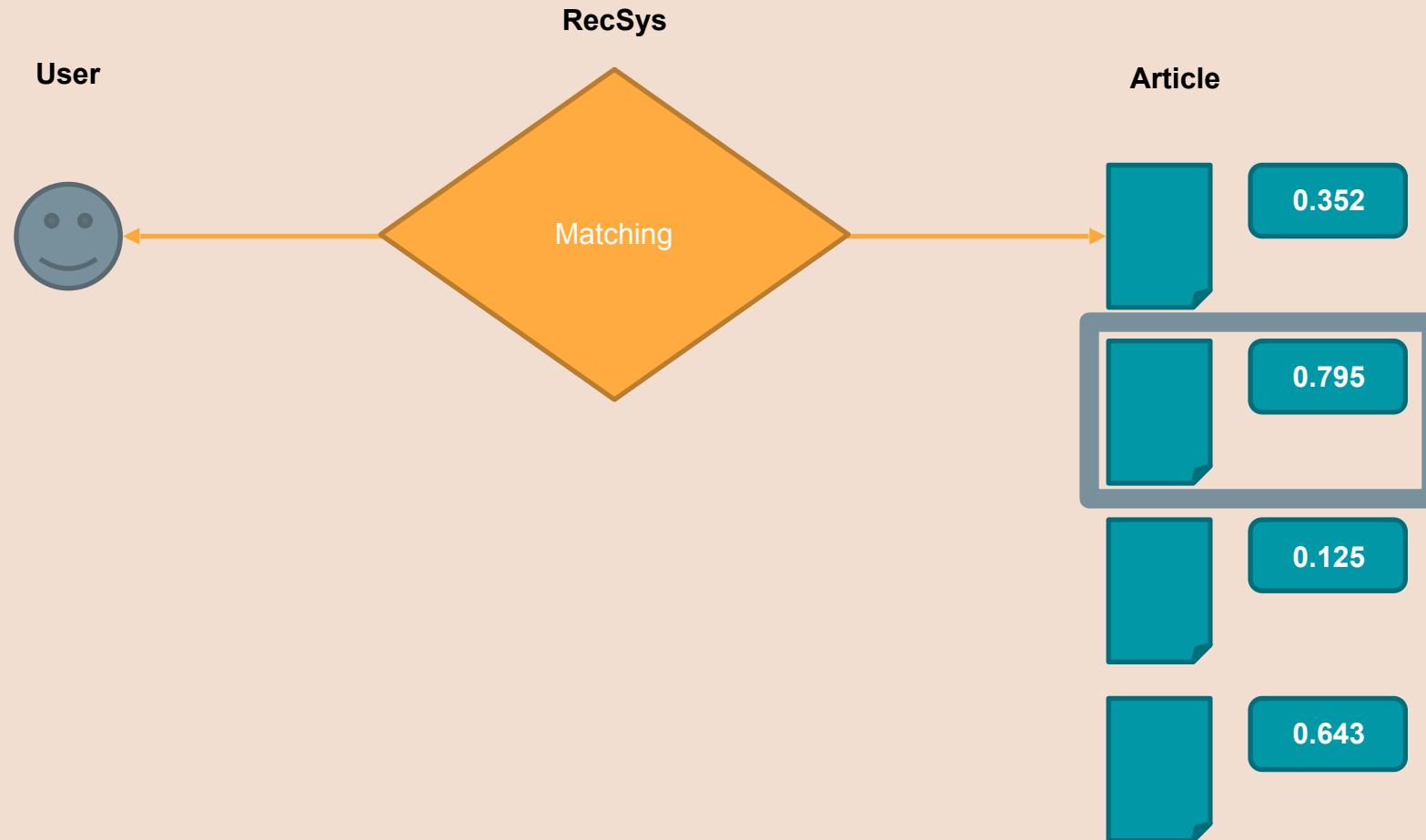
Analysing content and creating accurate summaries is still a manual task. Providing a single summary, let alone multiple summaries, for every FD article is a monumental task and would require a serious investment in time or money, which should be spent on creating more high-quality journalism. TL;DR will automate this process by building a model using the latest AI/ML techniques that analyses content to provide multiple summaries, which can be combined with our recommendation engine or help increase the availability of content. The result will be a landing page that not only contains the information you're interested in, but also presents it in a way that matches your reading style. This will be a valuable addition for current subscribers, and a new product to attract new customers looking for a quick and user-friendly update of the latest financial news.

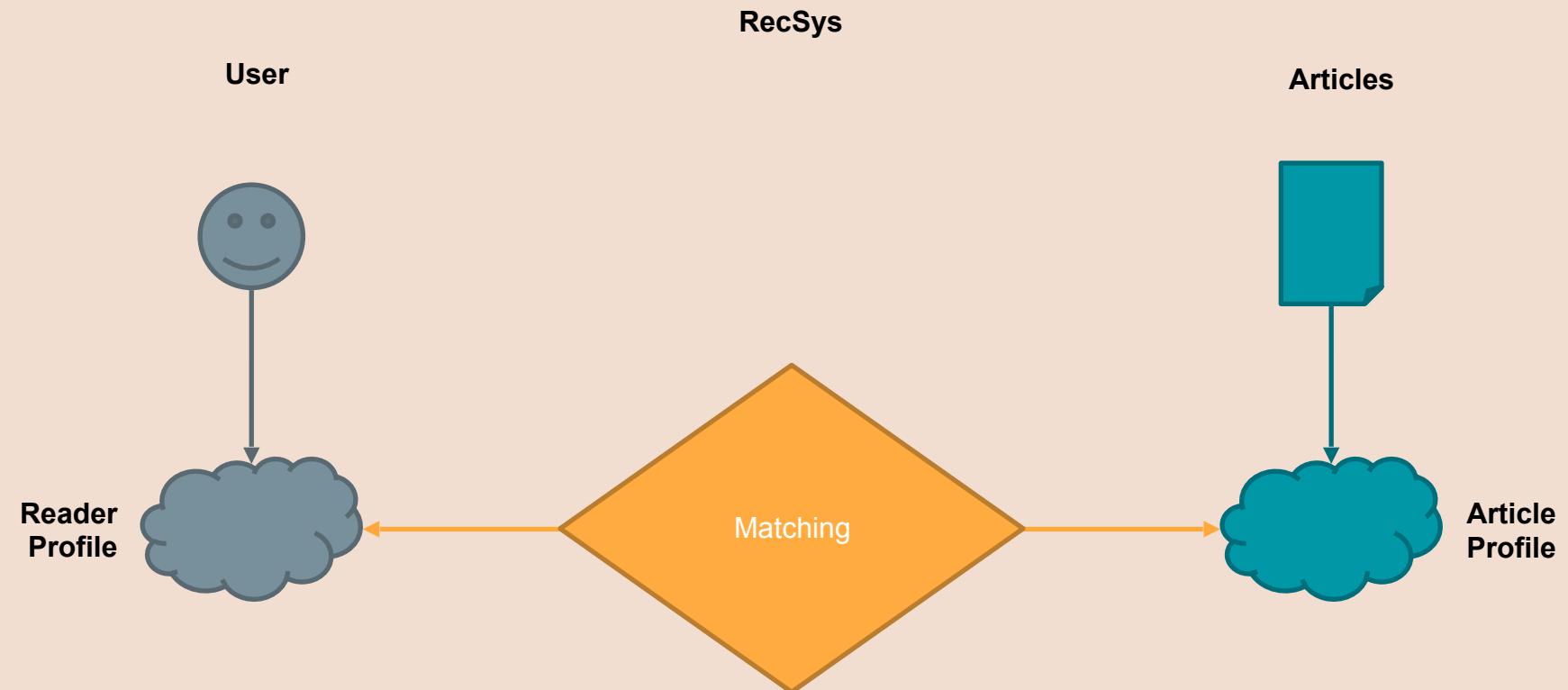


SMART Journalism

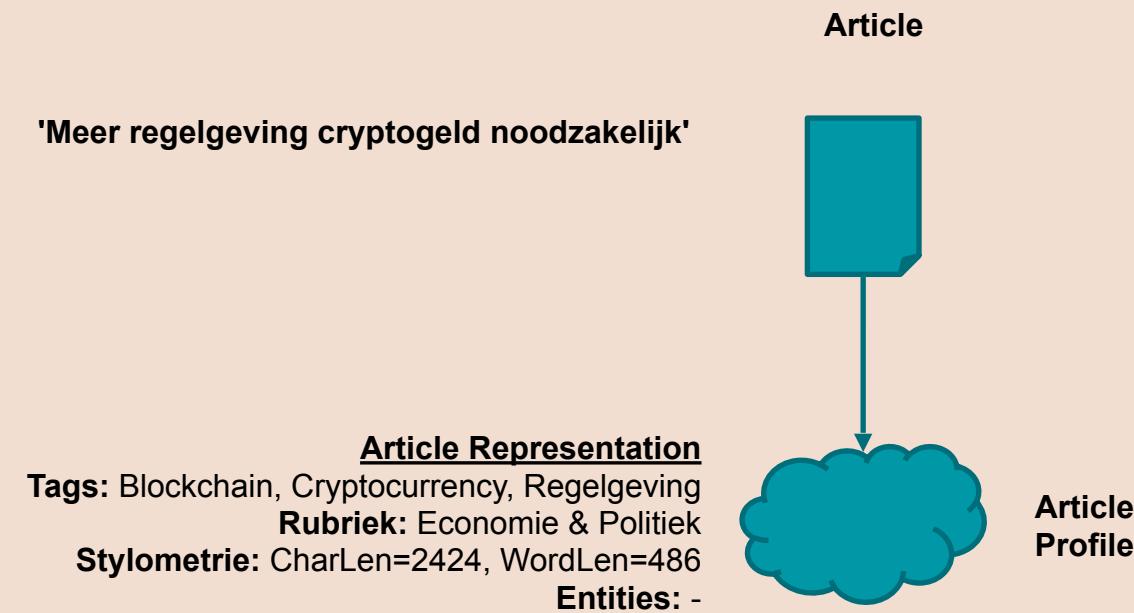
- Moonshot; personalized summarization
- How to get there:
 - Content Understanding
 - Content-based Recommender System; <user, article>
 - Personalized snippet retrieval; <user, snippet-in-article>
 - Snippet-to-summary abstractor (?)

Tech

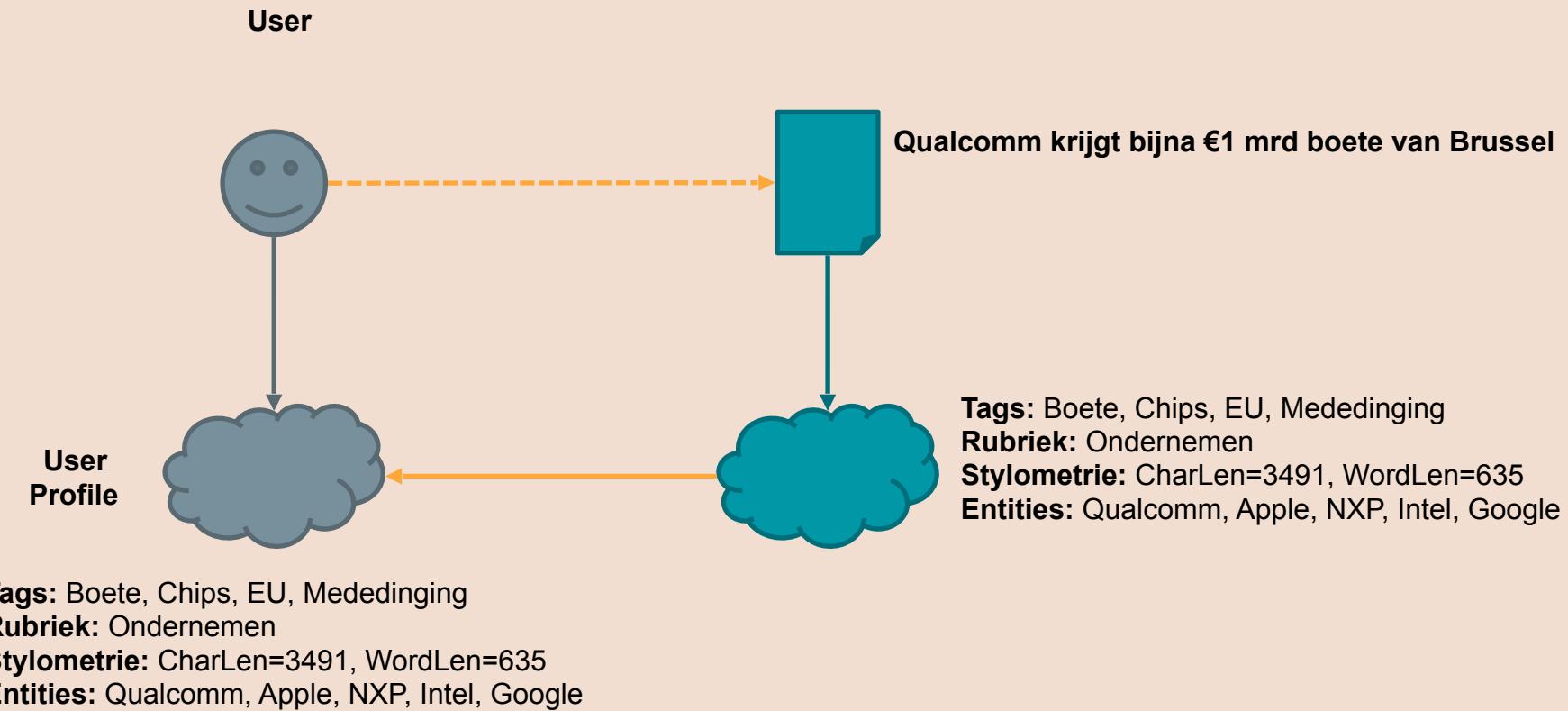


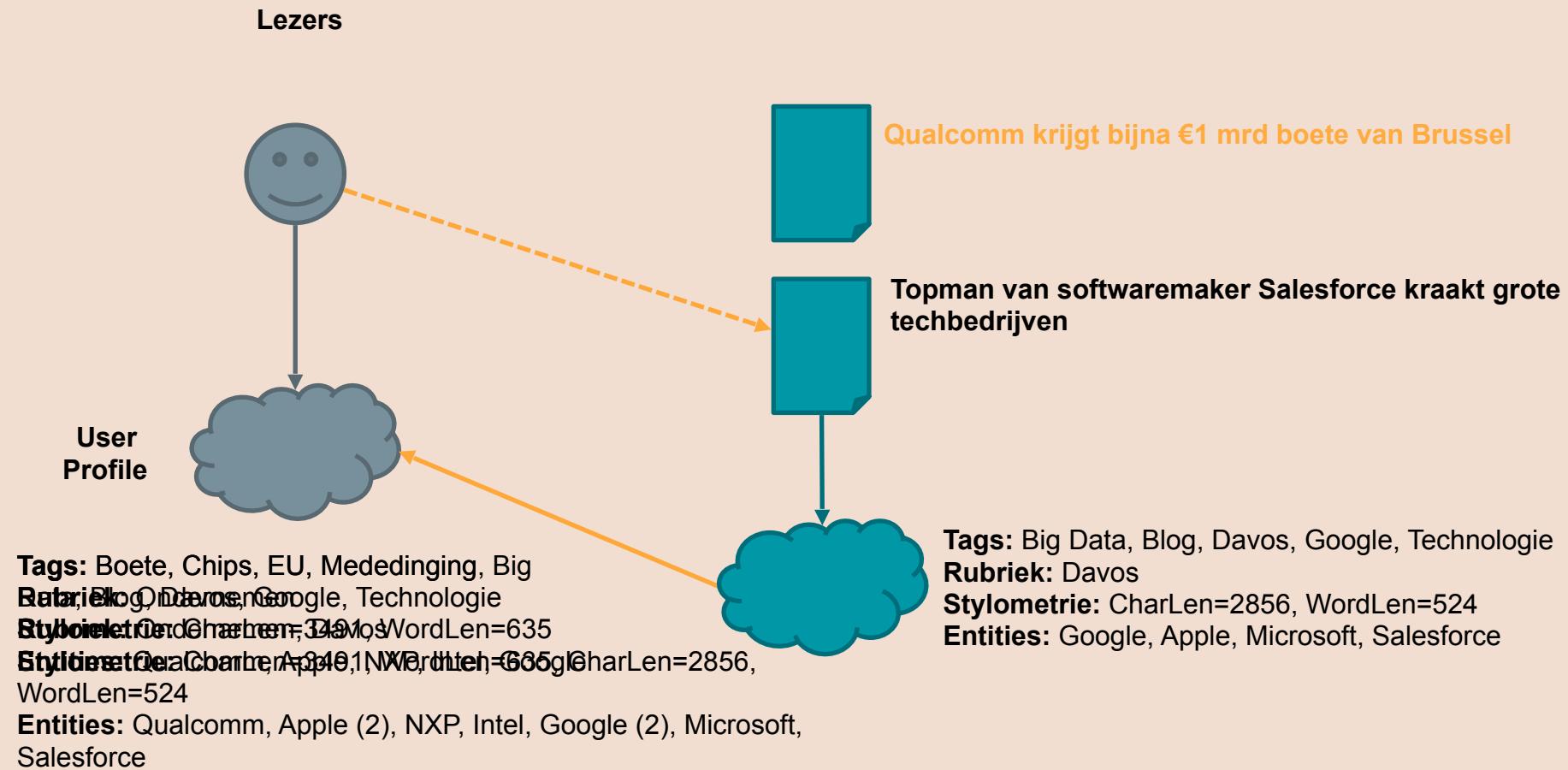


Article Representation



User Profile





Model

- Content-based RecSys
 - Using **point-wise LTR**
 - Model: **GBDT** (xgboost)
- **Labels:** clicks (i.e., click = 1, non-click = 0)
 - *Implicit feedback* (dangers and assumptions ahead)
- **Features:** user, article, user-article features ($\sim 14k$)
- **Trained** nightly

Product

Goal

- Help users to *discover content* [1]

Markten

12 min

Handel aandelenopties Euronext hervat na storing

Aandelenopties zijn door een storing niet verhandelbaar.

 [Delen](#) [Bewaren](#)

Damrak

52 min

Air France-KLM fors onderuit op vlak Damrak

ABN Amro en Altice doen het goed na een nieuw koopadvies.



Beursverslag

 [Delen](#) [Bewaren](#)

Column

57 min

De planologie van dominee Hemel

Column | Nederland moet af van de dogmatische top-downplanologie, vindt Hilde Sennema

 [Delen](#) [Bewaren](#)

Onderwerpen aanpassen

Nieuw onderwerp...

[Toevoegen](#)

Amsterdam

 [Ontvolg](#)

Automatisering

 [Ontvolg](#)

Computers

 [Ontvolg](#)

Databanken

 [Ontvolg](#)

Duurzaamheid

 [Ontvolg](#)[Toon meer](#)

Breaking nieuws alert

Ontvang een e-mail alert bij belangrijk FD Nieuws



Bewaarde artikelen

Mijn nieuws

ICT-DIENSTEN



'IT-beslisser slecht op de hoogte van belangrijke beveiligingsmaatregelen'

Bewaren

x

PROFIEL



Talent zonder geduld voor flauwekul

Bewaren

x

ICT-DIENSTEN



Bunq-oprichter wordt een van de groten in webhosting

Bewaren

x

BANKEN



Apple Pay nu ook bij Bunq en spoedig bij ABN Amro en Rabo

Bewaren

x

MILIEU EN KLIMAAT



Jonathan Safran Foer: 'Eet overdag geen vlees en kaas'

Bewaren

x

Building a RecSys for FD

- Definition of Done:

Our RecSys needs to provide relevant and useful results, and be trustworthy.

- **Relevance:** IR evaluation metrics
- **Usefulness:** What we want our RecSys to achieve
- **Trust:** Whether we see that people like/use our RecSys

1. Relevance

- Using different (ranking) metrics, e.g.;
 - NDCG
 - MAP
 - RPrec
- Evaluating models offline
- Comparing offline vs. online performance

Relevance: NDCG

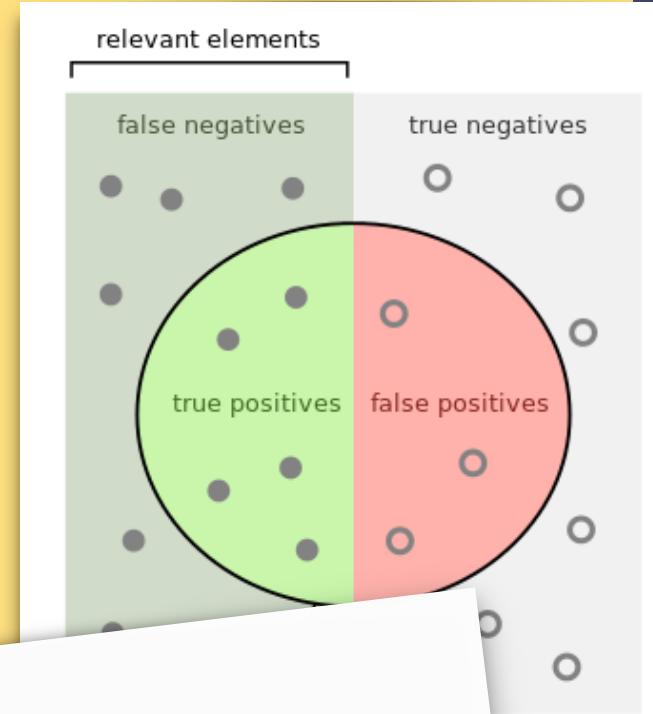
- Item's score: *Gain*
- Sum scores: *Cumulative Gain*
- Penalize scores at lower ranks: *Discounted Cumulative Gain*
 - (e.g., divide CG by log of item position)
- *Normalized* by “ideal” ranking: **NDCG**

MAP

- *Precision*: $TP / TP+FP$
- *Average Precision*: Precision at each rank
- *Mean Average Precision*: average AP across queries

$$\text{AveP} = \frac{\sum_{k=1}^n (P(k) \times \text{rel}(k))}{\text{number of relevant documents}}$$

where $\text{rel}(k)$ is an indicator function equaling 1 if the item at rank k is a relevant document, zero otherwise.^[3] Note that the average is over all relevant documents and the relevant documents not retrieved get a precision score of zero.



<https://nlp.stanford.edu/IR-book/html/htmledition/evaluation-of-ranked-retrieval-results-1.html>

R-Precision

- Precision@k
- where $k = R$ (= number of relevant documents)

Code

Issues 6

Pull requests 0

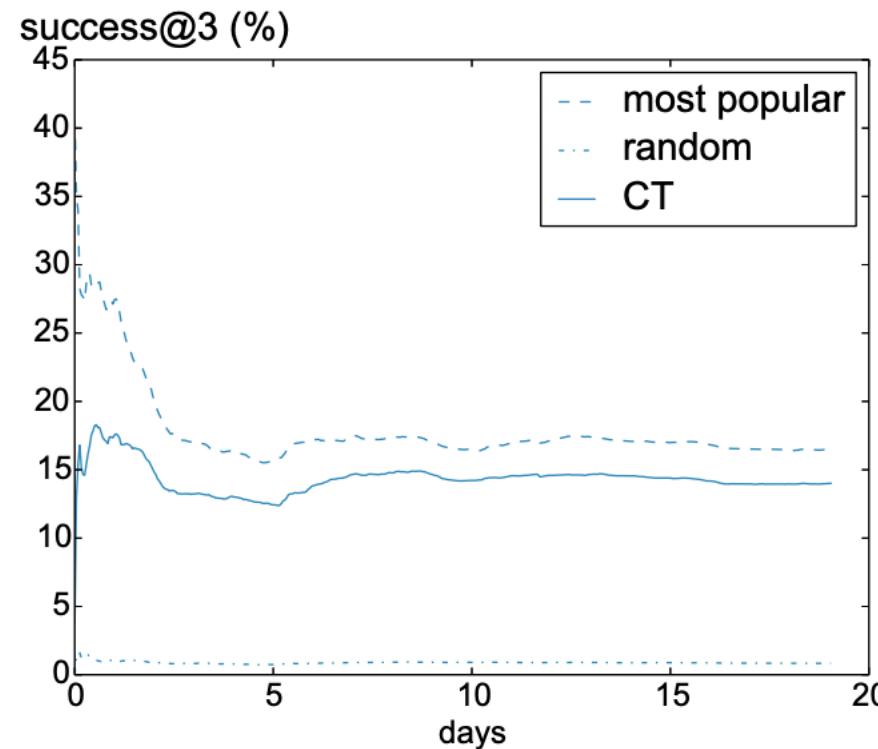
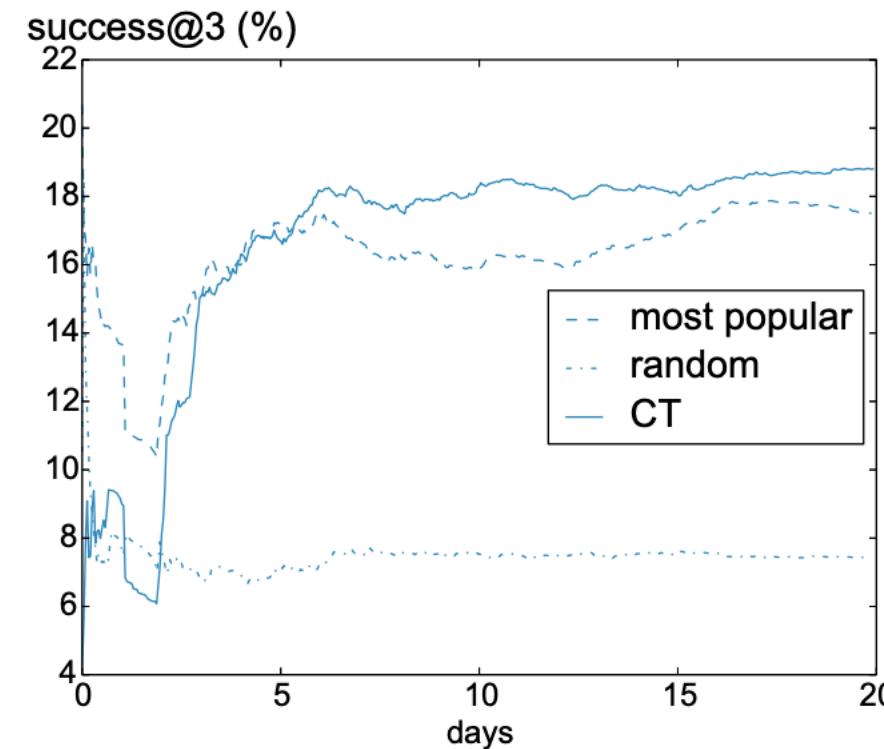
Projects 0

Wiki

Security

Insights

pytrec_eval is an Information Retrieval evaluation tool for Python, based on the popular trec_eval. <http://ilps.science.uva.nl/>

(a) Offline *predicted* accuracy(b) Online *actual* accuracy

.gitmodules

Initial commit

2 years ago

.travis.yml

Python3.7 fix (retro-compatible) (#10)

7 months ago

2. Usefulness

-  Part algorithmic
 - *What can we build?*
-  Part business/values
 - *What do we want our RecSys to do?*



Algorithmic: Tailor-made RecSys

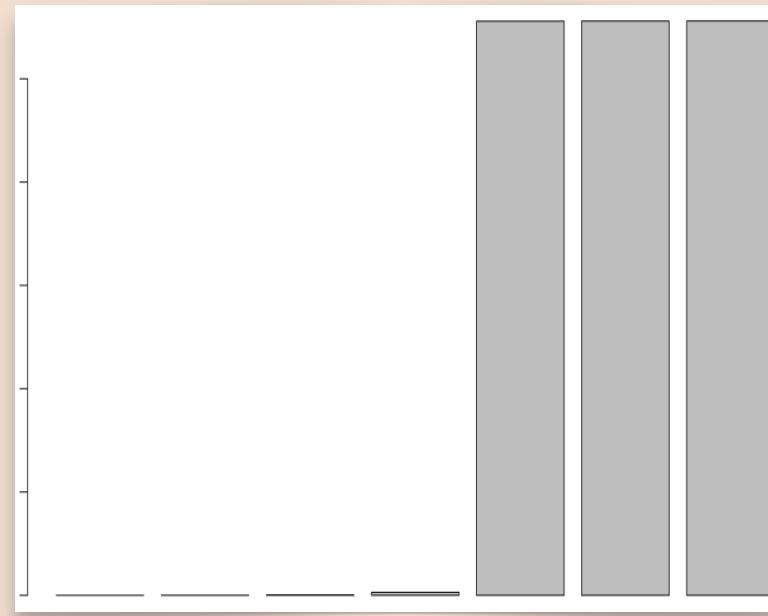
- Diversity
- Popularity
- Coverage
- Surprise
- Speed/Recency
- ...

Popularity

- *“It is generally not useful to recommend very popular items as they are generally already known by the user”* [1]
- **Presentation bias**
 - Skewed clicks
 - Issue in historic data and online evaluation

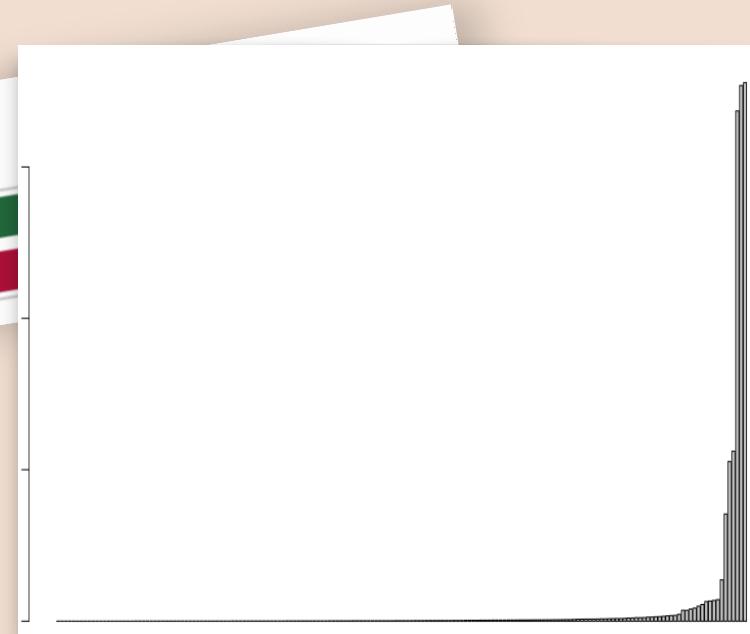
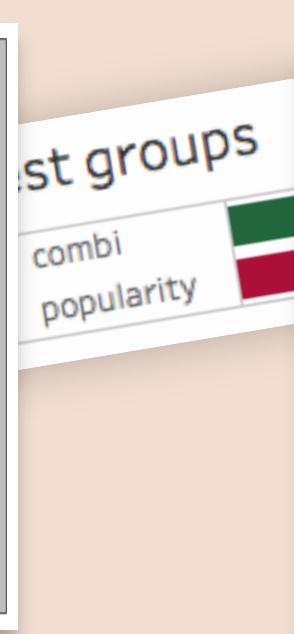
Coverage

- Unleash the long tail?



Most read:

Same 6 articles for everyone



Personalized:

178 articles



Business: *What do we want/expect?*

- **⚠ WIP** Study where journalists, data scientists, product, developers were **interviewed** to identify shared perceptions, expectations, attitudes towards algorithms.



Business: perception of algorithms

- Distinguish between **audience** & **journalism-related** values.
- Results TBD, but think of aspects such as:
 - broadness vs. personalizedness (**audience**)
 - usability (**audience**)
 - objectivity/neutralty (**journalism**)
 - speed/immediacy (**journalism**)
 - etc.



**But what about the
filter bubble?**

Short story

- Tailor-made RecSys: we are in control
- RecSys as “yet another ranking”
- The slightly longer story...



Should we worry about filter bubbles?

Frederik J. Zuiderveen Borgesius

Institute for Information Law (IViR), University of Amsterdam, Amsterdam, The Netherlands

Damian Trilling

Department of Communication Science, University of Amsterdam, Amsterdam, Netherlands

Judith Möller

Department of Communication Science, University of Amsterdam, Amsterdam, Netherlands

Balázs Bodó

*Institute for Information Law (IViR), University of Amsterdam, Amsterdam, Netherlands,
bodo@uva.nl*

Claes H. de Vreese

Department of Communication Science, University of Amsterdam, Amsterdam, Netherlands

Natali Helberger

Institute for Information Law (IViR), University of Amsterdam, Amsterdam, Netherlands

Published on 31 Mar 2016 | DOI: 10.14763/2016.1.401

Abstract: Some fear that personalised communication can lead to information cocoons or filter bubbles. For instance, a personalised news website could give more prominence to conservative or liberal media items, based on the (assumed) political interests of the user. As a result, users may encounter only a limited range of political ideas. We synthesise empirical research on the extent and effects of self-selected personalisation, where people actively choose which content they receive, and pre-selected personalisation, where algorithms personalise content for users choice. We conclude that at present there is little empirical evidence

1. Should we worry?

[We] focus on empirical evidence of the spread of personalised news services and its likely effects on political polarisation and political information.

1. Should we worry?

- It's difficult to bubble yourself
 - Both offline:
 - “Those who use a lot of partisan information also use **an above-average amount of mainstream news.**”
 - “[M]ost people by far still get their news via **traditional sources**, most notably public-service television.”
 - And online:
 - “People who choose personalisation are **more likely to use an above-average amount of general-interest news** as well.”
 - “A recent study suggests that the **influence of [the Facebook] algorithm is lower than the influence of the user’s choices.**”

1. ! Take homes

- “[T]here is **no empirical evidence** that warrants any strong worries about filter bubbles.”
- “One lesson we should have learned from the past is that **panic does not lead to sane policies**. More **evidence is needed on the process and effects of personalisation**, so we can shift the basis of policy discussions from fear to insight.”

2. Some empirical evidence

- On average, personalization is...
due to
- Varies widely
ing.
- Only found in results
of searching with a logged
in user.

Measuring Personalization of Web Search

ANIKÓ HANNÁK, Northeastern University

PIOTR SAPIĘŻYŃSKI, Technical University of Denmark

ARASH MOLAVI KAKHKI, Northeastern University

DAVID LAZER, ALAN MISLOVE, and CHRISTO WILSON, Northeastern University

Web search is an integral part of our daily lives. Recently, there has been a trend of personalization in Web search, where different users receive different results for the same search query. The increasing level of personalization is leading to concerns about *Filter Bubble* effects, where certain users are simply unable to access information that the search engines' algorithm decides is irrelevant. Despite these concerns, there has been little quantification of the extent of personalization in Web search today, or the user attributes that cause it.

In light of this situation, we make three contributions. First, we develop a methodology for measuring personalization in Web search results. While conceptually simple, there are numerous details that our methodology must handle in order to accurately attribute differences in search results to personalization. Second, we apply our methodology to 200 users on Google Web Search and 100 users on Bing. We find that, on average, 11.7% of results show differences due to personalization on Google, while 15.8% of results are personalized on Bing, but that this varies widely by search query and by result ranking. Third, we investigate the user features used to personalize on Google Web Search and Bing. Surprisingly, we only find measurable personalization as a result of searching with a logged in account and the IP address of the searching user. Our results are a first step towards understanding the extent and effects of personalization on Web search engines today.

General Terms: Design, Measurement

Additional Key Words and Phrases: Internet Filter Bubble, Personalization

1 INTRODUCTION

Web search services like Bing and Google Web Search (Google Search) are an integral part of our daily lives; Google Search alone receives 17 billion queries per month from U.S. users [8]. People use Web search for a number of reasons, including finding authoritative sources on a topic, keeping abreast of breaking news, and making purchasing decisions. The search results that are returned, and their order, have significant implications: ranking certain results higher or lower can dramatically affect business outcomes (e.g., the popularity of search engine optimization services), political elections (e.g., U.S. Senator Rick Santorum's battle with Google [60]), and foreign affairs (e.g., Google's ongoing conflict with Chinese Web censors [67]).

Recently, major search engines have implemented *personalization*, where the Web search operator modifies the results—or their order—based on the user who is making the query [19, 56]. As previous work has noted [42], an effective personalized search engine is able to decide autonomously whether or not a user is interested in a specific webpage and, if so, display that result at a higher rank. For example, users searching for “pizza” in New

This article is an extension of the paper “Measuring Personalization of Web Search”, published at WWW 2013 [25]. This article extends the original by adding measurement results from Bing and DuckDuckGo, as well as adding several new experimental treatments not found in the conference paper.

2. Method

1.

1. Get 200 volunteers with Google accounts
2. Have them issue the same set of queries
3. Compare results

2.

1. Construct Google bot accounts
 - Vary aspects such as location, demographics, click behavior, browsing + search history, etc.
2. Have them issue the same set of queries
3. Compare results

2. Findings

- On average, **11.7% of results show differences due to personalization on Google.**
- **Top ranks tend to be less personalized than bottom ranks.**

2. Findings

-  A great deal of **personalization based on location** (especially for company names, where users received different store locations).
-  The least personalized results tend to be factual and health related queries.

Most Personalized	Least Personalized
gap hollister hgtv boomerang home depot greece pottery barn human rights h2o nike	what is gout dance with dragons what is lupus gila monster facts what is gluten ipad 2 cheri daniels psoriatic arthritis keurig coffee maker maytag refrigerator

2. Findings

- ✓ Logged in vs. “cleared cookies” account
- ✓ Geolocation
- ✗ Gender
- ✗ Age
- ✗ Search history
- ✗ Click history
- ✗ Browsing history

3. Bursting bubbles

Symposium Article

Promoting Civil Discourse Through Search Engine Diversity

Social Science Computer Review
2014, Vol. 32(2) 145-154
© The Author(s) 2013
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0894439313506838
ssc.sagepub.com



Elad Yom-Tov¹, Susan Dumais², and Qi Guo³

Abstract

The ability of modern web services such as news aggregators and search engines to tailor their results to the tastes of individuals, together with people's preference for reading opinions which reinforce their own viewpoints, have raised concerns that people are nowadays exposed to a narrow range of viewpoints, a phenomenon referred to as the "filter bubble". In this paper we focus on increasing exposure to varied political opinions with a goal of improving civil discourse. We develop a method to algorithmically encourage people to read diverse political opinions and test it when people actively seek information. First, analyzing data from a popular search engine we show that people are indeed more likely to read opinions consistent with their own. Interestingly, they are more likely to read news from opposing sites when the language model of a particular news item is close to the language model of their own political leaning. Based on this finding, we describe a method for assisting people to read divergent opinions by choosing documents of opposing viewpoints that have a language model closer to their own language model. We test our method on a number of web searchers and show that pages of the opposing side which were more similar than the average persons' own language model tended to be clicked 38% more than those below. We also describe the long-term effects of our method, showing that people who were shown more diverse results continued reading more diverse results and overall became more interested in news.

Keywords

search engines, filter bubble, selective exposure, political discourse, diversity, information retrieval, media bias

Introduction

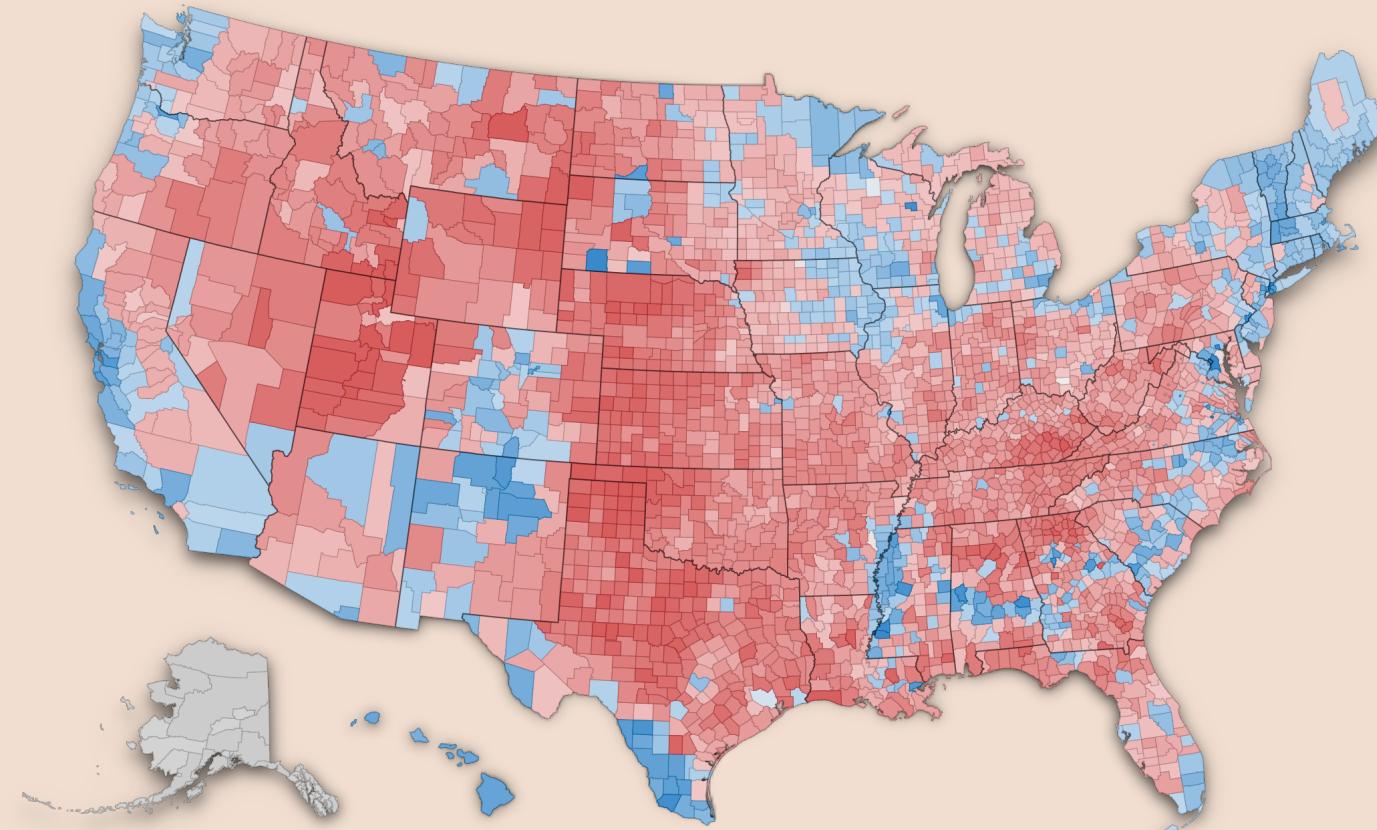
Modern web services, including search engines, increasingly tailor their results to individuals, attempting to match information to the perceived preferences of each person according to their characteristics and past behavior. While this tailoring can be useful for users, it can also lead to a "filter bubble", where individuals are exposed only to information that reinforces their existing beliefs and viewpoints, and are less likely to encounter information that challenges them or presents alternative perspectives. This lack of exposure to diverse viewpoints can limit people's understanding of different issues and can contribute to polarization and division in society. One way to address this issue is to promote search engine diversity, which involves presenting users with a wider range of search results from different sources and perspectives. This can help users to gain a more balanced and comprehensive understanding of the topics they are interested in, and can encourage them to engage in more civil and constructive discourse. The goal of this paper is to explore the potential of search engine diversity as a tool for promoting civil discourse and improving political discourse.

3. Aim

*Increase exposure to varied political opinions
with a goal of improving civil discourse*

3. Method

- Classify searchers into *political leaning* (using geo data)



3. Method

- Infer *political leaning* of news sources from user behavior.
- Identify polarized search queries (with strong *political leanings* — in both directions).

Table 3. Queries With the Highest and Lowest Leaning Score, Grouped by Subject Matter.

Republican	Democratic
what is the tea party movement, what is the tea party, tea party, tea party movement	the tea party movement
obamacare, death panels, patient protection and affordable care act, affordable health care act of 2010, the affordable care act, affordable health care act	affordable care act summary, death panel, obama care summary, ppaca
Obama nobel prize, obama nobel peace prize	barack obama nobel peace prize, obama peace prize
stimulus package, recovery act, American recovery and reinvestment act	american recovery and reinvestment act of 2009, 2009 stimulus package, arra
2010 census	
war on women	
	journolist
	bowles-simpson
	shirley sherrod
	Shirley Sherrodsh for clunkers

3. Method

- **Treatment group:** Insert **red** results for **blue** users, and **blue** results for **red** users
- **Control group:** Do not adjust results

3. Methode

1. Short term: Compare clicks/behavior between control & treatment.
2. Long term: Measure during two weeks, per user;
 1. **Polarization:** Difference of user's *leaning*-score compared to average *leaning* across all sources.
 2. **Engagement:** Average number of queries + average read articles.

3. Findings 1

- Less clicks on inserted *opposing* sources.
- But:

“Results pages of the opposing viewpoint which had a similarity higher than the average tended to be clicked 38% more than those below the average.”

4. Findings 2

- **Polarization:**
 - *Treatment*: Average *leaning* ‘moves’ ~25% to centre
 - *Control*: Negligable difference (~1%)
- **Engagement:**
 - *Treatment*: Number of queries: +9% / articles read: +4%
 - *Control*: Small reduction in both (~2.5%)

4. How do algorithmic recommendations compare to human ?

INFORMATION, COMMUNICATION & SOCIETY, 2018
<https://doi.org/10.1080/1369118X.2018.1444076>

 Routledge
Taylor & Francis Group

 OPEN ACCESS 

Do not blame it on the algorithm: an empirical assessment of multiple recommender systems and their impact on content diversity

Judith Möller^a, Damian Trilling^a, Natali Helberger^b and Bram van Es^c

^aDepartment of Communication Science, University of Amsterdam, Amsterdam, Netherlands; ^bInstitute for Information Law, University of Amsterdam, Amsterdam, Netherlands; ^ceScience Center, Amsterdam, Netherlands

ABSTRACT

In the debate about filter bubbles caused by algorithmic news recommendation, the conceptualization of the two core concepts in this debate, diversity and algorithms, has received little attention in social scientific research. This paper examines the effect of multiple recommender systems on different diversity dimensions. To this end, it maps different values that diversity can serve, and a respective set of criteria that characterizes a diverse information offer in this particular conception of diversity. We make use of a data set of simulated article recommendations based on actual content of one of the major Dutch broadsheet newspapers and its users ($N=21,973$ articles, $N=500$ users). We find that all of the recommendation logics under study proved to lead to a rather diverse set of recommendations that are on par with human editors and that basing recommendations on user histories can substantially increase topic diversity within a

ARTICLE HISTORY

Received 30 October 2017
Accepted 20 February 2018

KEYWORDS

News; recommender systems; diversity metrics; filter bubbles; automated content classification

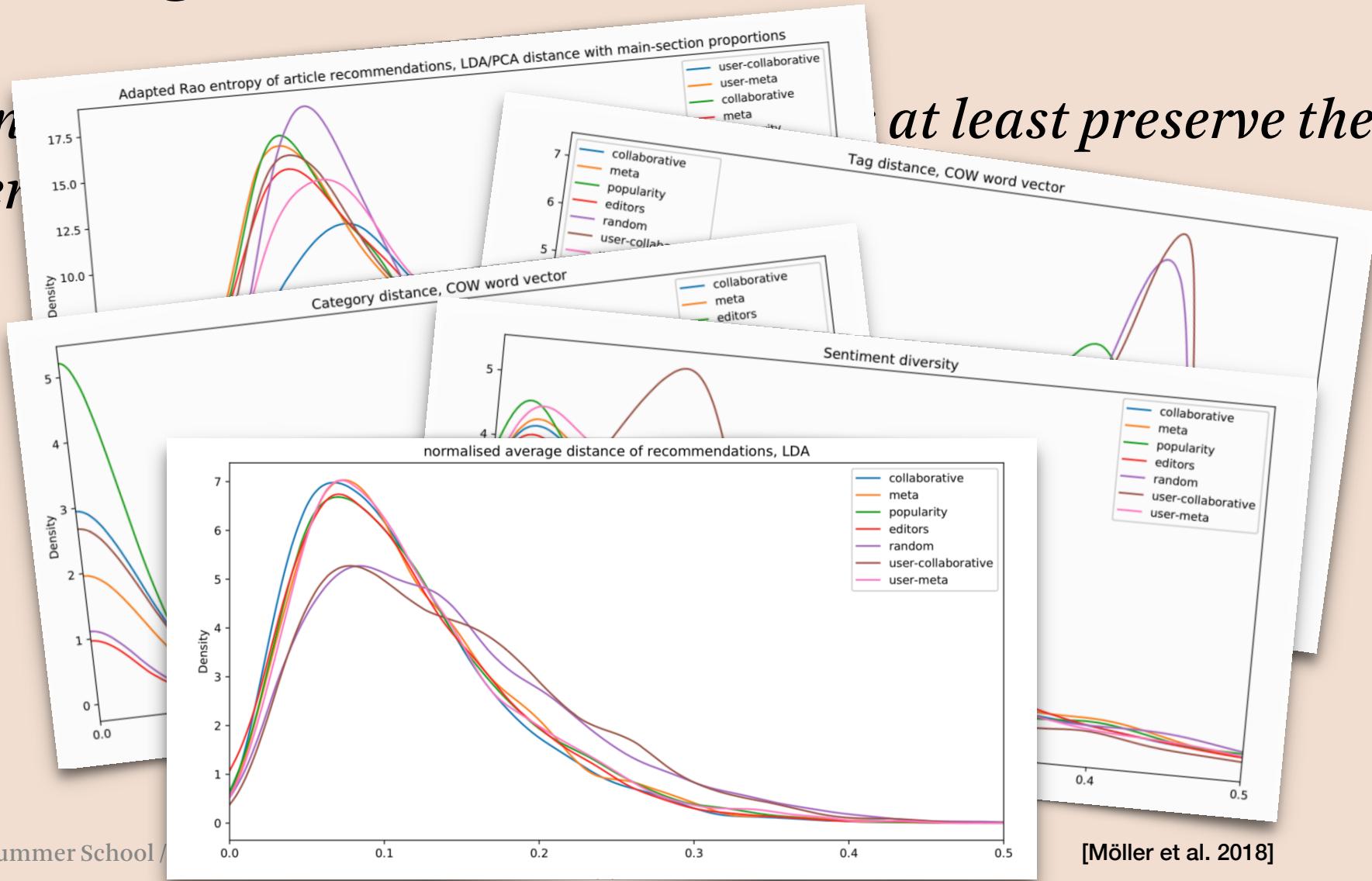
4. Method

-  Generate article recommendations for news articles using different (off the shelf) recommender systems (CF & CB).
-  Compare to hand-picked article recommendations.
- Measure “diversity” of recommended articles:
 - At content level
 - At tag level
 - At category level
 - At sentiment/subjectivity level

4. Findings

“Conven-
topic/ser-

at least preserve the



4. ! Take home

- Diversity is preserved with conventional recommender systems.

/end of rant

⚠️ Besides

- It is trivial to model and incorporate aspects such as diversity (as you'll (hopefully) learn during our hands-on 🙌...))

3. Trust

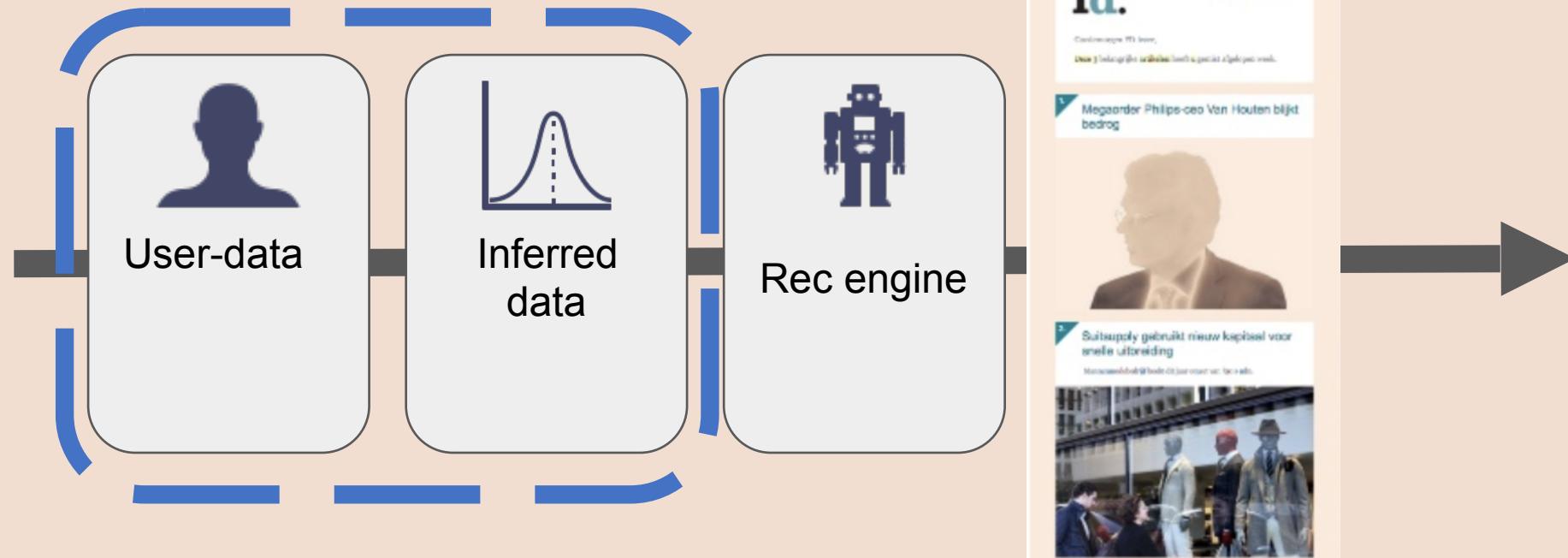
- Returning visitors / users
- Measure CTR
- Questionnaires [1]

Explaining user profiles

Why fair and transparent?

- User studies have shown:
 - Our users want personalized content
 - Our users care for transparency
- FD
 - Verifiability is one of the core values for FD Mediagroup
 - Transparency enables verifiability

The Context



T

Reading News with a Purpose: Explaining User Profiles for Self-Actualization

Emily Sullivan*, Dimitrios Bountouridis*, Jaron Harambam[†], Shabnam Najafian*, Felicia Loescherbach^{††},
 Mykola Makhortykh[‡], Domokos Kelen^{‡‡}, Daricia Wilkinson[¶], David Graus^{**}, Nava Tintarev*

*Delft University of Technology, The Netherlands

[†]Institute for Information Law, University of Amsterdam, The Netherlands

[‡]Amsterdam School of Communication Research, University of Amsterdam, The Netherlands

^{††}Vrije University, Amsterdam, The Netherlands

^{‡‡}Institute for Computer Science and Control, Hungarian Academy of Sciences, Hungary

[¶]Clemson University, Clemson, USA

^{**}FD Mediagroep

e.e.sullivan-mumm@tudelft.nl

ABSTRACT

Personalized content provided by recommender systems is an integral part of the current online news reading experience. However, news recommender systems are being criticized for their 'black-box' approach to data collection and processing, and for their lack of explainability and transparency. This paper focuses on explaining user profiles constructed from aggregated reading behavior data, used to provide content-based recommendations. By doing so, the paper makes a first step toward consolidating epistemic values of news providers and news readers. We present an evaluation of an explanation interface reflecting these values, and find that providing users with different goals for self-actualization (i.e., *Broaden Horizons* vs. *Discover the Unexplored*) influences their reading intentions for news recommendations.

KEYWORDS

explainability, user profile, user control, self-actualization, news recommender systems

ACM Reference Format:

Emily Sullivan, Dimitrios Bountouridis, Jaron Harambam, Shabnam Najafian, Felicia Loescherbach, Mykola Makhortykh, Domokos Kelen, Daricia Wilkinson, David Graus, & Nava Tintarev. 2019. Reading News with a Purpose: Explaining User Profiles for Self Actualization. In *Proc. of 27th Conf. on User Modeling, Adaption & Personalization Adjunct (UMAP'19 Adjunct)*, June 9–12, 2019, Larnaca, Cyprus. ACM, NY, NY. 5 pages.
<https://doi.org/10.1145/3314183.3323456>

1 INTRODUCTION

Personalized experiences powered by recommender systems permeate our lives. However, the rise of distrust and skepticism around the collection and use of personal data generates an increased interest in *explainability* and *transparency* of black-box recommender systems [12, 13, 31]. Many works [14, 19, 35] show that explanations are integral for users to understand their consumption preferences and achieve their *epistemic goals* i.e., goals for knowledge development. Moreover, researchers have begun exploring methods that support the exploration and understanding of users' current, aspirational, and self-actualized goals to provide transparency in recommender systems [11, 22]. Yet, the task of opening the black box of recommender systems remains notoriously hard to achieve [8, 28]. For the domain of *online news*, this might limit the readers' control over information diets [15]. Limited explainability can also diminish users' engagement with recommended content [21], as well as raising concerns about the societal role of news media [9] and consequences of increased algorithmization [26, 33].

To address these worries, this paper focuses on the goal of explaining recommendations in the context of *online news*. We explore methods of explaining one aspect of how a content-based recommenders work: the *user profile*. We aim to automatically summarize and visualize the recommender's high dimensional internal representations of users. These profiles are automatically generated from their reading behavior, through leveraging news article's topics, entities, and tags. Explaining user profiles have been demonstrated to be effective [3], particularly when

STEP 1

DISCOURSE
Understanding the problem

STEP 2

FRAMEWORK
Systematic layering of explanations

STEP 3

EVALUATION
Data exploration and evaluation

STEP 1

DISCOURSE
*Understanding the
problem*

STEP 2

FRAMEWORK
*Systematic
layering of
explanations*

STEP 3

EVALUATION
*Data exploration
and evaluation*

Discourse

- Who is the target audience of the explanations?
- What is the goal?
- What purpose do the explanations serve?

STEP 1

DISCOURSE
Understanding the problem

STEP 2

FRAMEWORK
Systematic layering of explanations

STEP 3

EVALUATION
Data exploration and evaluation

Framework

Goal

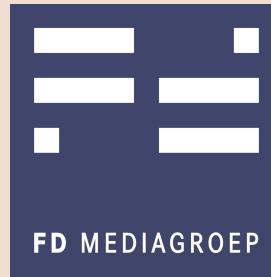


Transparency

Answers the question

*“What is my interaction
with the system so I can
understand my past?”*

Framework



Values

Broadness, diversity, autonomy, objectivity,
match with the user needs, controllability



I want to
be an expert



I want to
stay informed



I want to
broaden my
horizon



I want to
discover the
unexplored

STEP 1

DISCOURSE
Understanding the problem

STEP 2

FRAMEWORK
Systematic layering of explanations

STEP 3

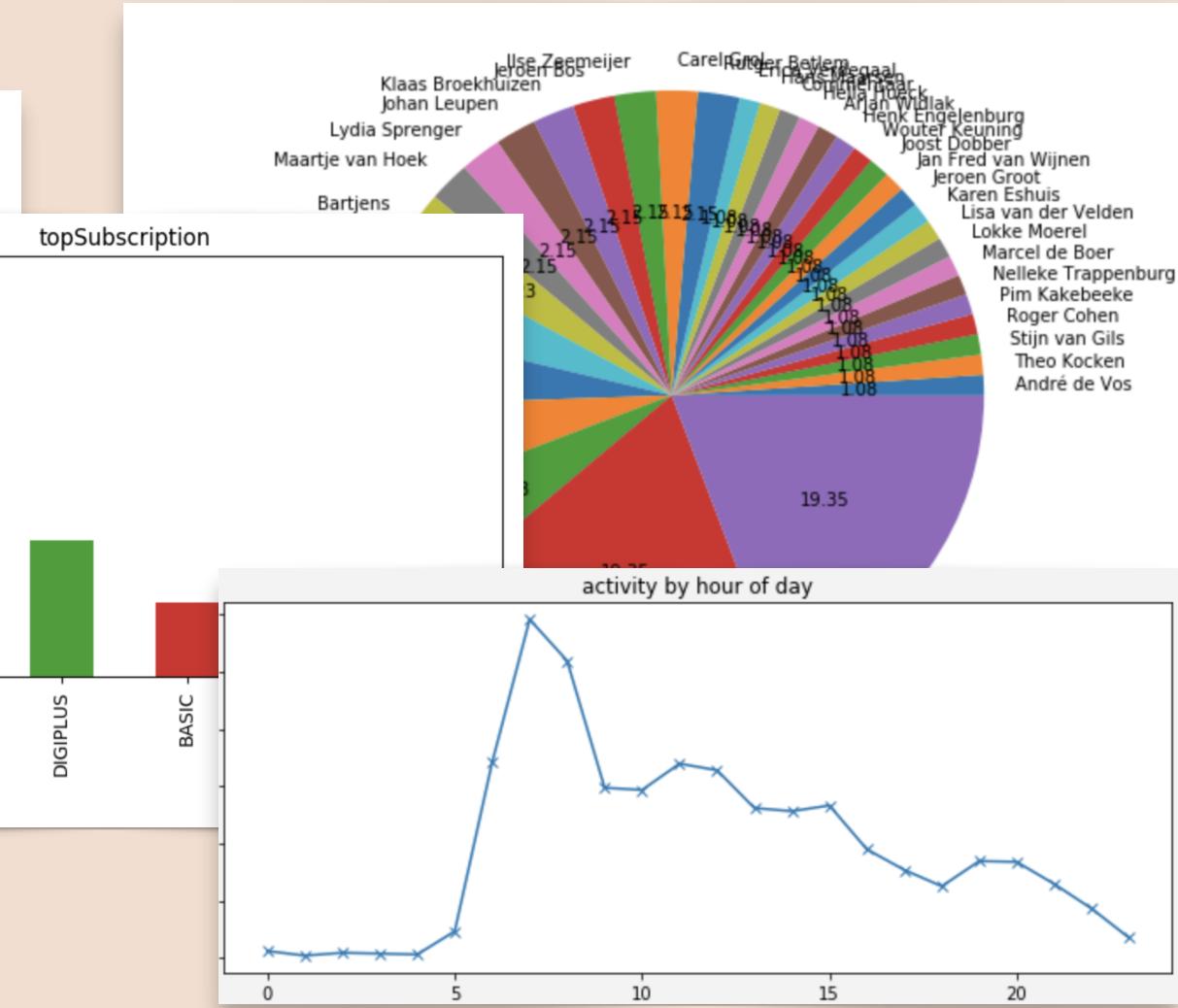
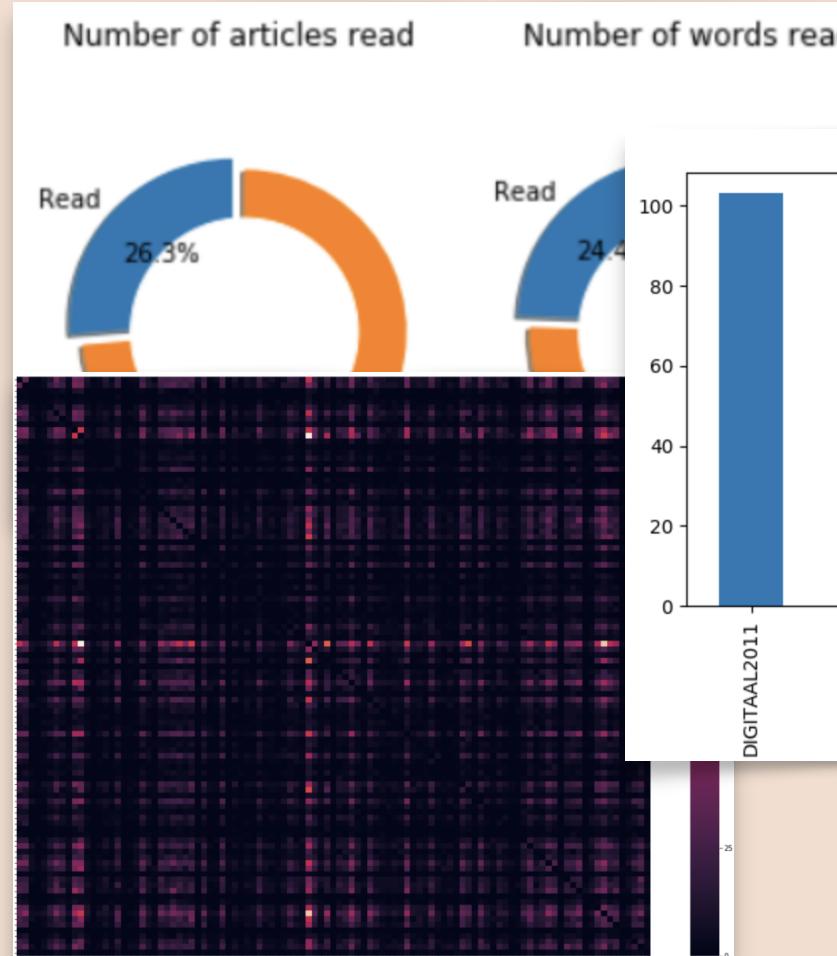
EVALUATION
Data exploration and evaluation

Data exploration

Data exploration



Data exploration



Data exploration

article ID	supertag	tag 1	tag 2	tag 3	tag 4	tag 5
901315	blog					
1097464	achtergrond	strategie		overname	zuid-	
1098018	opinie politiek		opinie	roland van der vorst		
1100737	personal finance					
1106152	opinie verenigde staten			opinie fraude	wetge	
1111744	zomerserie					
1125210	alternatief	strategie		investeringen	zuid-	
1126839	fd gazellen	technologie		werkgelegenheid	innov	
1142007	interview	verenigde staten		aandelen		
1147378	achtergrond	belasting	china	panama papers		
1156193	#brexitssongs	verenigde staten		eu	brexi	
1157570	interview	raad van bestuur		boskalis		
1159722	opinie industrie		opinie energie	duurzaamheid		
1161567	in beeld					
1164052	bartjens	beleggingsanalist		scheepvaart		
1186413	bartjens	luchtvaart		personenvervoer	ierla	
1189318	blog	technologie	innovatie		financieel	
1215672	opinie arbeidsmarkt	opinie onderwijs		wetge		
1220962	update aandelen		effectenbeurs	investeringen		
1227333	fd gazellen	technologie	zuid-holland	fd ga		
1228326	analyse farmacie		tweede kamer	analyse minis		
1234179	achtergrond	accountancy	toezicht		regel	
1234577	achtergrond	china	azië	achtergrond	kred	
1239514	vastgoed					
1241514	column aandelen		column dranken	bartjens		
1244258	column aandelen		effectenbeurs	column		
1250059	analyse aandelen		overname	accell	fiets	
1251524	opinie opinie	infrastructuur	luchtvaart		milie	
1252984	opinie opinie	innovatie		toezichthouder	farm	
1254287	interview	interview		toerisme	reist	

2. Should I Stay, or Should I Go? – The Clash (1982, Epic, hoogste positie 17)

Britse punkband zag het nationale dilemma in de jaren tachtig al aankomen.

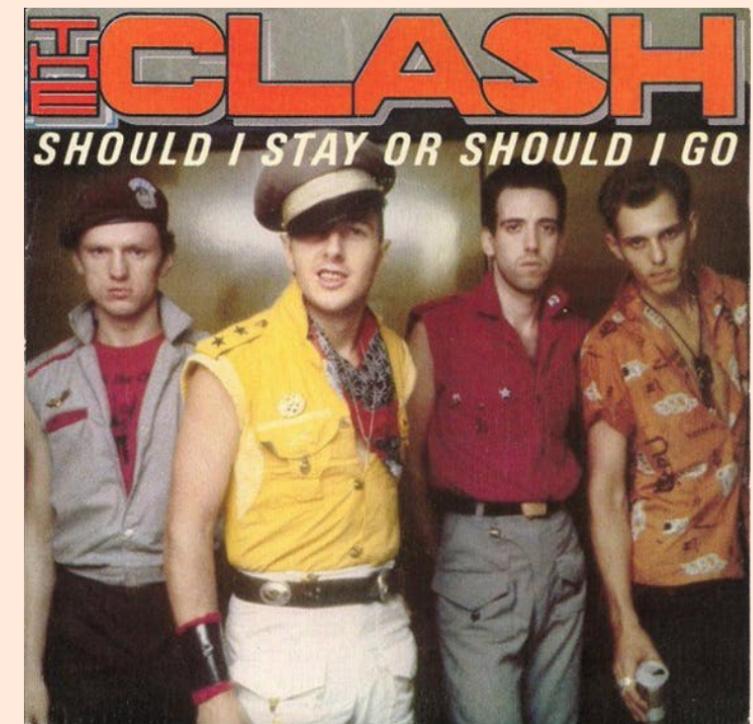
Should I stay or should I go now?

Should I stay or should I go now?

If I go there will be trouble

And if I stay it will be double

So come on and let me know



3. With or without you - U2 (1987, Island Records, hoogste positie 4)

User study

User study



System
Evaluation



Your goal is to Broaden your Horizons.

There may be topics you do not normally read about, but you may actually find interesting. Exploring this helps to build a broad perspective on the issues that matter to you.



Your goal is to Discover the Unexplored.

There may be topics that you haven't explored before that may actually become new interests. Exploring new topics can promote creativity and objectivity.

User study



System
Evaluation



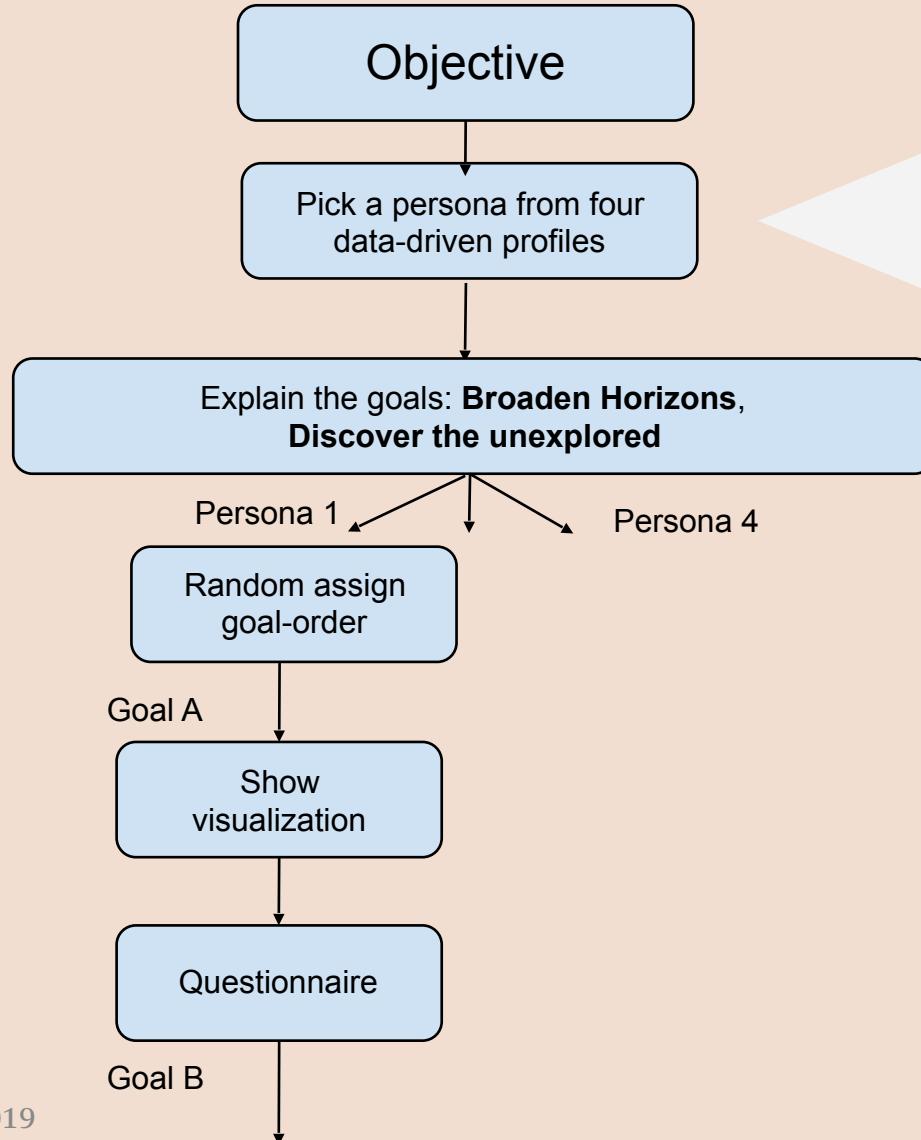
Aim:

Study whether being offered an **explanation of reading behavior** and a **particular goal**, would influence the user's **intended reading behavior**

User study



System
Evaluation



Circle color:

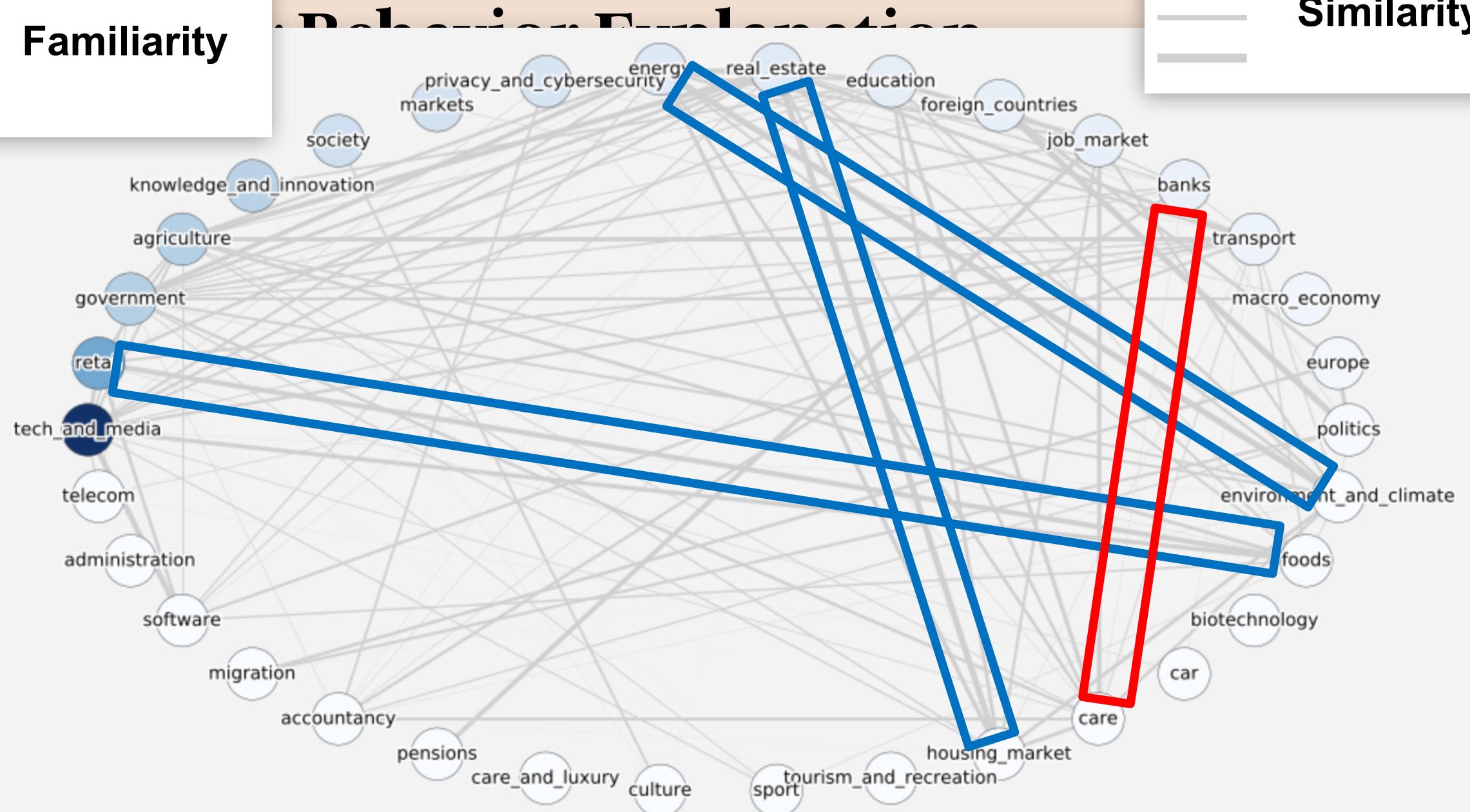


Line thickness:



Familiarity

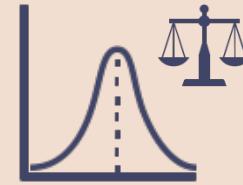
Similarity



Hypotheses



System
Evaluation



1. Goal Framework:

Broaden Horizon chosen topics are **more similar** and **more familiar** than Discover the unexplored



2. Broaden horizon:

Users choose topics that have **high similarity** and **high familiarity** compared to the non-selected topics.



3. Discover the unexplored:

Users choose topics that have **low similarity** and **low familiarity** compared to the non-selected topics.

Results

(Executive summary)

Results



System
Evaluation



1. Goal Framework:

Broaden Horizon chosen topics are **more similar** and **more familiar** than Discover the unexplored



2. Broaden horizon:

Users choose topics that have **high similarity** and **high familiarity** compared to the non-selected topics.



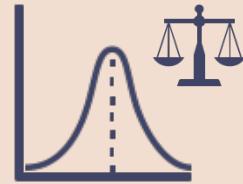
3. Discover the unexplored:

Users choose topics that have **low similarity** and **low familiarity** compared to the non-selected topics.

Results



System
Evaluation



1. Goal Framework:

Broaden Horizon chosen topics are **more similar** and **more familiar** than Discover the unexplored



2. Broaden horizon:

Users choose topics that have **high similarity** and **high familiarity** compared to the non-selected topics.



3. Discover the unexplored:

Users choose topics that have **low similarity** and **low familiarity** compared to the non-selected topics.

Results



System
Evaluation



1. Goal Framework:

Broaden Horizon chosen topics are **more similar** and **more familiar** than Discover the unexplored



2. Broaden horizon:

Users choose topics that have **high similarity** and **high familiarity** compared to the non-selected topics.



3. Discover the unexplored:

Users choose topics that have **low similarity** and **low familiarity** compared to the non-selected topics.

Results



System
Evaluation



1. Goal Framework:

Broaden Horizon chosen topics are **more similar** and **more familiar** than Discover the unexplored



2. Broaden horizon:

Users choose topics that have **high similarity** and **high familiarity** compared to the non-selected topics.



3. Discover the unexplored:

Users choose topics that have **low similarity** and **low familiarity** compared to the non-selected topics.

Results



System
Evaluation



1. Goal Framework:

Broaden Horizon chosen topics are **more similar** and **more familiar** than Discover the unexplored



2. Broaden horizon:

Users choose topics that have **high similarity** and **high familiarity** compared to the non-selected topics.



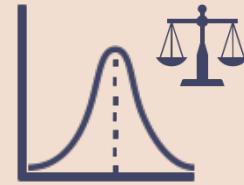
3. Discover the unexplored:

Users choose topics that have **low similarity** and **low familiarity** compared to the non-selected topics.

Results



System
Evaluation



1. Goal Framework:

Broaden Horizon chosen topics are **more similar** and **more familiar** than Discover the unexplored



2. Broaden horizon:

Users choose topics that have **high similarity** and **high familiarity** compared to the non-selected topics.



3. Discover the unexplored:

Users choose topics that have **low similarity** and **low familiarity** compared to the non-selected topics.

Results



System
Evaluation



1. Goal framework

Problem horizon chosen topics are **more similar** and **more familiar** than discover the unexplored



2. Broaden horizons

Users choose topics with **high similarity** and **low familiarity** compared to the non-selected topics.

Discover the unexplored:

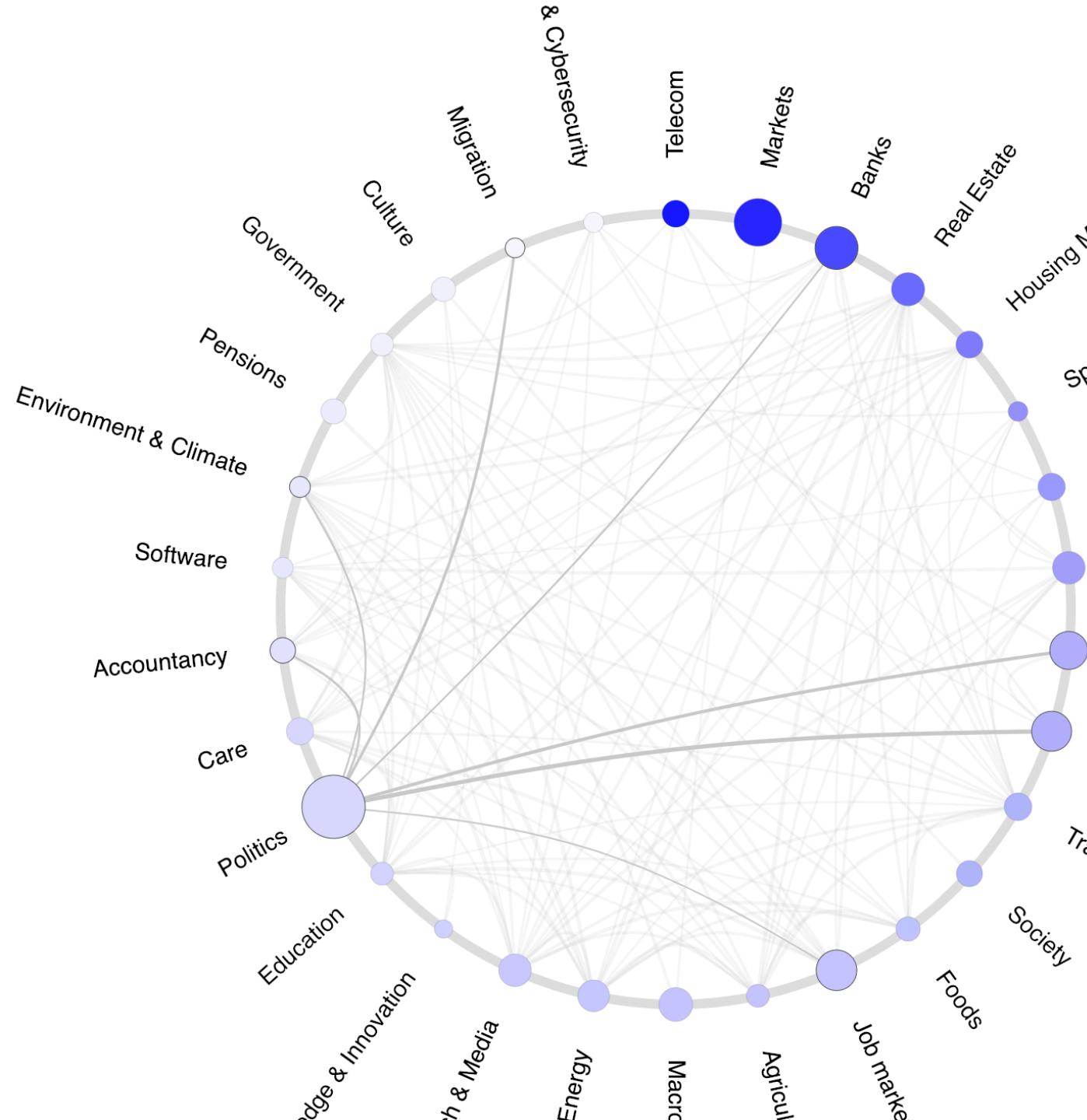
Users choose topics that have **low similarity** and **low familiarity** compared to the non-selected topics.

What have we done

- Novel, scalable and generalizable framework of user-profile explanations
- Exploration of reader data; domain-knowledge and data-driven ontology
- Interface mockup
- User-study to evaluate the value-driven explanations
- “Explainability” as a product

Future

- Further designing implementation of our framework
- Formalize the epistemic goals together with editors
- Extend user studies/ focus groups to all value-driven goals



In summary

- Shown you where/how we personalize @ FD Mediagroep
- Shown you how we evaluate and can determine usefulness of RecSys using IR metrics
- Told you that these metrics are not all there is to it
- Told you it is not so difficult to incorporate diversity, discount for popularity, etc.
 - What we'll work on after the break