

Personalised Blendle

Anne Schuth
data scientist



Personalised Blendle

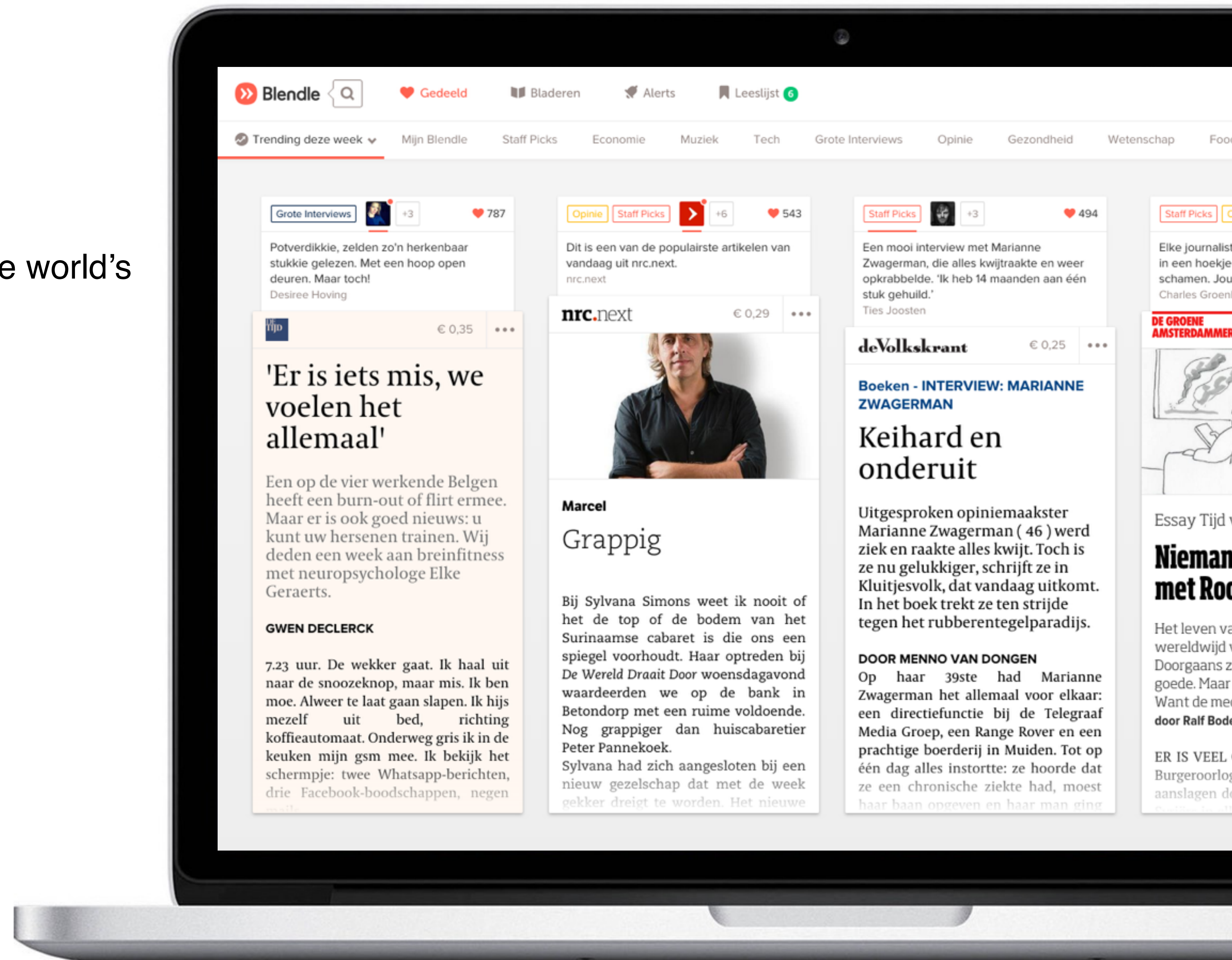
Thanks to Martijn Spitters who created many of these slides

Anne Schuth
data scientist

Blendle

✓ Mission

Help you discover and support the world's best journalism



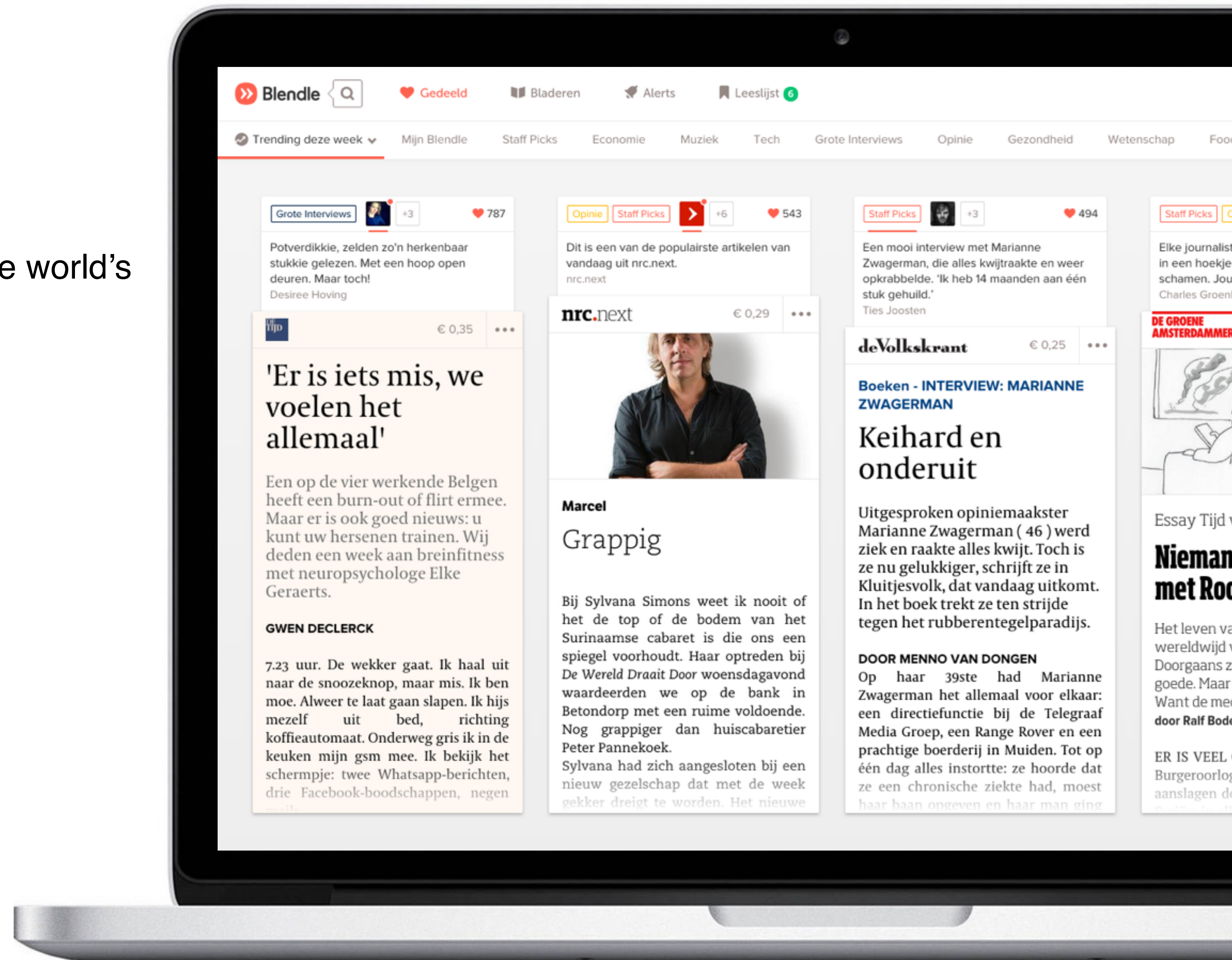
Blendle

✓ Mission

Help you discover and support the world's best journalism

✓ Micropayment platform

articles, issues, subscriptions



Blendle

✓ Mission

Help you discover and support the world's best journalism

✓ Micropayment platform

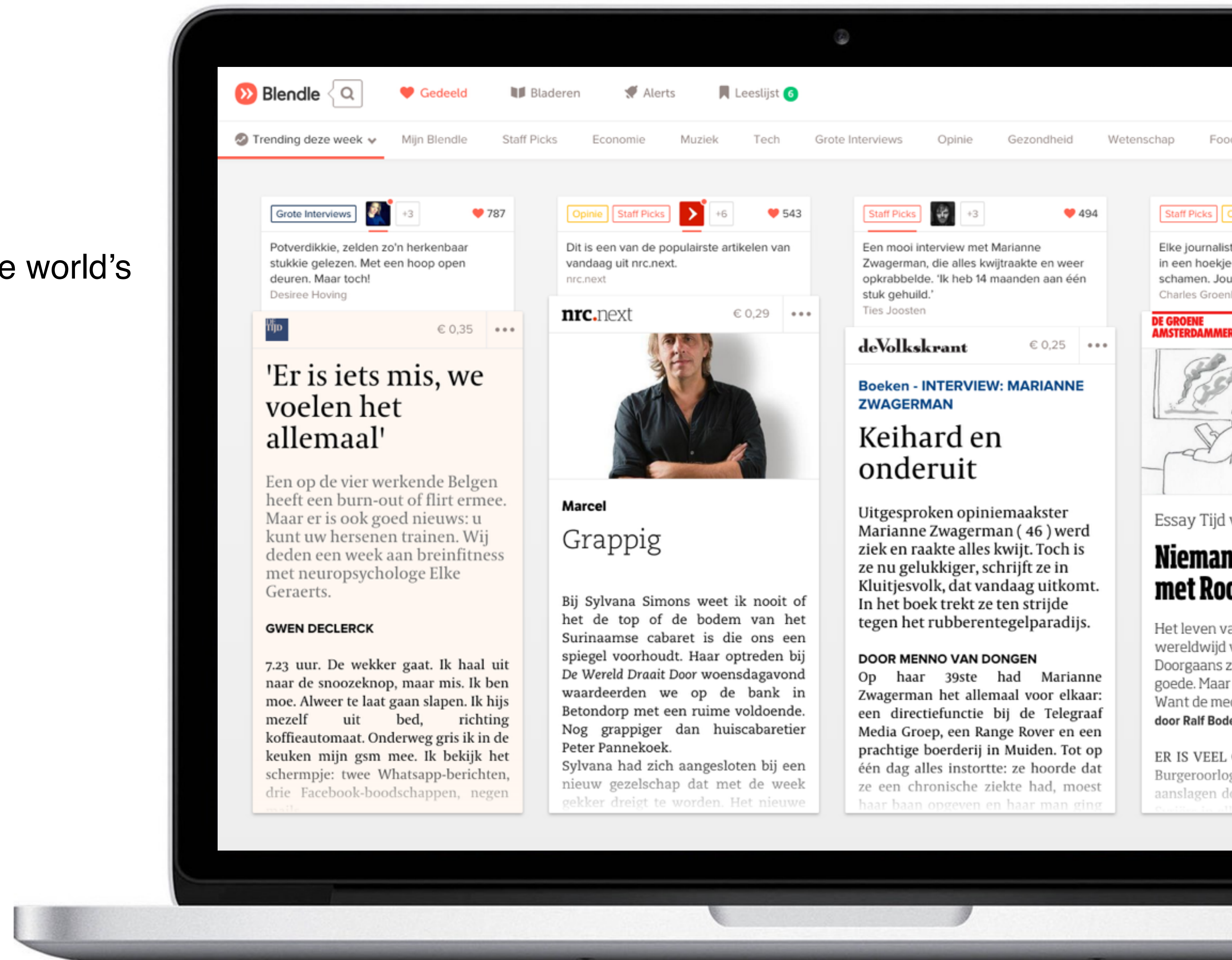
articles, issues, subscriptions

✓ International

May 2014 : The Netherlands

Sept 2015 : Germany

March 2016 : United States



Blendle

✓ Mission

Help you discover and support the world's best journalism

✓ Micropayment platform

articles, issues, subscriptions

✓ International

May 2014 : The Netherlands

Sept 2015 : Germany

March 2016 : United States

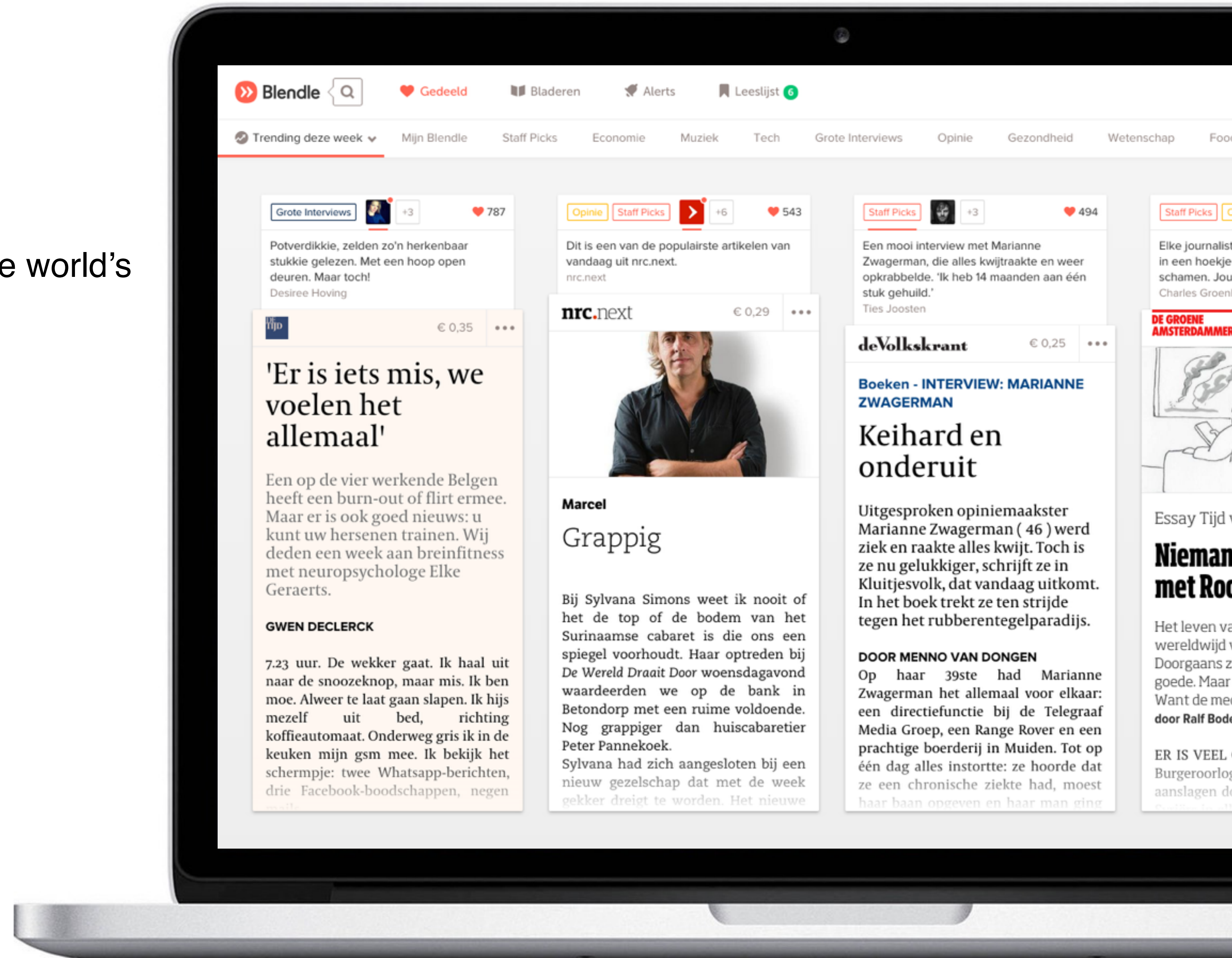
✓ Numbers

650.000 users

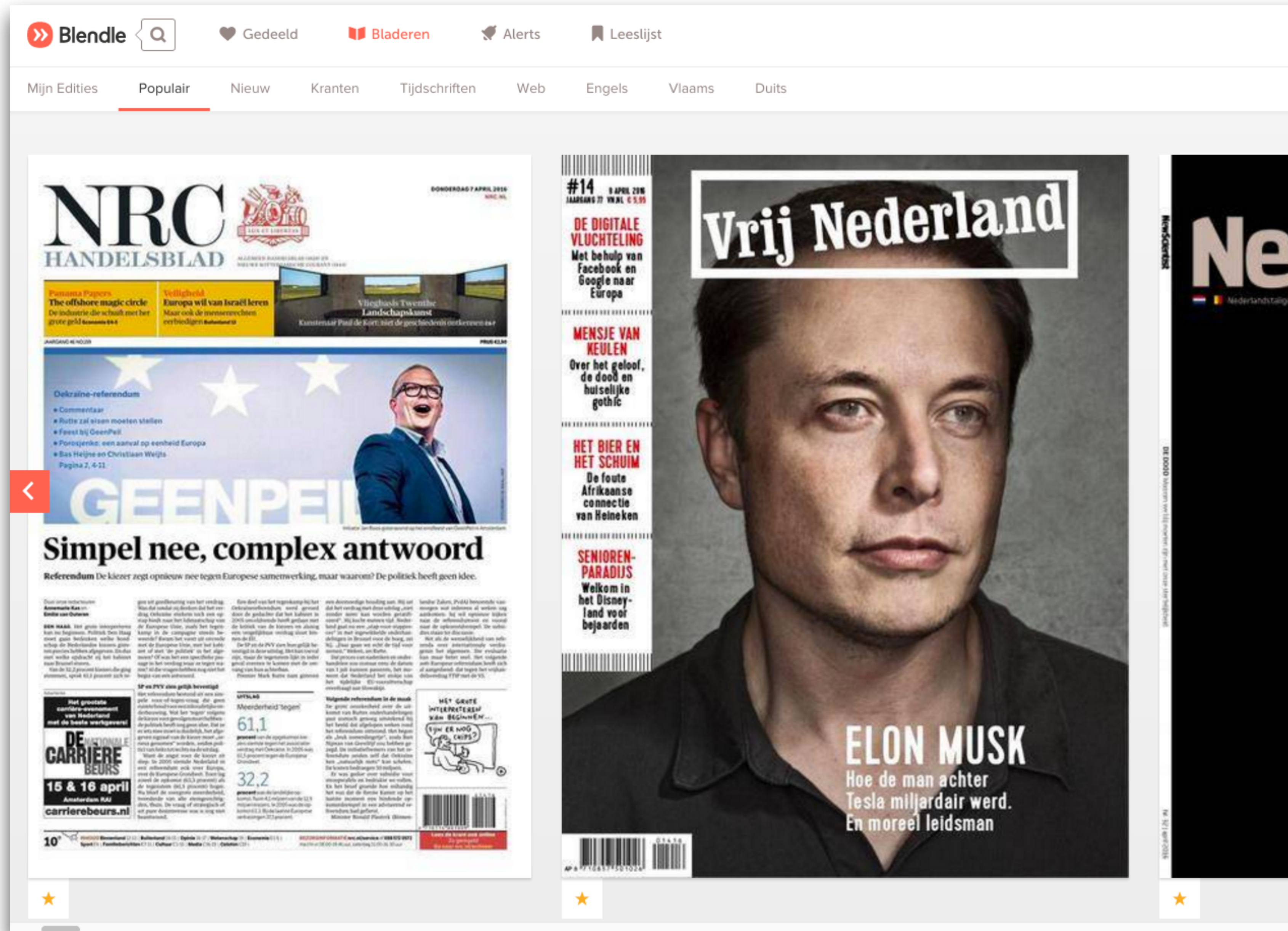
1 in 5 is paying

70 employees

(15 journalists, 50 developers,
+ 30 clippers)



In Blendle you can browse through all quality newspapers and magazines



You only pay for what you read, with a **single click**

Blendle Gedeeld Bladeren Alerts Leeslijst

Populair Vrij Nederland van gisteren

Vrij Nederland € 0,89

ELON MUSK

**PORTRET
DE ONDERNEMER ALS
MOREEL LEIDER**

Ja, er zit geld in moraal. Elon Musk presenteerde de Tesla Model 3, de coole en betaalbare elektrische auto die de wereld moet veranderen, en hij wil de mensheid redden door Mars te koloniseren. Het heeft hem geen windeieren gelegd, dat redden van de mensheid.

SANDER PLEIJ | PORTRET DAN WINTERS

Hoeveel tijd kost een vrouw? Zo luidde de vraag van Elon Musk, directeur van Tesla (iPhone onder de elektrische auto's) en SpaceX (rakettenfabriek die de mensheid naar Mars wil emigreren); een rijke vent die het beste met de mensheid voor heeft.

Hij zat met een probleem, zo vertrouwde hij zijn biograaf toe, hij was net voor de tweede keer gescheiden en zocht nu een nieuwe vriendin, maar hij vreesde dat een

DUURZAMER VERVOER De onthulling van Model 3 van de Tesla. Binnen een dag waren er 100.000 reserveringen

EXCENTRIEKE CELEBRITY Musk met zijn tweede echtgenote, filmmaker

€ 4,49 Koop editie

Hij gebruikt anderen en zichzelf als middel: hing hij rond met patsers, met andere inter- nairs, met filmsterren, hij ging wonen in Bel-



onderdelen zelf te maken. Niks aaneenklagen: liever heel grote fabrieken met alles en iedereen vlak bij elkaar want dat betekent: niet wachten, geen tijdverlies, directe communicatie. Binnenkort opent hij de Gigafactory, een nieuwe batterijfabriek met de grootste omvang ter wereld en nul giftige uitstoot. Alles lijkt onnos vertoonend: geen reclame maken - met de ideologische reden dat reclame manipuleert en overmatig is. Auto's moeten niet jaarlijks met een nieuw model komen, nee, het huidige model wordt constant beter - ah je wilt wachten op een nieuw model kun je voor altijd wachten. Bovendien ondergaat je Tesla continu software-updates. Iets als automatisch van rijbaan veranderen is een nachtelijke download. Over vijf jaar, denkt Musk, zijn de zover dat de Tesla's zichzelf kunnen rijden.

Ook onconventioneel: raketten maken en lanceren vanaf het afgelegen atol Kwajalein in de Grote Oceaan. Alles kan hij ook en onde moekt hij zelf. De technische handboeken over ruimtevaart leest hij zelf. Dat de werkdruk gigantisch is, moge duidelijk zijn, maar hij blijft vrienden en vrienden w respect. Als je tachtig u baas werkt, blijf je dat een er zelf hondt werkt. M ze doen het niet voor de mensheid. Het doel is zo zeuren over vijf jaar alheid op spel staat? Het futiel, heiligt alle middelen Musk stelt simpel: 'Je ge re auto's of raketten cre week, het koloniseren v beuren in voortgang p Opofferen, allemaal.

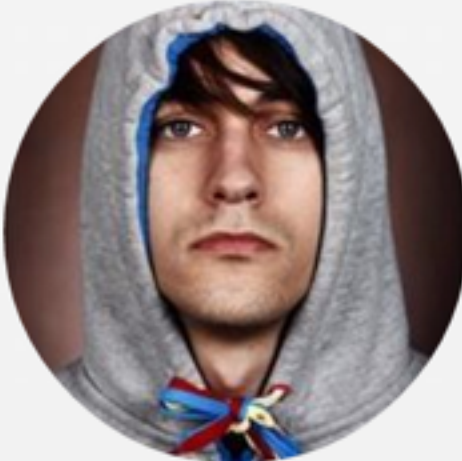
Musk schijnt zichzelf e soort computers te bes over leren als het 'down algoritmen in je brein' kinderen met een 'downloadcode' heeft. Me noemt hij een bodap:

Articles are shown in their **original layout and style**, and **without advertisements**



Blendle is a **social network**, where users can **follow friends**, to see today's **must-reads**

blendle   Gedeeld  Bladeren  Alerts



Alexander Klöpping
Technoviking.

✓ Volgend 58622 Volgers 1272 Volgend

Alexander Klöpping
3 dagen geleden

Mooi stuk over de ongelofelijke groei (naar verluid zetten ze vorig jaar 3 miljard om) van het Amsterdamse Booking.com

 timbesselink en 7 anderen

ELSEVIER Juist 



LIZANNE SCHIPPEN

Booking.com slaagt erin om één ding heel goed te doen

Terwijl reisbureaus wegwijnen, publiceert Booking.com duizelingwekkende groeicijfers. De grootste boekingssite ter wereld is begonnen door een Twentse technaut. 'Ze dachten dat hij gek was'

Alexander Klöpping
4 dagen geleden

Sponsormanagers denken te vaak "wat vind ik zelf leuk", en dan komen ze bij sport uit. Maar cultuurstellingen beginnen binnen te lopen.

 Marieke Monden en 1 ander

Het PAROOL € 0,25

Veel waar voor je sponsorgeld

LORIANNE VAN GELDER

Cultuur sponsoren is volgens onderzoek veel beter voor het imago dan het steunen van een sportclub. Gesubsidieerde culturele instellingen staan er meer voor open dan vroeger. Maar doen ze niet te veel concessies?

A udi, 'de grijpgrage kapitalist', wilde een autoshowroom in een museum. Een schande. 1999 was het en de gemeenteraad stond op zijn kop. Een geheime sponsorovereenkomst tussen het Stedelijk Museum en autofabrikant Audi was uitgelekt en de cultuurwereld was te klein. De toenmalige museumdirecteur Rudi Fuchs was met

Alexander Klöpping
4 dagen geleden

LOL. "Wie zijn deze erupties?"

 Koen Rutjens en 1 ander

Het PAROOL



Aldus

Dansen op vulkaan Bárðarbunga

Reisbureau Askja organiseert een eruptiereis als vulkaan Bárðarbunga op IJsland. Lieve Wolfs vertelt.

ROB ROMBOUTS

U hoopt op een uitbarsting
"Voor ons is het een gelukkig ongeluk als de Bárðarbunga uitbarsting komt."

Blendle is a **social network**, where users can **follow friends**, to see today's **must-reads**

Currently the most important means to personalize!

The screenshot displays the Blendle social network interface. At the top, there is a navigation bar with the Blendle logo, a search icon, and links for 'Gedeeld', 'Bladeren', and 'Alerts'. The main content area features a user profile for Alexander Klöpping, identified as 'Technoviking'. His profile includes a circular profile picture, his name, and statistics: 58,622 followers and 1,272 people he is following. Below the profile, there is a feed of articles. The first article is from 'Juist' by Lizanne Schippen, titled 'Booking.com slaagt erin om één ding heel goed te doen'. The second article is from 'Het PAROOL' by Lorianne van Gelder, titled 'Veel waar voor je sponsorgeld'. The third article is also from 'Het PAROOL', titled 'Dansen op vulkaan Bárðarbunga'. The interface is clean and modern, with a focus on user profiles and a curated feed of articles.

Blendle button for payments per article and subscriptions

NieuwsCultuur & LevenMediaBEWAAR



Waarom een eigen krantje? Omdat het kan

De bron

ARTIKEL Wie of wat zette je leven op het juiste spoor? In een serie interviews vraagt Sacha Bronwasser mensen naar hun inspiratiebron. Zo zag satiricus Wim de Bie hoe het uitgeven van een gestencild blad de levensvervulling werd van zijn vader.

Door: Sacha Bronwasser 7 april 2016, 02:00



'Bloggen doe ik niet meer, maar tegenwoordig beschouw ik Twitter als mijn 'krantje'. Dat is natuurlijk beperkt, met die 140 tekens, maar ik hang er hele artikelen en filmpjes aan - heerlijk. Weer een technisch middel dat vrijheid verschaft, daarom ben ik er zo enthousiast over. De vrijheid om je uit te drukken, die is altijd essentieel.

'Ik heb wat krantjes meegenomen uit mijn archief.'

 86% vindt dit artikel waardevol

 Slechts € 0,25

 Niet goed? Geld terug!

Doorlezen met  Blendle

Ingelogd als Martijn ▾

AANBEVOLEN ARTIKELN

'Kees is leuker dan ik, dat heb ik nooit een bezwaar gevonden'
 5 maart 2015

'Alles is hetzelfde, behalve dat de personages wat dikker zijn geworden'
 19 mei 2015

Mike Boddé verkiest muziek boven de grap
 22 maart 2016

ADVERTENTIE



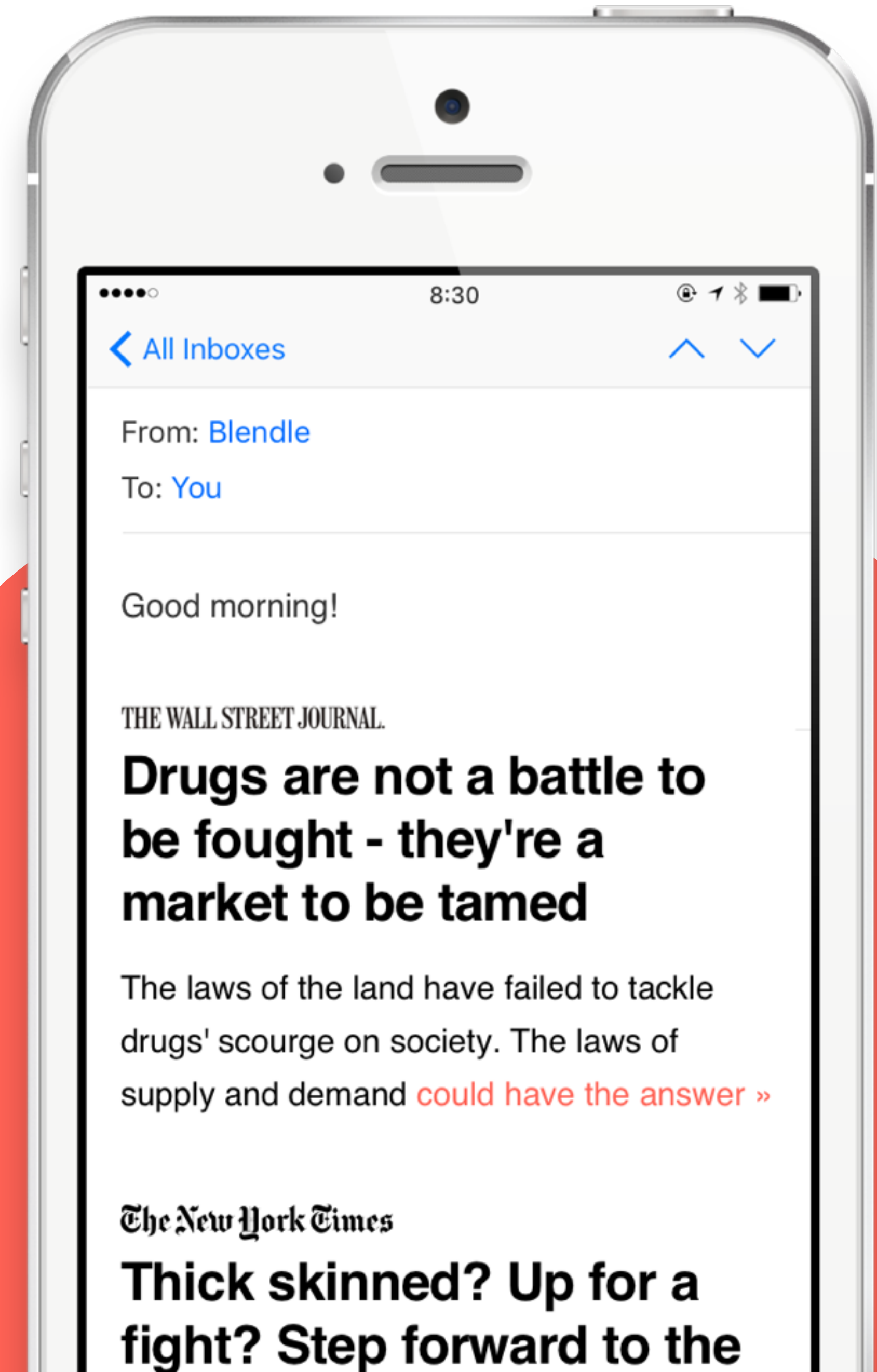
Mobielvrij rijden bevordert de verkeersveiligheid.

→ LEES MEER

MEEST GELEZEN MEDIA

1. Het rijk moet meer doen

Our **editors** select the best articles for our email newsletter **every day**



Data at Blendle

- **Articles**
 - **3M in total**
 - **5K new every day**

Data at Blendle

- **Articles**
 - 3M in total
 - 5K new every day
- **Users**
 - 650K users
 - 1 in 5 converts to paying user

Data at Blendle

- **Articles**
 - 3M in total
 - 5K new every day
- **Users**
 - 650K users
 - 1 in 5 converts to paying user
- **Events**
 - 750M in total
 - 2M new every day

Data at Blendle

- **Articles**

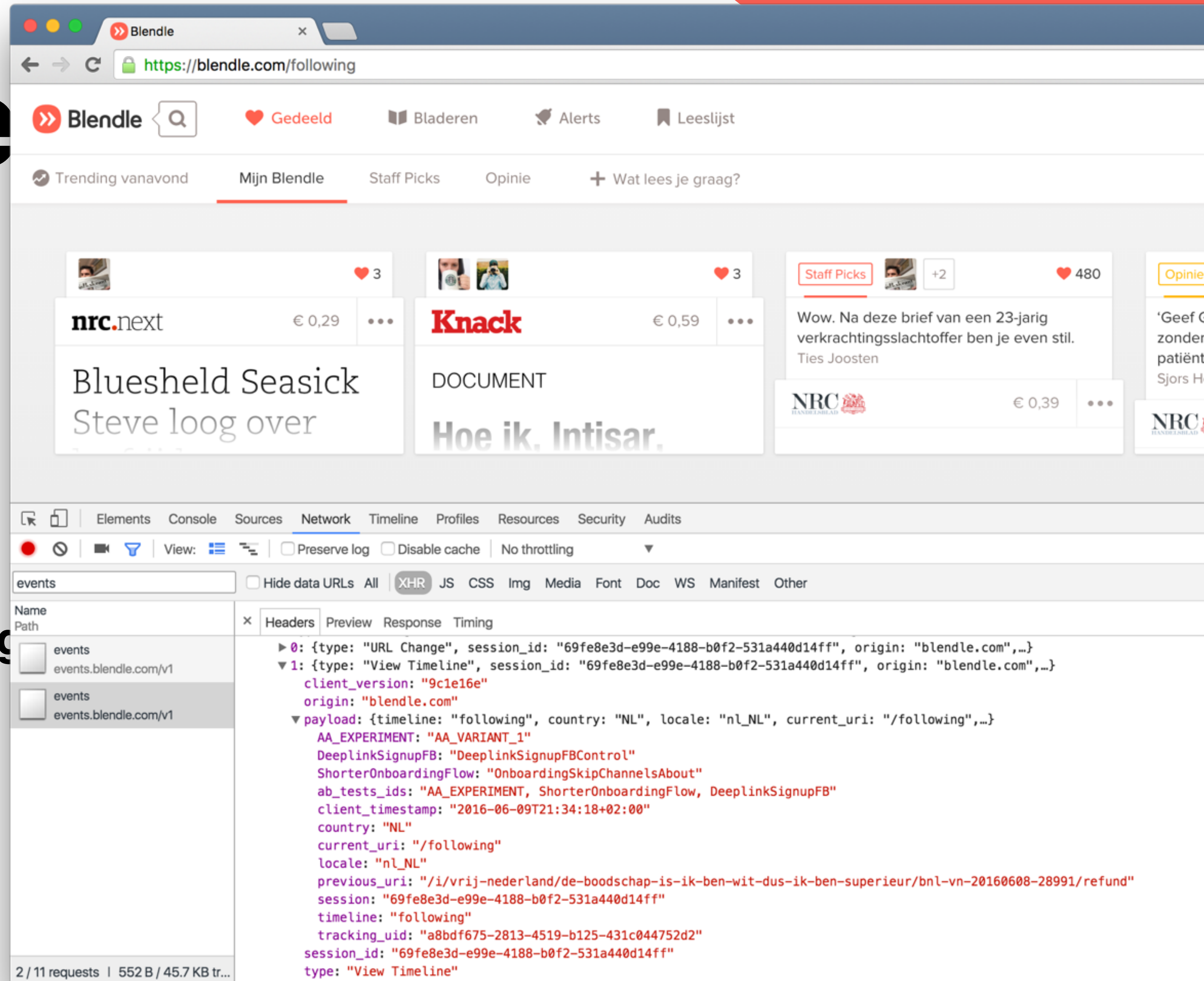
- 3M in total
- 5K new every day

- **Users**

- 650K users
- 1 in 5 converts to paying

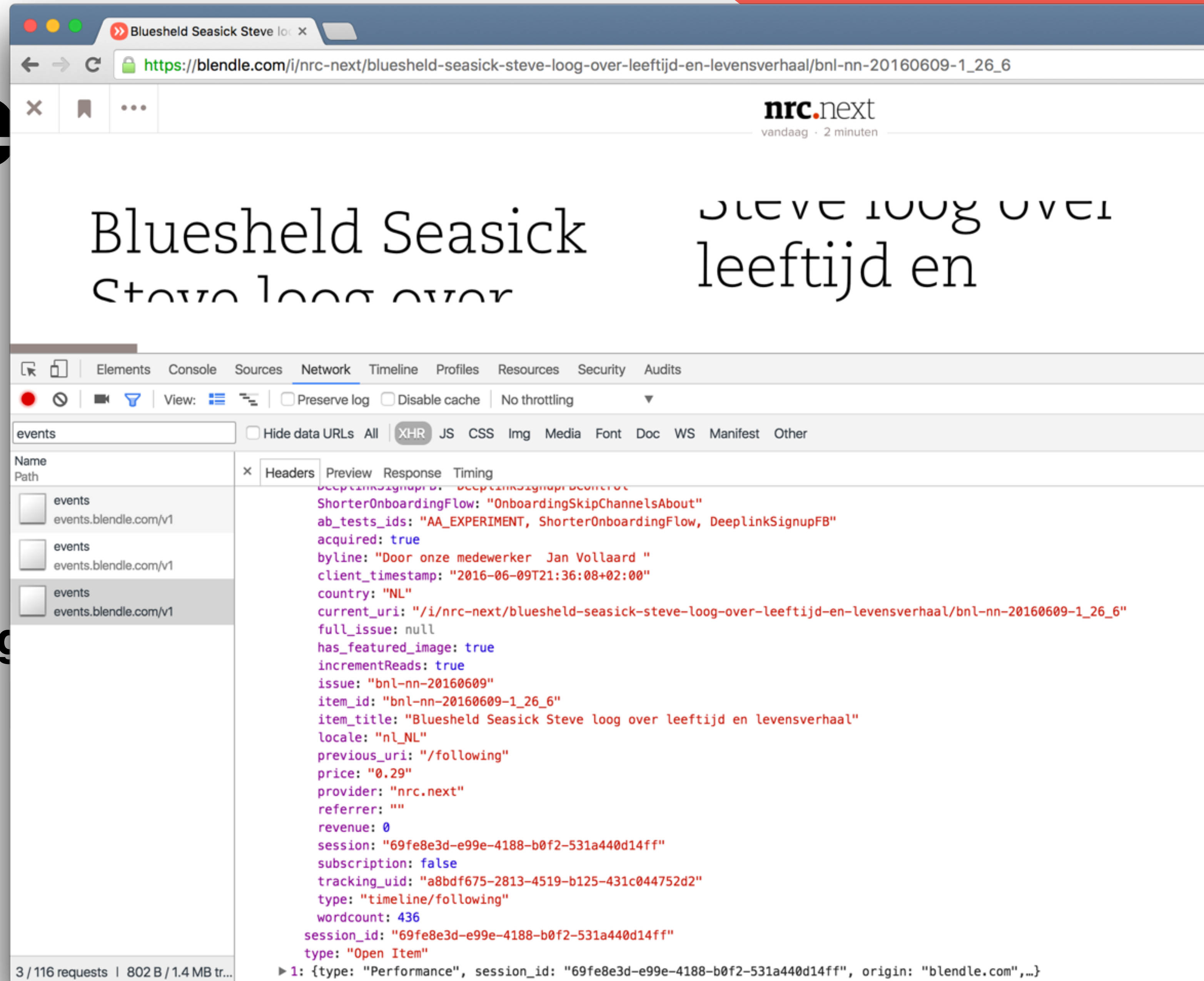
- **Events**

- 750M in total
- 2M new every day



Data at Blendle

- **Articles**
 - 3M in total
 - 5K new every day
- **Users**
 - 650K users
 - 1 in 5 converts to paying
- **Events**
 - 750M in total
 - 2M new every day

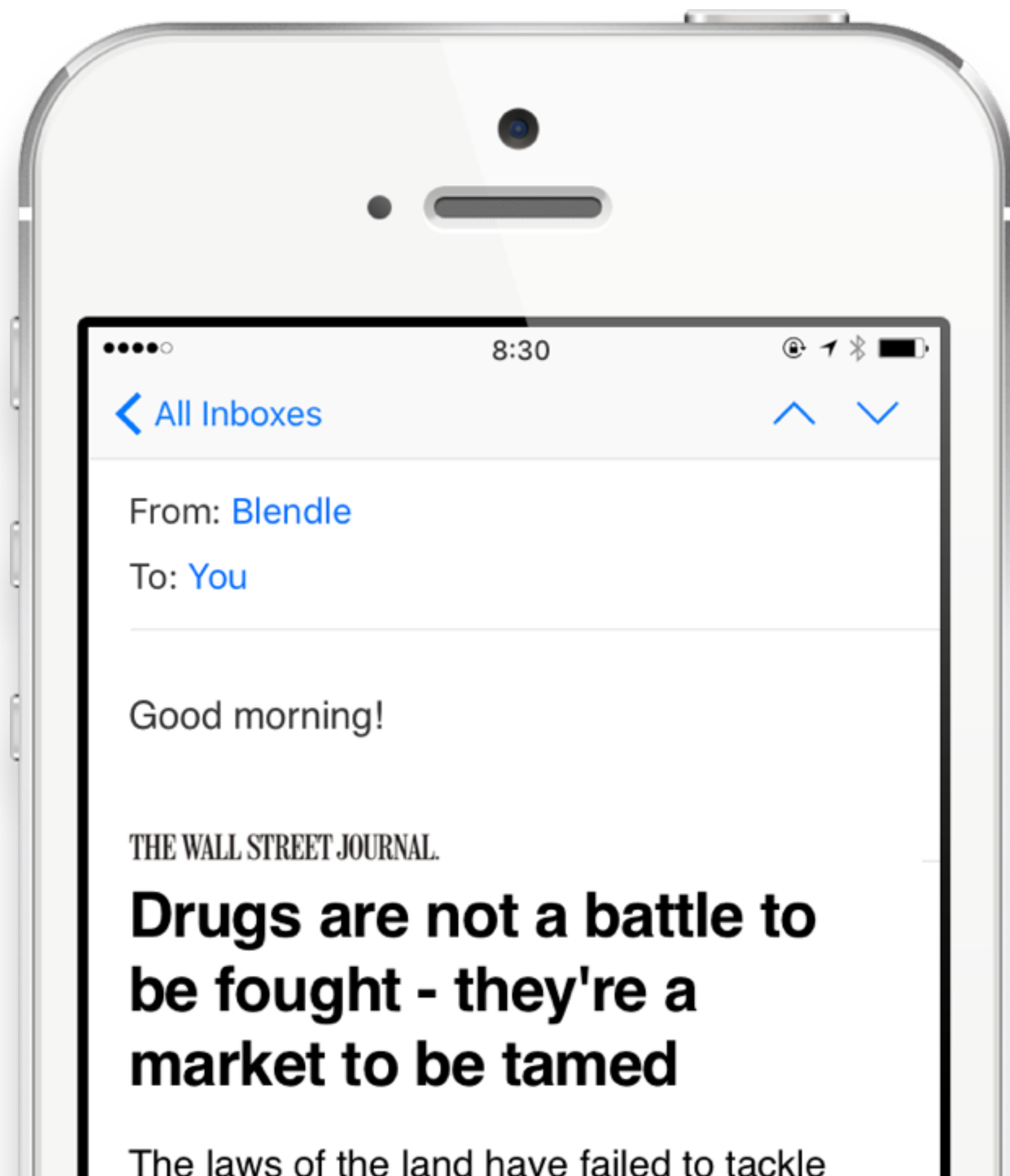


Personalised recommendations

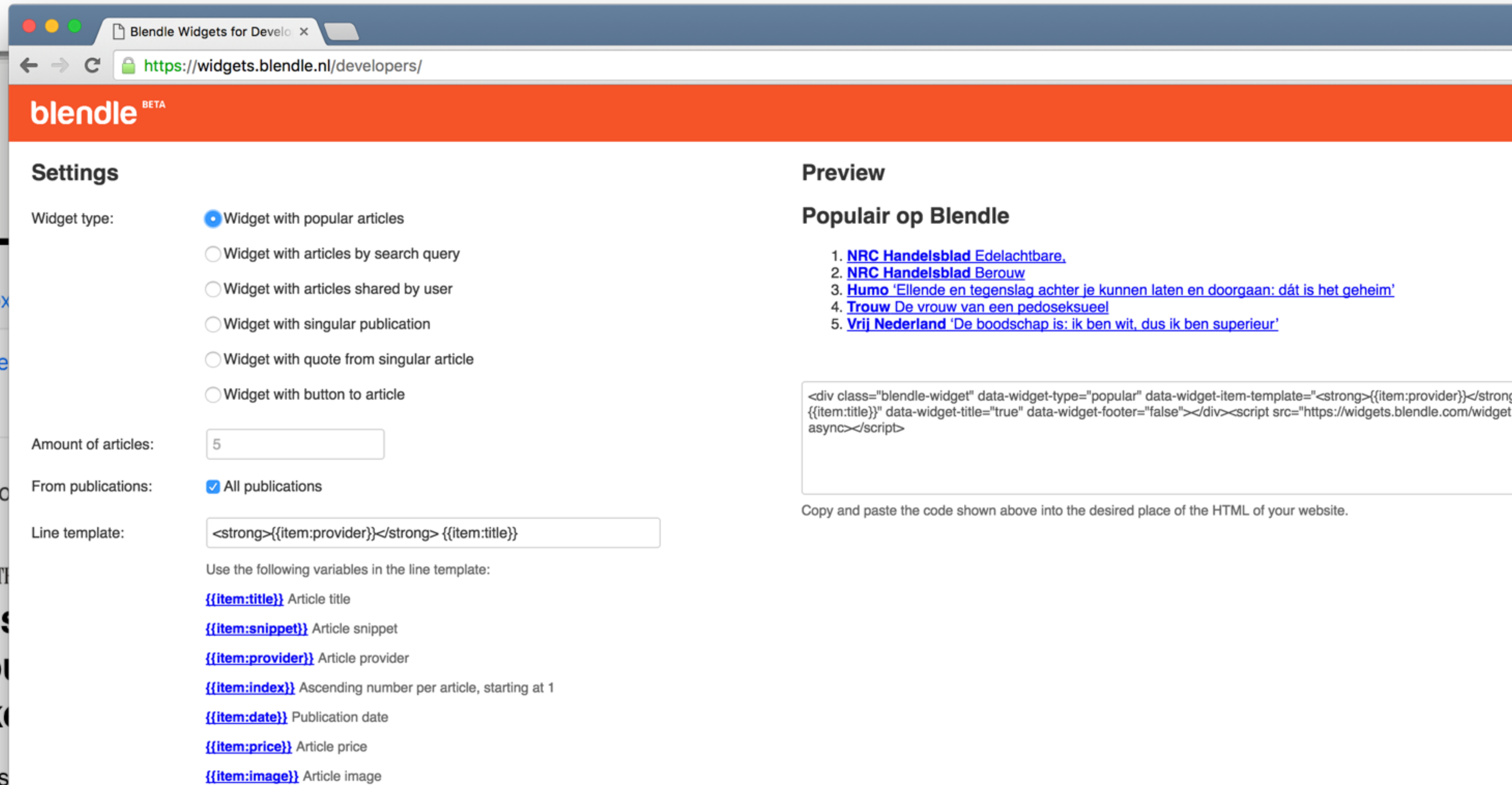
The right articles to the right users



Rankings at Blendle



Rankings at Blendle



Blendle Widgets for Developers

https://widgets.blendle.nl/developers/

blendle BETA

Settings

Widget type:

- ☒ Widget with popular articles
- ☐ Widget with articles by search query
- ☐ Widget with articles shared by user
- ☐ Widget with singular publication
- ☐ Widget with quote from singular article
- ☐ Widget with button to article

Amount of articles:

From publications: ☒ All publications

Line template:

Use the following variables in the line template:

- [{{item:title}}](#) Article title
- [{{item:snippet}}](#) Article snippet
- [{{item:provider}}](#) Article provider
- [{{item:index}}](#) Ascending number per article, starting at 1
- [{{item:date}}](#) Publication date
- [{{item:price}}](#) Article price
- [{{item:image}}](#) Article image

Preview

Populair op Blendle

- [NRC Handelsblad Edelachtbare,](#)
- [NRC Handelsblad Berouw](#)
- [Humo 'Ellende en tegenslag achter je kunnen laten en doorgaan: dát is het geheim'](#)
- [Trouw De vrouw van een pedoseksueel](#)
- [Vrij Nederland 'De boodschap is: ik ben wit, dus ik ben superieur'](#)

```
<div class="blendle-widget" data-widget-type="popular" data-widget-item-template="<strong>{{item:provider}}</strong> {{item:title}}</div><script src="https://widgets.blendle.com/widget-async"></script>
```

Copy and paste the code shown above into the desired place of the HTML of your website.

Rankings at Blendle

blendle BETA

Settings

Widget type:

Amount of articles:

From publications:

Line template:

< All Inbox

From: Blendle

To: You

Good morning

THE WALL STREET JOURNAL

Drugs

be for

mark

The laws

f

p

YouTube

t

ig

Grazia - Home

www.grazia.nl

BESTEL JE GRAZIA

BLENDLE

ADVERTORIAL

SAY WHAT? DE VAJAYJAY-MASSAGE VAN 350 EURO

GRATIS MY JEWELLERY ARMBAND VOOR JOU

TOYBOY JORDAN LAAT LARA STONE WEER SHINEN

KHLOÉ KARDASHIAN IS KLAR MET HAAR OUDE LEVEN

Amber Heard

lifestyle

Happy TV: Kat en baby hebben hevige ruzie om een dekentje

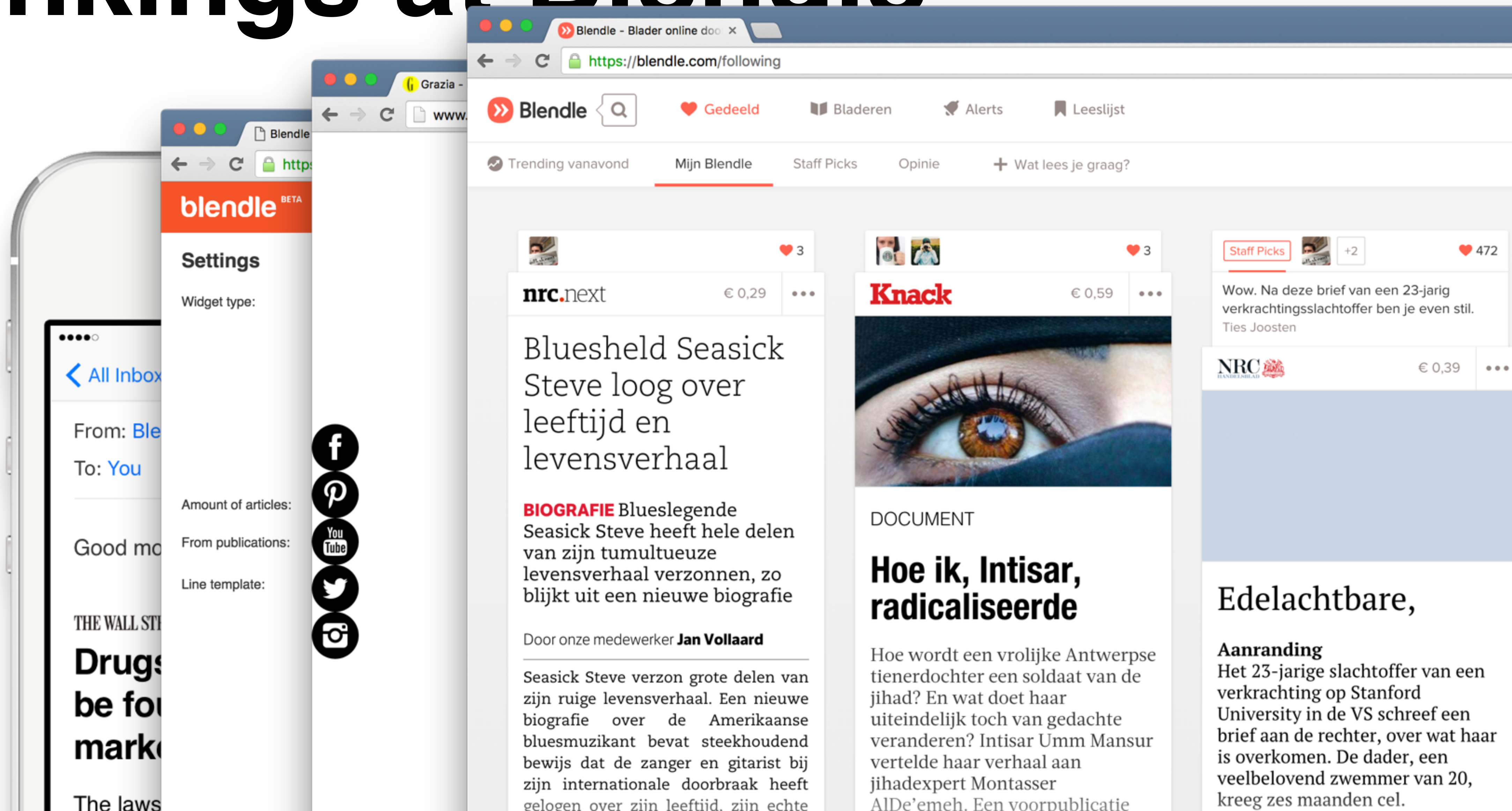
showbizz

Raffaëla is trots: 'Ik ben nu vier jaar clean'

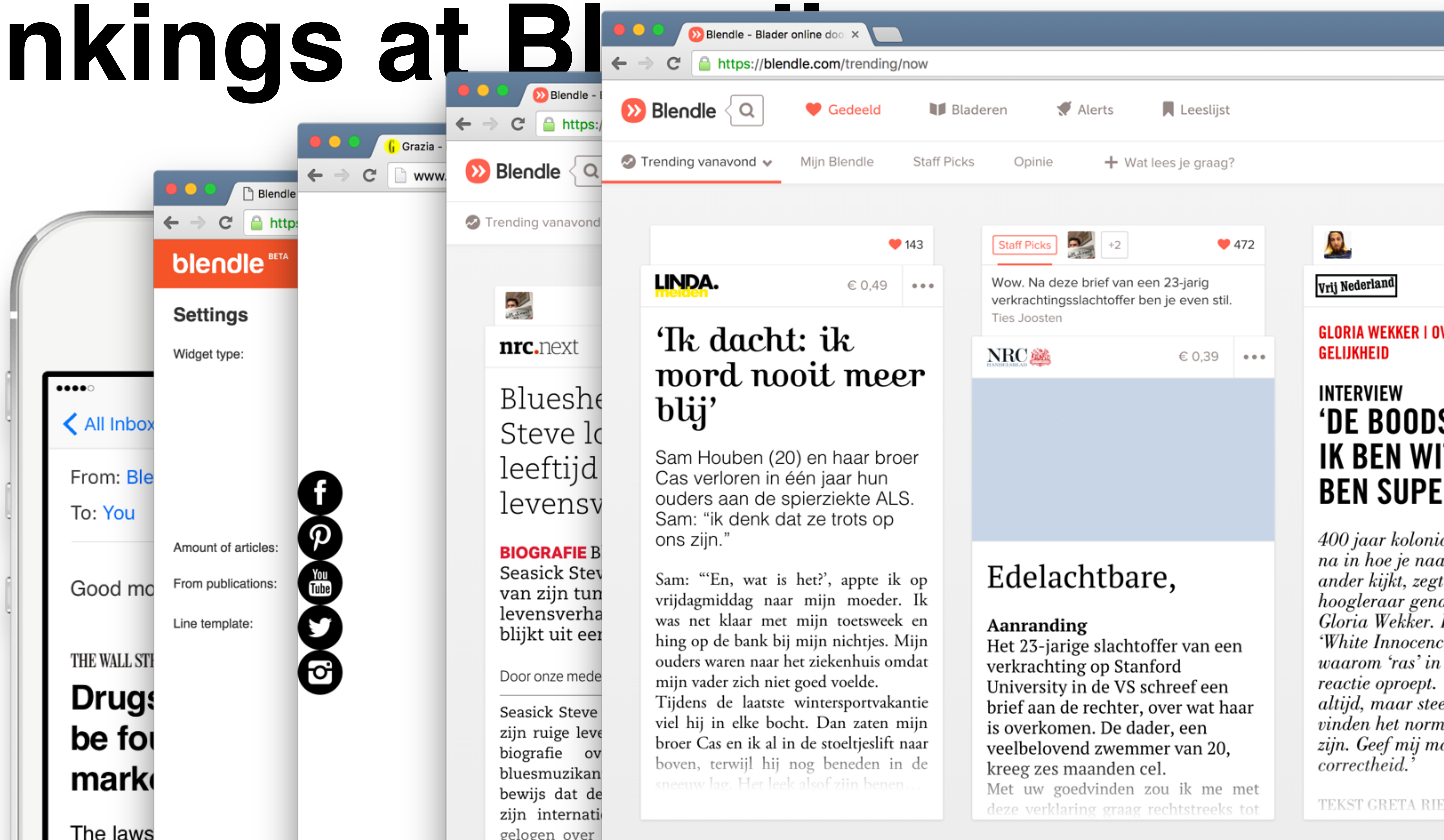
showbizz

Zien: Josje deelt bikini-kiekje

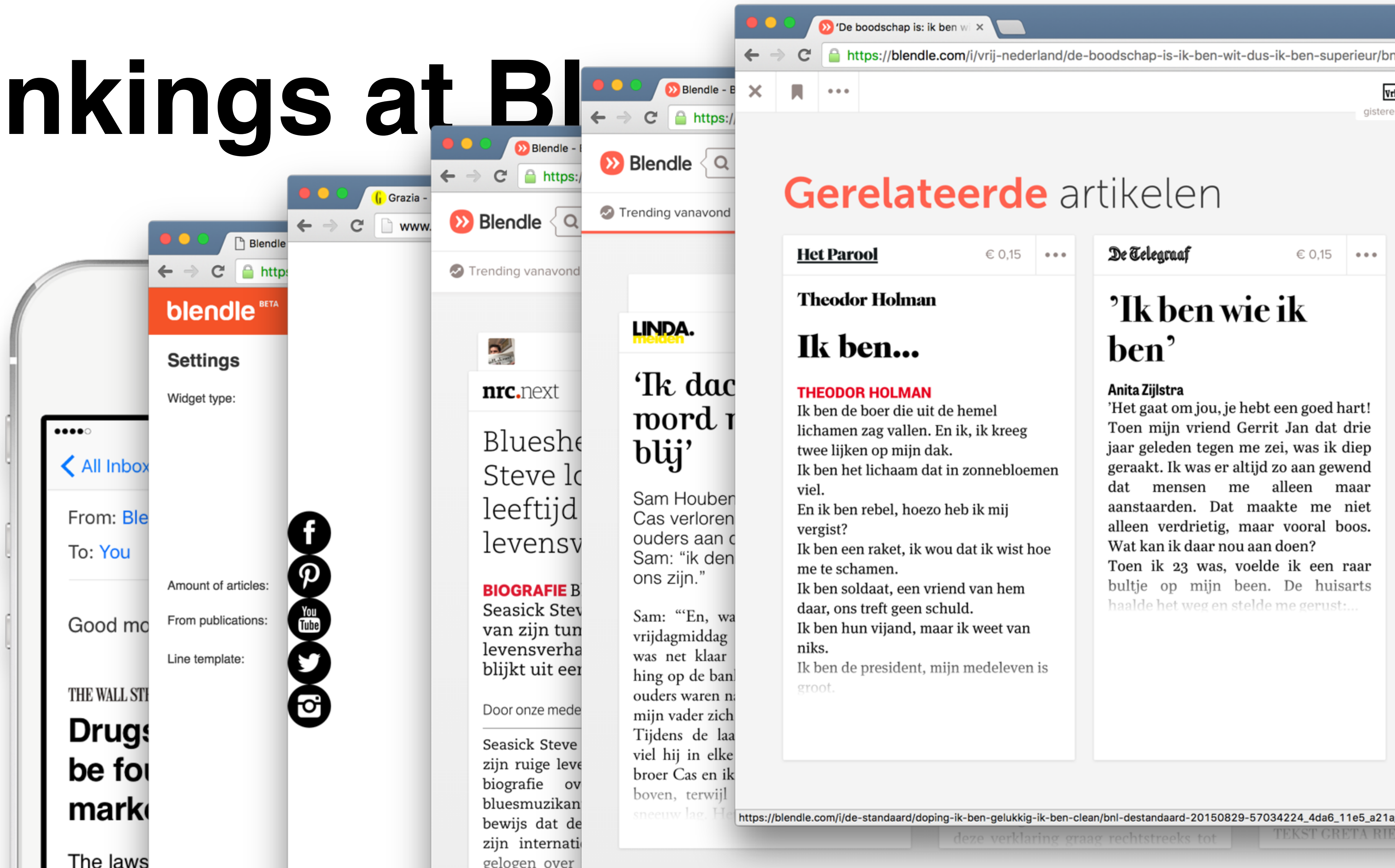
Rankings at Blendle



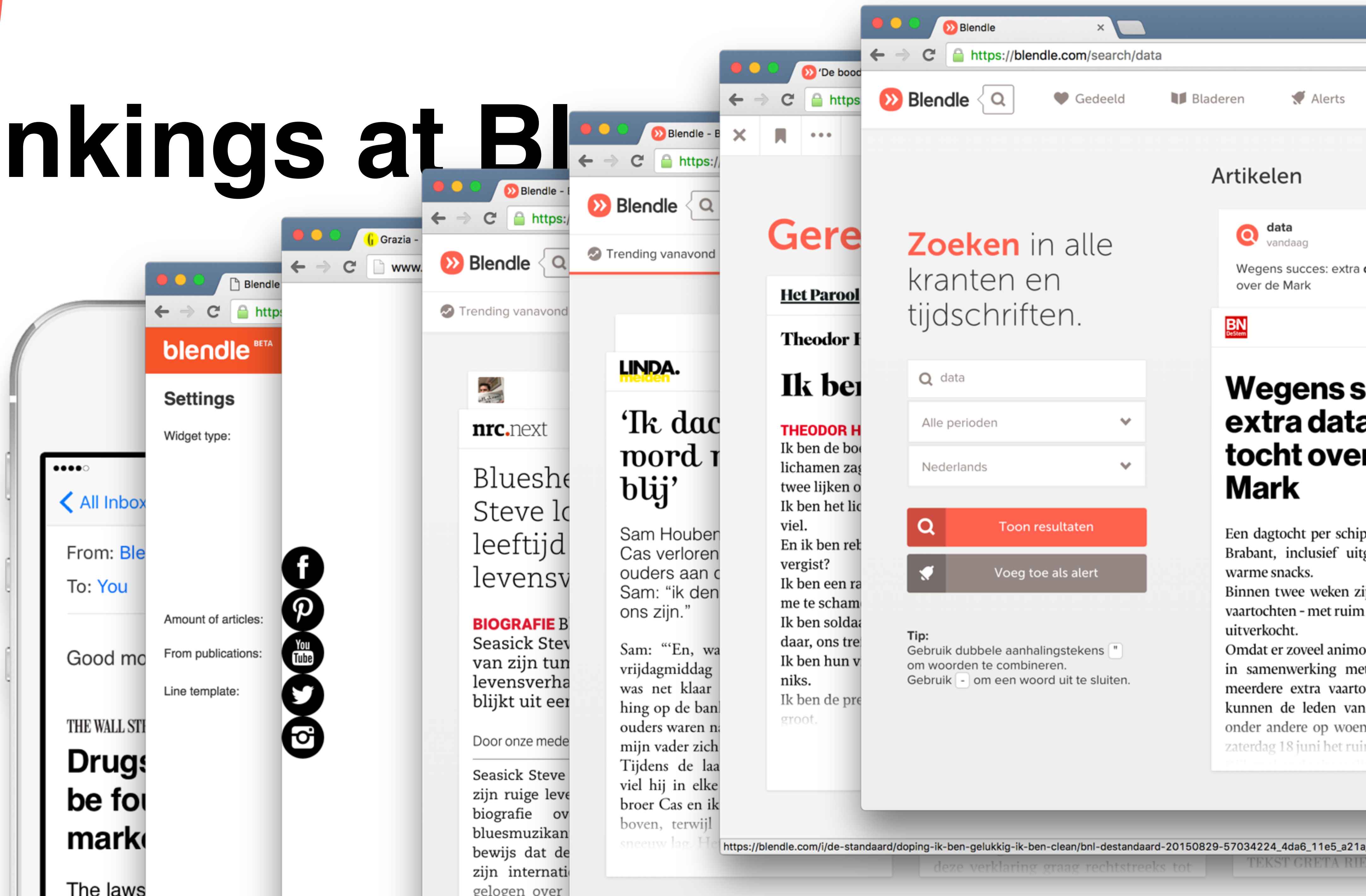
Rankings at Blendle



Rankings at Blendle



Rankings at Blendle



Rankings at Blendle

Rankings at Blendle

- Should be personalised

Rankings at Blendle

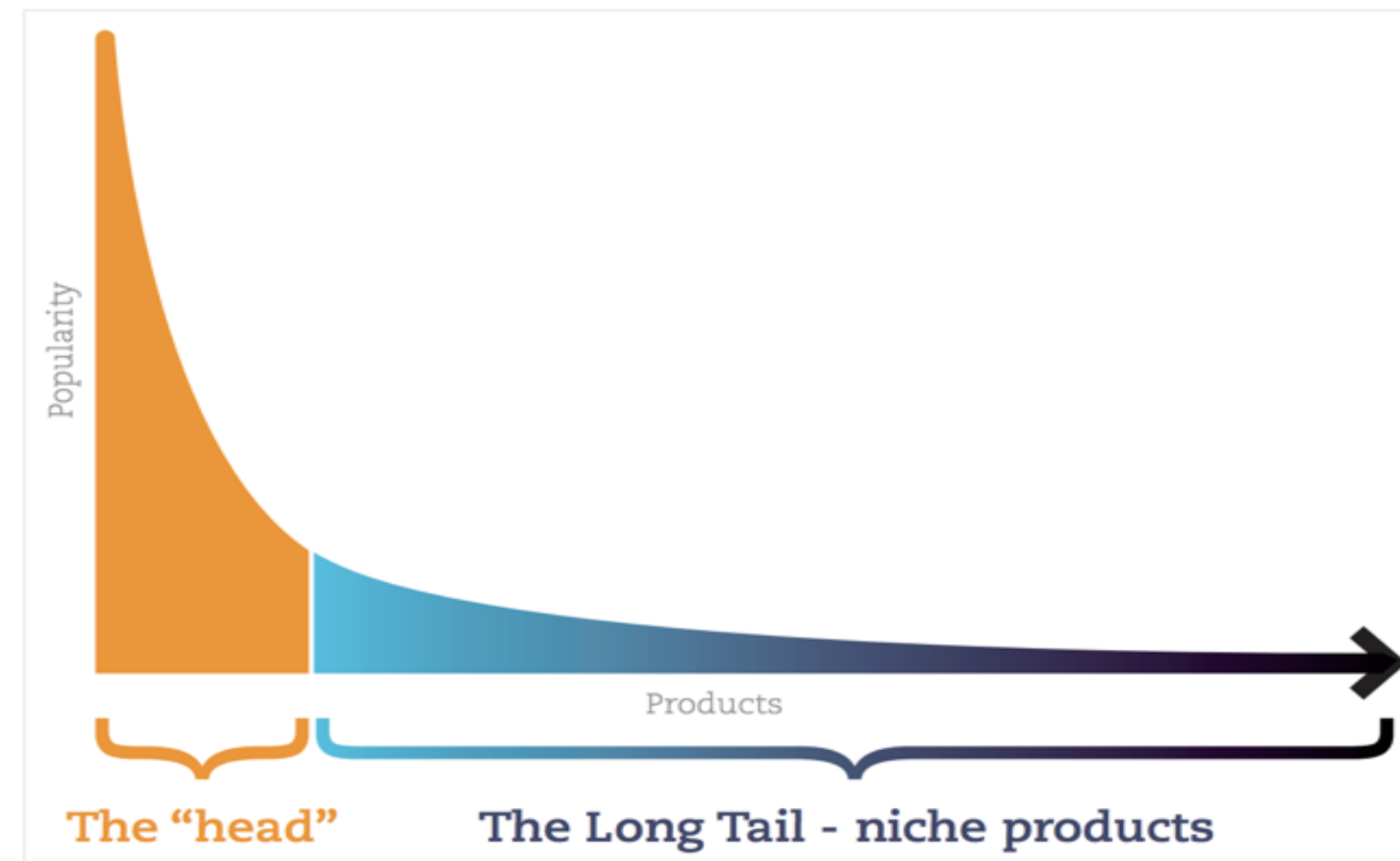
- Should be personalised
 - Lots of content

Rankings at Blendle

- Should be personalised
 - Lots of content
 - Lots of different tastes

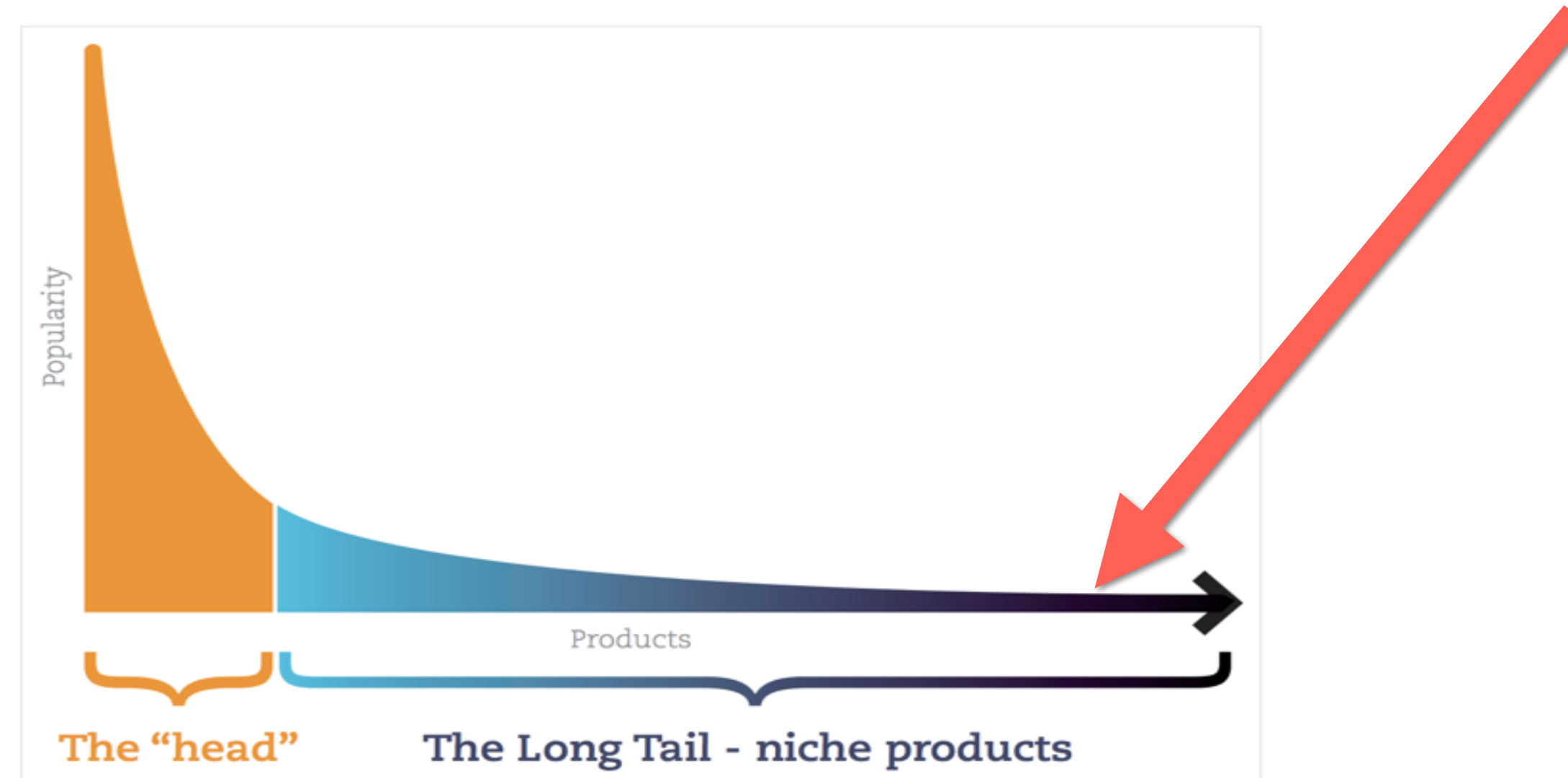
Rankings at Blendle

- Should be personalised
 - Lots of content
 - Lots of different tastes



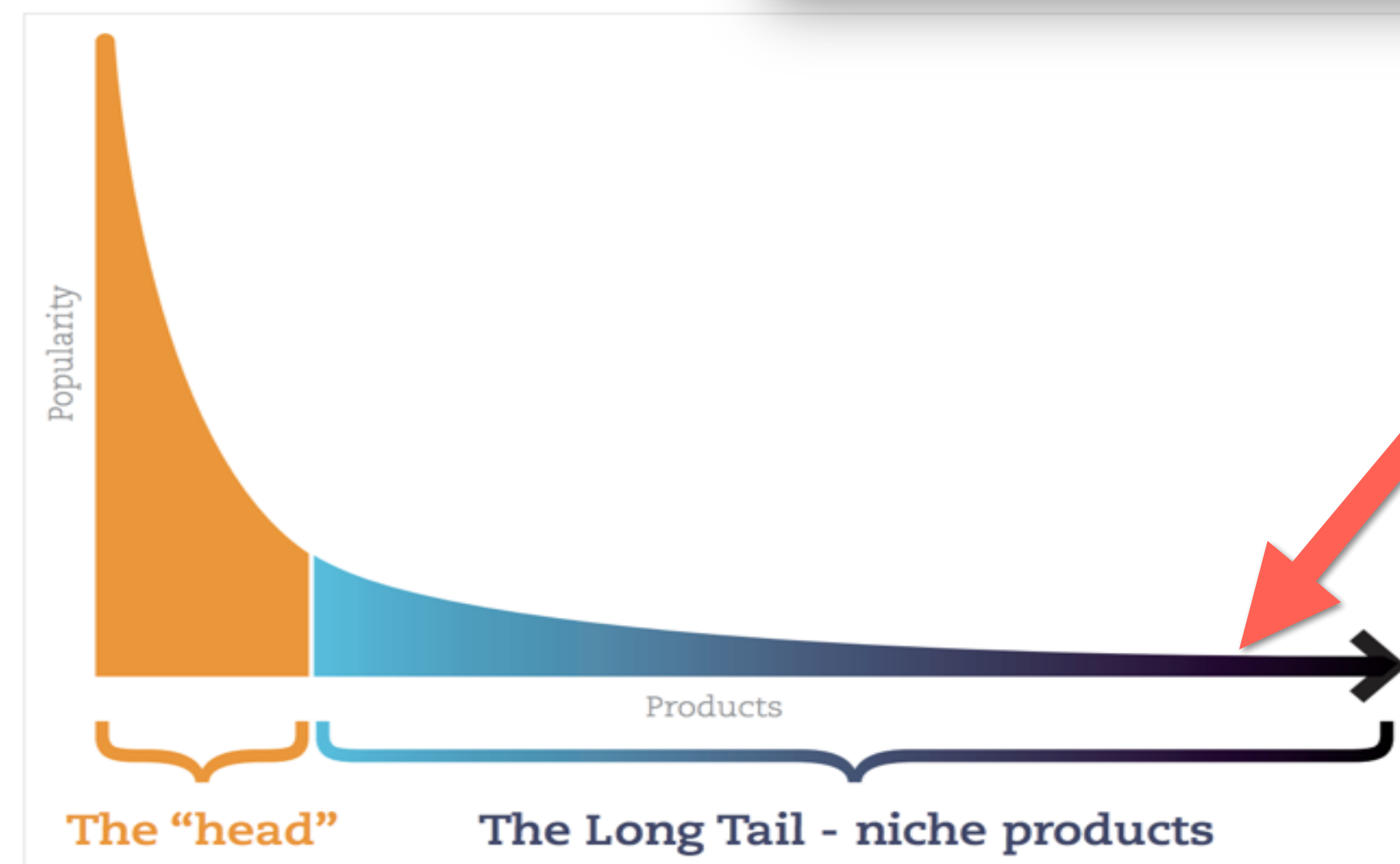
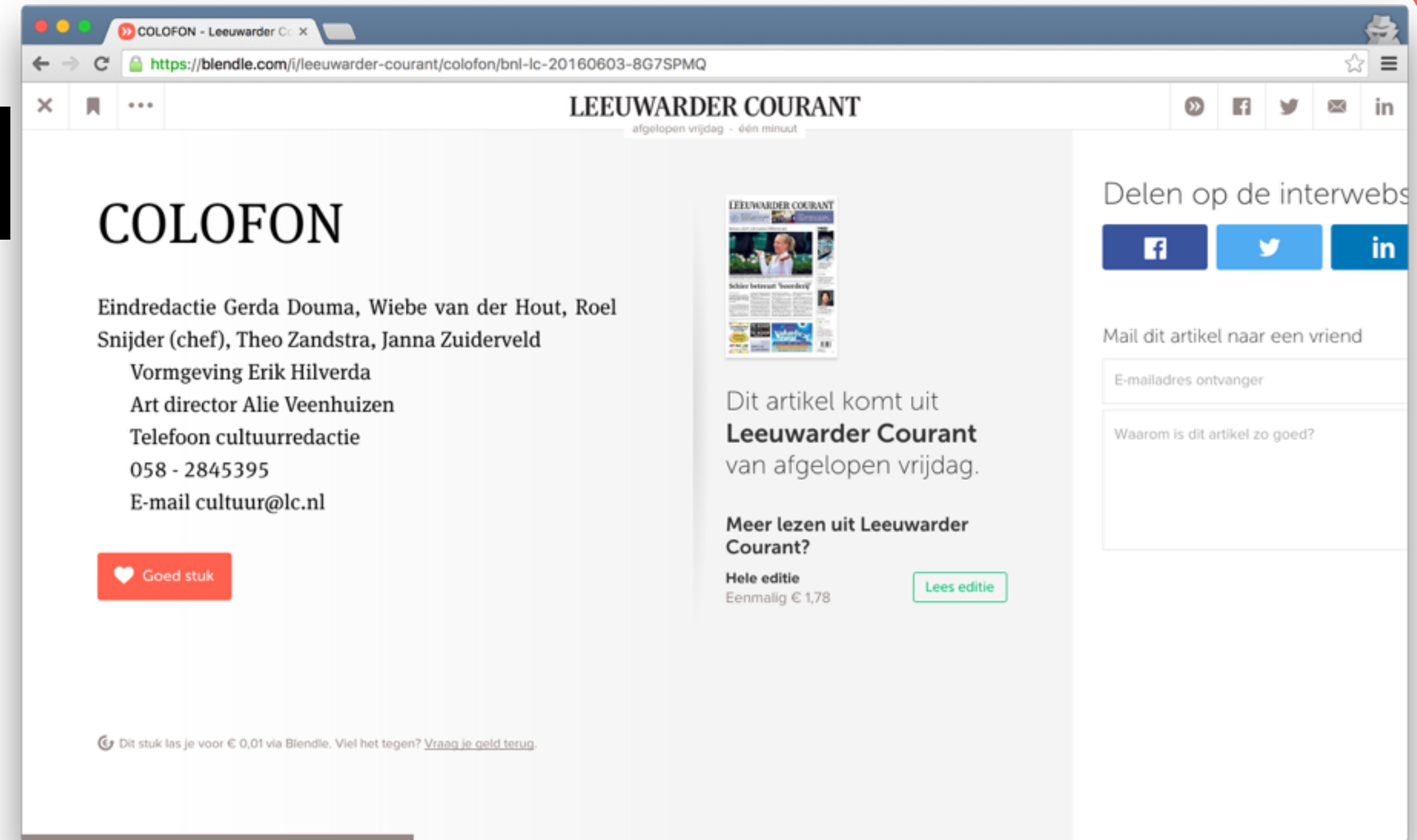
Rankings at Blendle

- Should be personalised
 - Lots of content
 - Lots of different tastes



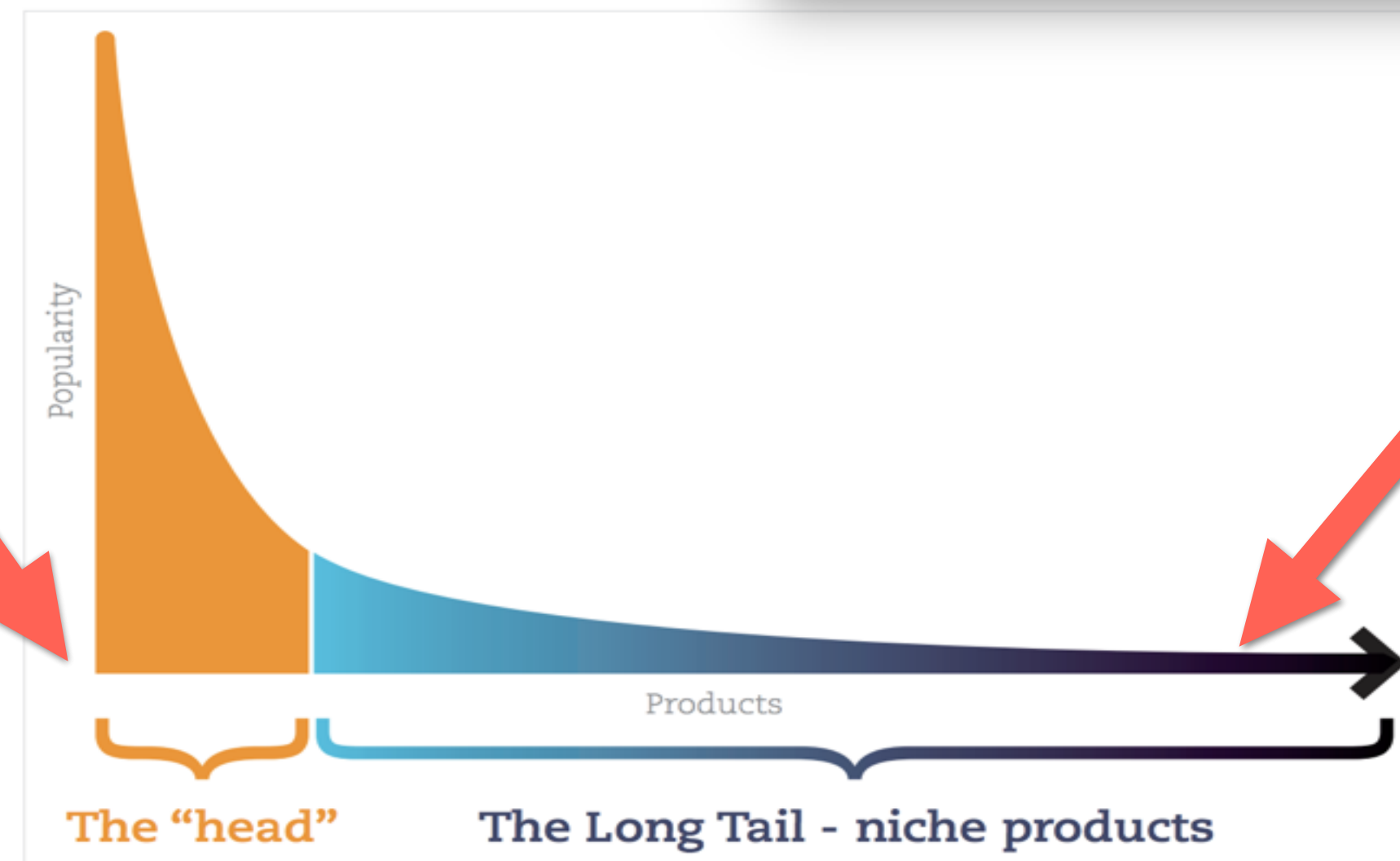
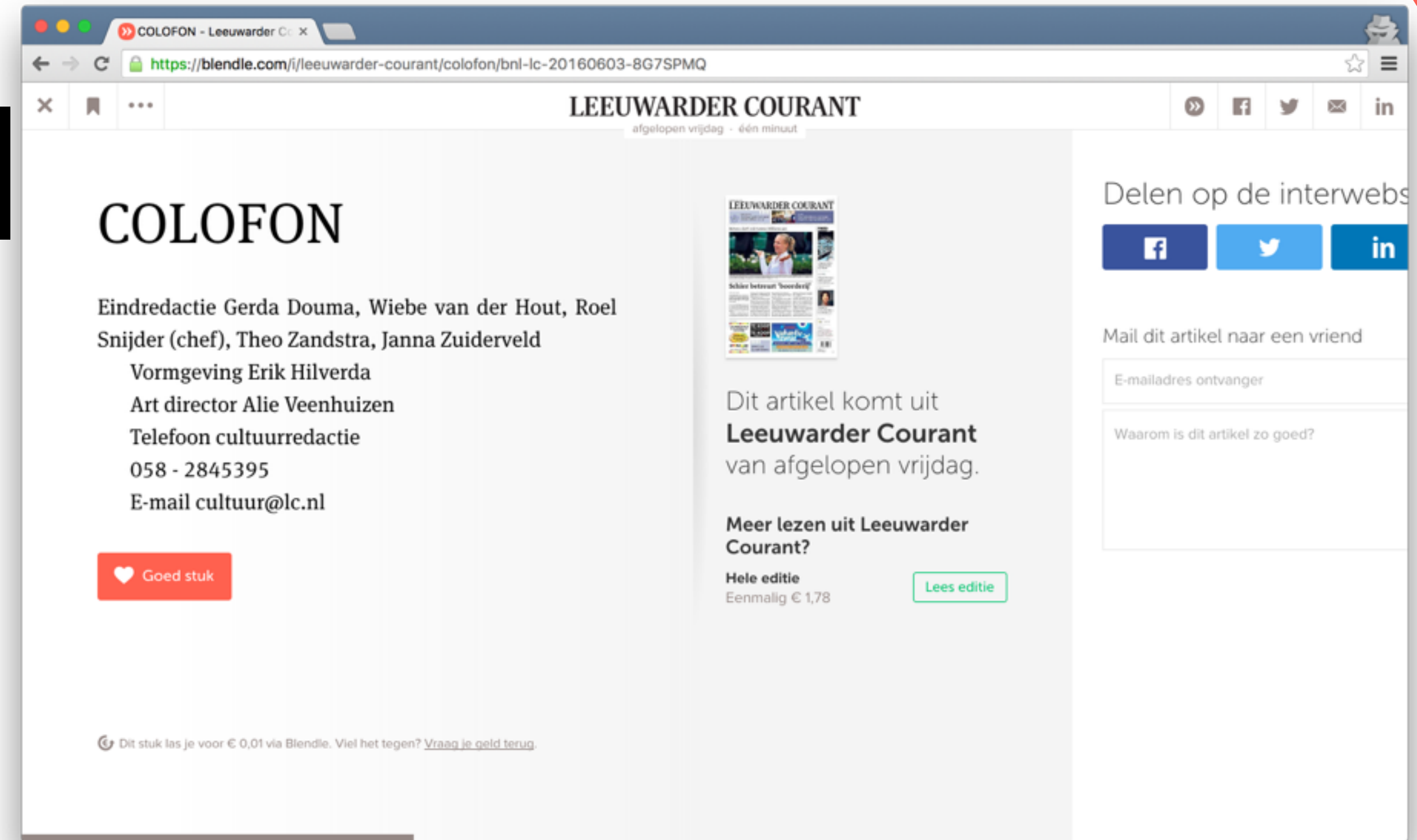
Rankings at Blendle

- Should be personalised
 - Lots of content
 - Lots of different tastes



Rankings at Blendle

- Should be personalised
 - Lots of content
 - Lots of different tastes



Lustige dames - LINDA. - x

https://blendle.com/i/linda/lustige-dames/bnl-linda-20160120-120550_lustige_dames

LINDA. woensdag 20 januari 2016 - 5 minuten

Lustige dames

Tv-vrouw Geraldine Kemper (25) en model Kim Feenstra (30) rollebollen graag met vrouwen, maar willen de mannen niet missen. "Als het maar geen softie is."

Tekst Eva Loesberg, fotografie Marc De Groot, styling Jos Van Heel, visagie Irena Ruben

VROUWEN

Kim "Ik val op mannen én vrouwen. Daar ben ik kort geleden pas eerlijk voor uitgekomen. Op mijn achttiende heb ik een lesbische relatie gehad."

Geraldine "Ik heb nog nooit een relatie gehad met een vrouw, maar ik speel wel met vrouwen."

Kim "Ja, misschien was die relatie ook meer spelen. Het was sex. Het was ontdekken. Het duurde drie maanden. Vrouwen zijn zo zacht. Fijn om te voelen."

Geraldine "Ik vind sex met een vrouw ook fijn. Ik word aangetrokken door een persoon – of het nou een man of een vrouw is kan me niet schelen."

Kim "Het hippie-idee staat me aan: love everyone; geen hokjes, vrijheid."

RODDEL

Kim "Sinds ik in *Sputten en Slikken* heb laten merken dat ik Geraldine heel aantrekkelijk vind, gaat het gerucht dat er iets tussen ons is gebeurd."

Geraldine "Maar daar gaan we verder niets over zeggen. Sommige dingen blijven privé."

Kim "Die blijven in de slaapkamer."

Geraldine "We doen mee aan deze LINDA.-shoot omdat we fotograaf Marc de Groot geweldig vinden. Met hem weet je dat het classy wo...

MODE

Geraldine "We k... zou ik dat nooit e... genoeg voel ik m..."

Kim "Of iets sexy draagt. Bij jou is..."

Geraldine "Jij zag... van Christian Lo..."

Kim "Ik vond he..."

Geraldine "Norm... truitje. Simpel. C... Bezweet. In een l... maxjes. En dan h..."

Kim "Ik heb een... Lieslaarzen vind...

COLOFON - Leeuwarder Courant - x

https://blendle.com/i/leeuwarder-courant/colofon/bnl-lc-20160603-8G7SPMQ

LEEWARDER COURANT afgelopen vrijdag - één minuut

COLOFON

Eindredactie Gerda Douma, Wiebe van der Hout, Roel Snijder (chef), Theo Zandstra, Janna Zuiderveld

Vormgeving Erik Hilverda

Art director Alie Veenhuizen

Telefoon cultuurredactie 058 - 2845395

E-mail cultuur@lc.nl

Goed stuk

Dit artikel las je voor € 0,01 via Blendle. Viel het tegen? [Vraag je geld terug.](#)

Delen op de interwebs

Mail dit artikel naar een vriend

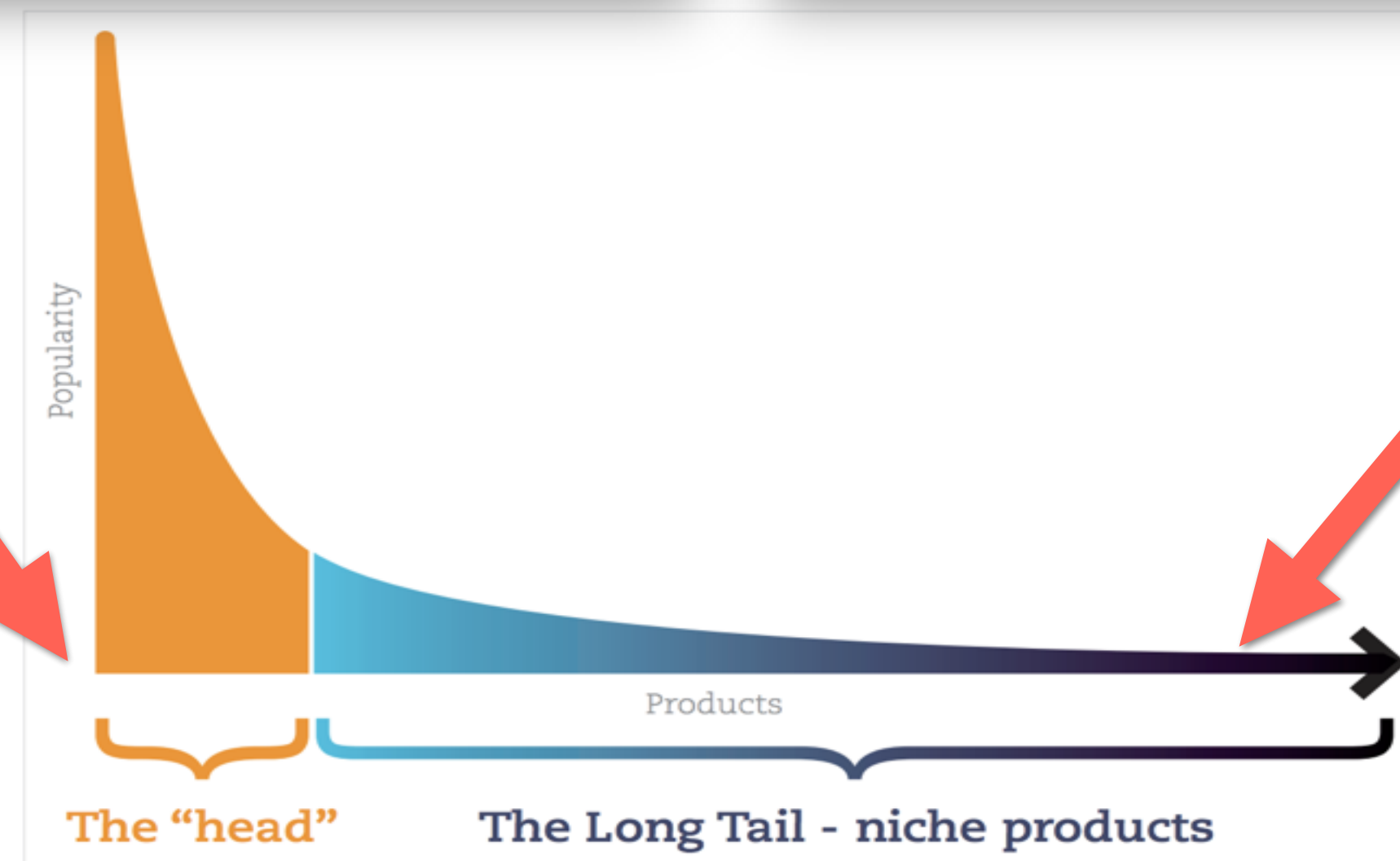
E-mailadres ontvanger

Waarom is dit artikel zo goed?

Dit artikel komt uit **Leeuwarder Courant** van afgelopen vrijdag.

Meer lezen uit Leeuwarder Courant?

Hele editie Eenmalig € 1,78 [Lees editie](#)



Rankings at Blendle

- Should be personalised
 - Lots of content
 - Lots of different tastes

Rankings at Blendle

- Should be personalised
 - Lots of content
 - Lots of different tastes
- Can be personalised

Rankings at Blendle

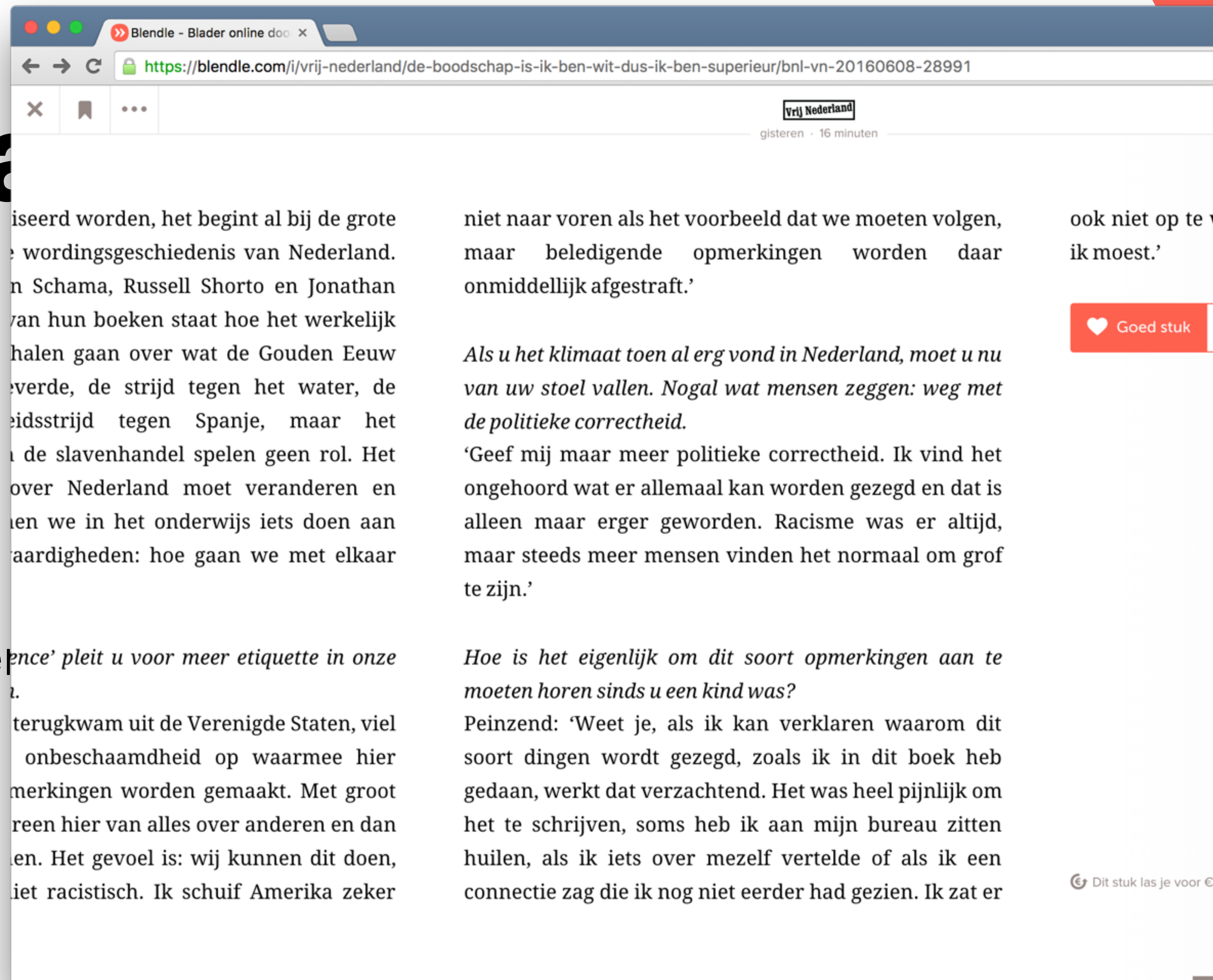
- Should be personalised
 - Lots of content
 - Lots of different tastes
- Can be personalised
 - We know what our users read

Rankings at Blendle

- Should be personalised
 - Lots of content
 - Lots of different tastes
- Can be personalised
 - We know what our users read
 - What they didn't like

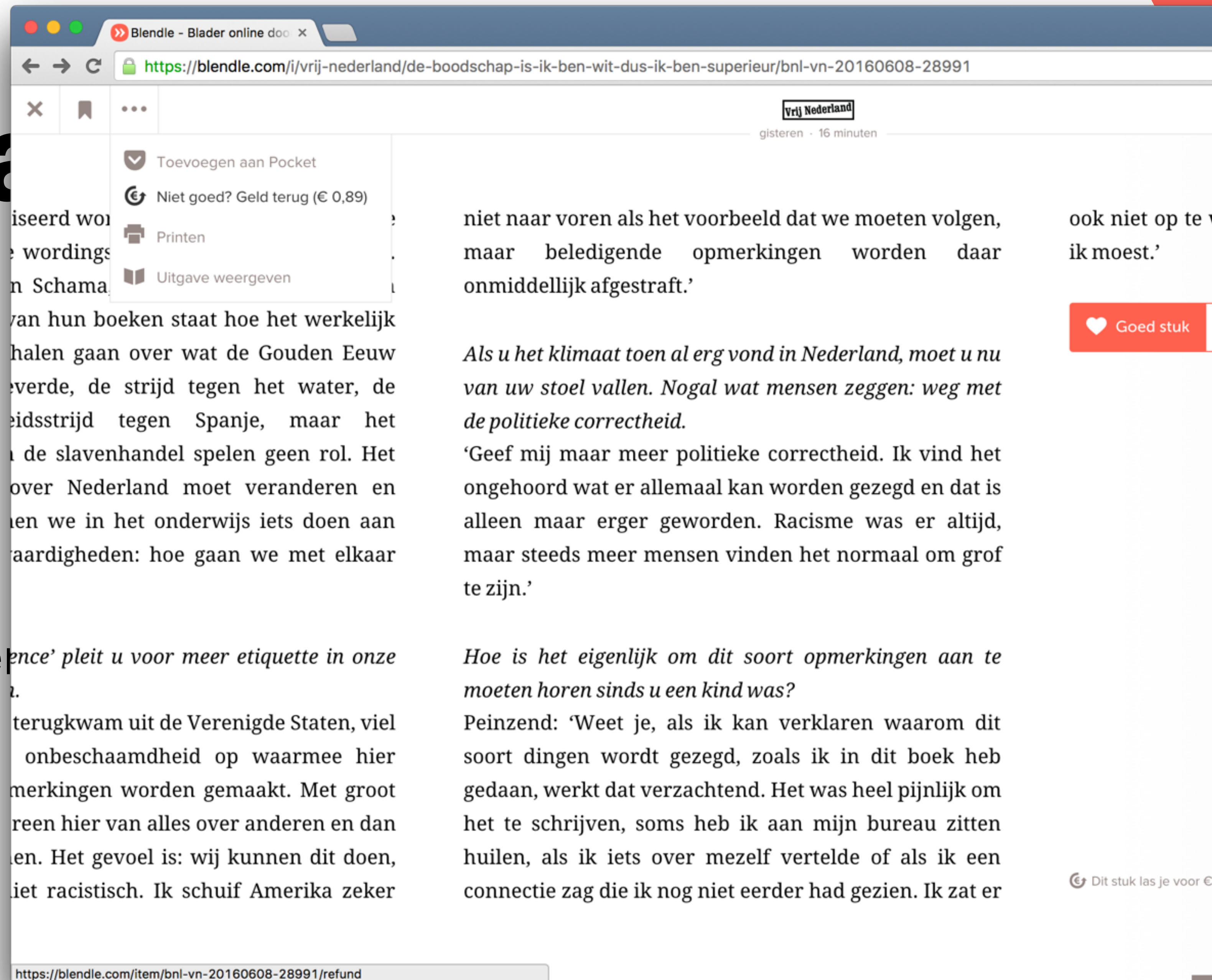
Rankings a

- Should be personalised
 - Lots of content
 - Lots of different tastes
- Can be personalised
 - We know what our users like
 - What they didn't like



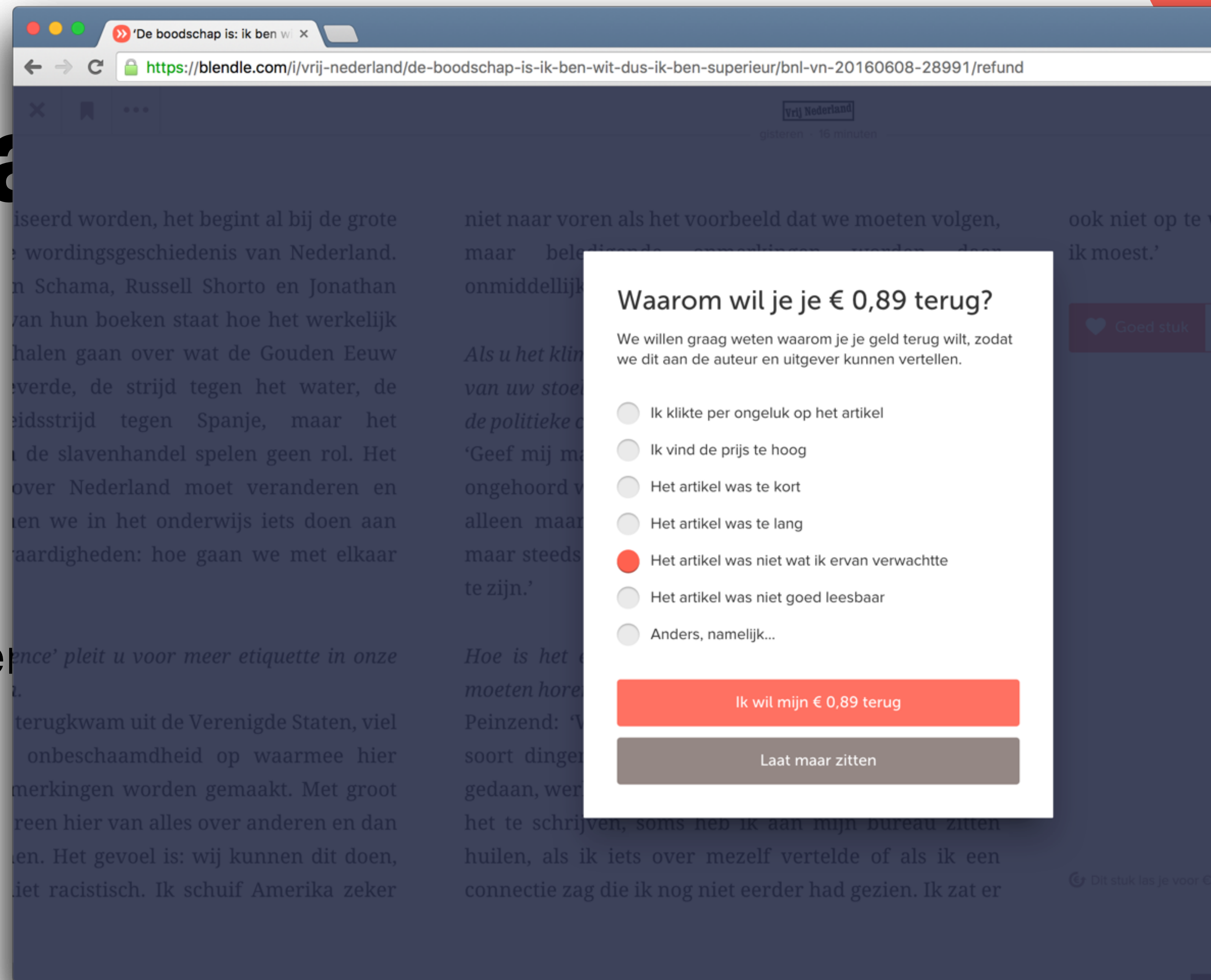
Rankings a

- Should be personalised
 - Lots of content
 - Lots of different tastes
- Can be personalised
 - We know what our users like
 - What they didn't like



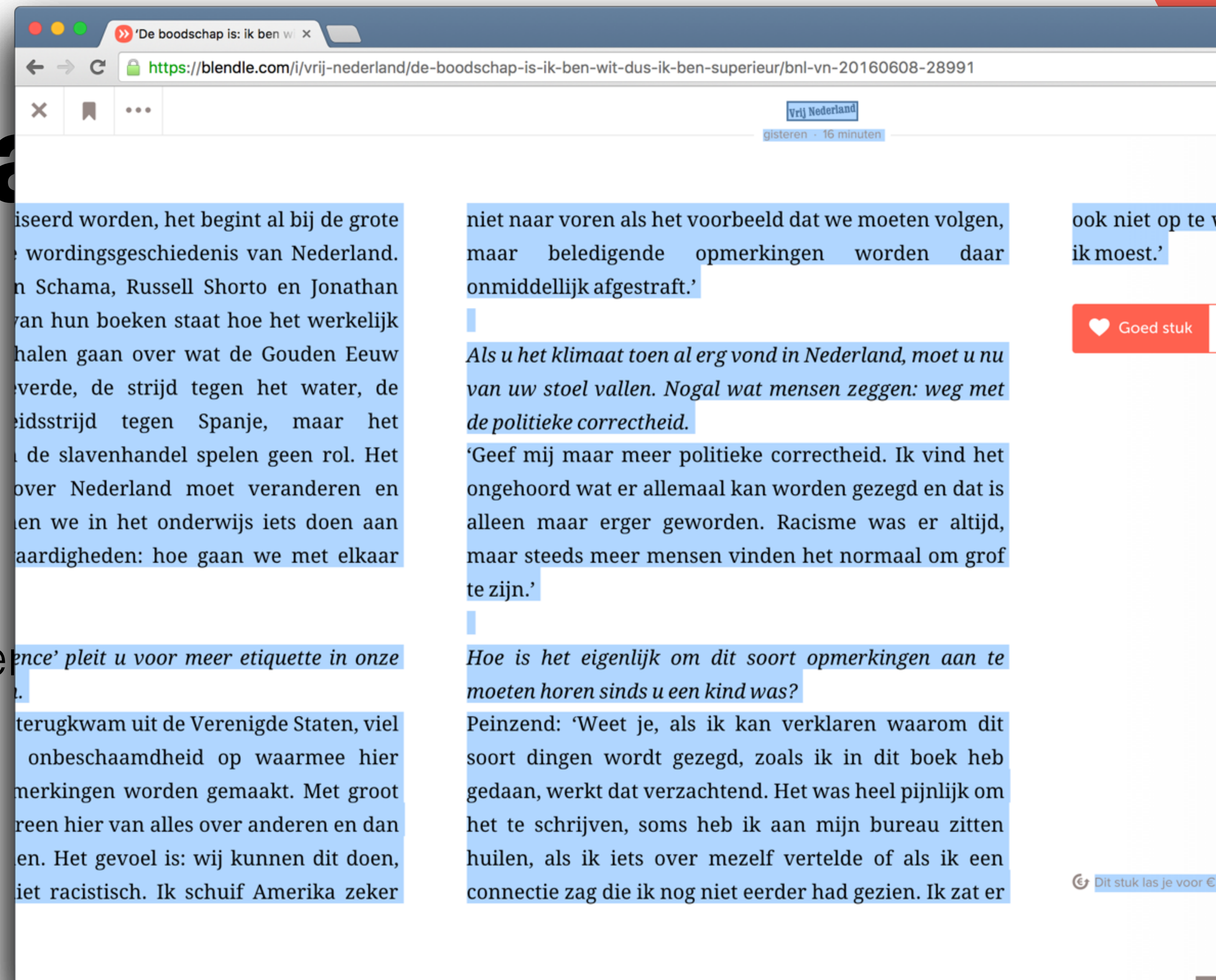
Rankings a

- Should be personalised
 - Lots of content
 - Lots of different tastes
- Can be personalised
 - We know what our users like
 - What they didn't like



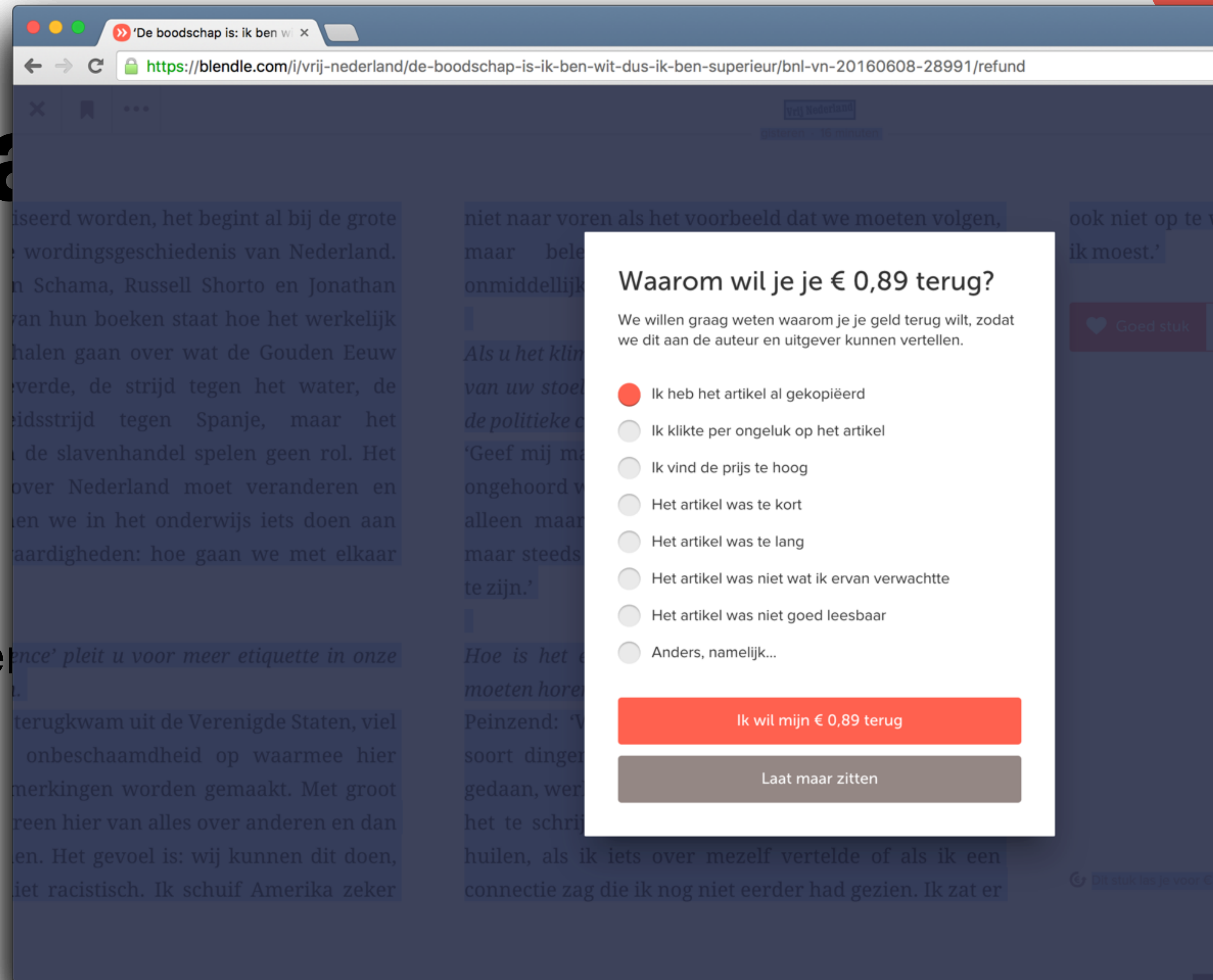
Rankings a

- Should be personalised
 - Lots of content
 - Lots of different tastes
- Can be personalised
 - We know what our users like
 - What they didn't like



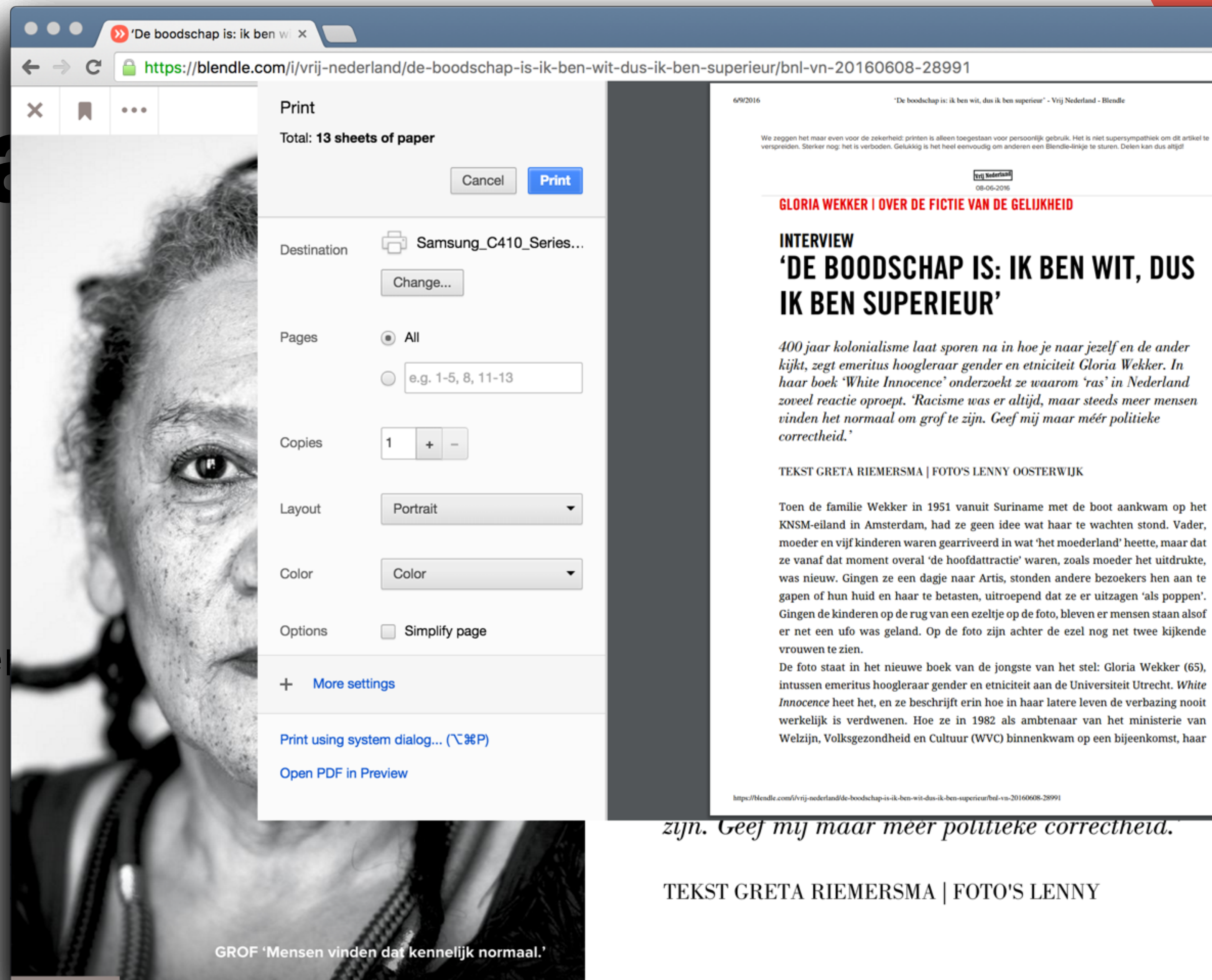
Rankings a

- Should be personalised
 - Lots of content
 - Lots of different tastes
- Can be personalised
 - We know what our users like
 - What they didn't like



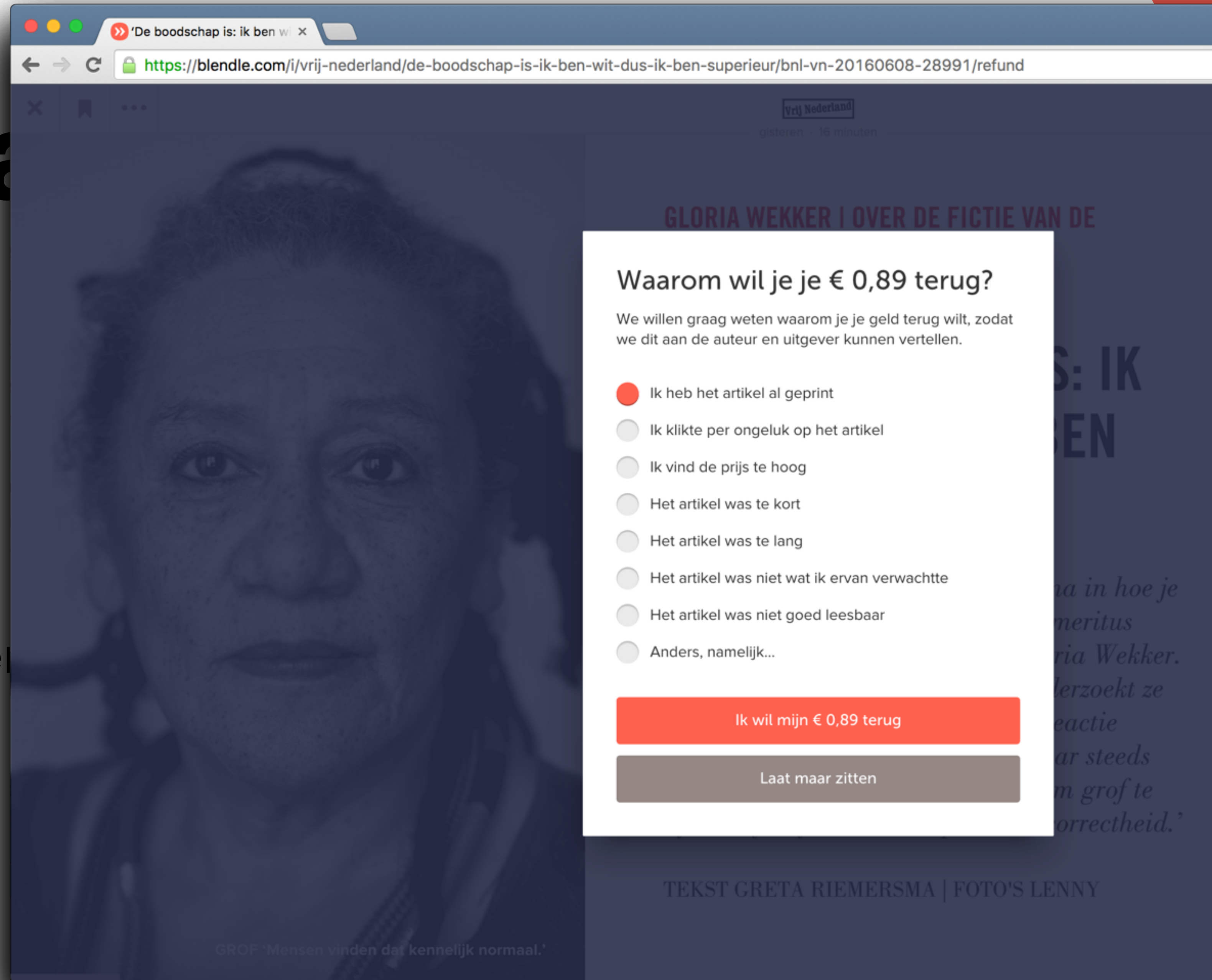
Rankings a

- Should be personalised
 - Lots of content
 - Lots of different tastes
- Can be personalised
 - We know what our users like
 - What they didn't like



Rankings a

- Should be personalised
 - Lots of content
 - Lots of different tastes
- Can be personalised
 - We know what our users like
 - What they didn't like



Rankings at Blendle

- Should be personalised
 - Lots of content
 - Lots of different tastes
- Can be personalised
 - We know what our users read
 - What they didn't like

Rankings at Blendle

- Should be personalised
 - Lots of content
 - Lots of different tastes
- Can be personalised
 - We know what our users read
 - What they didn't like
 - When they read

Rankings at Blendle

- Should be personalised
 - Lots of content
 - Lots of different tastes
- Can be personalised
 - We know what our users read
 - What they didn't like
 - When they read
 - ...

Personalised recommendations



Music

- **Repeated** consumption
- More **niche** than movies/series
- **Tens of millions** of songs and users
- **Long** (often unlimited) shelf life
- Feedback: implicit (listening behaviour)

Personalised recommendations



Music

- **Repeated** consumption
- More **niche** than movies/series
- **Tens of millions** of songs and users
- **Long** (often unlimited) shelf life
- Feedback: implicit (listening behaviour)



Movies and series

- **Single** consumption
- Less niche than music
- 75.000.000 users
- **Long** shelf life
- Feedback: explicit ratings

Personalised recommendations



Music

- **Repeated** consumption
- More **niche** than movies/series
- **Tens of millions** of songs and users
- **Long** (often unlimited) shelf life
- Feedback: implicit (listening behaviour)



Movies and series

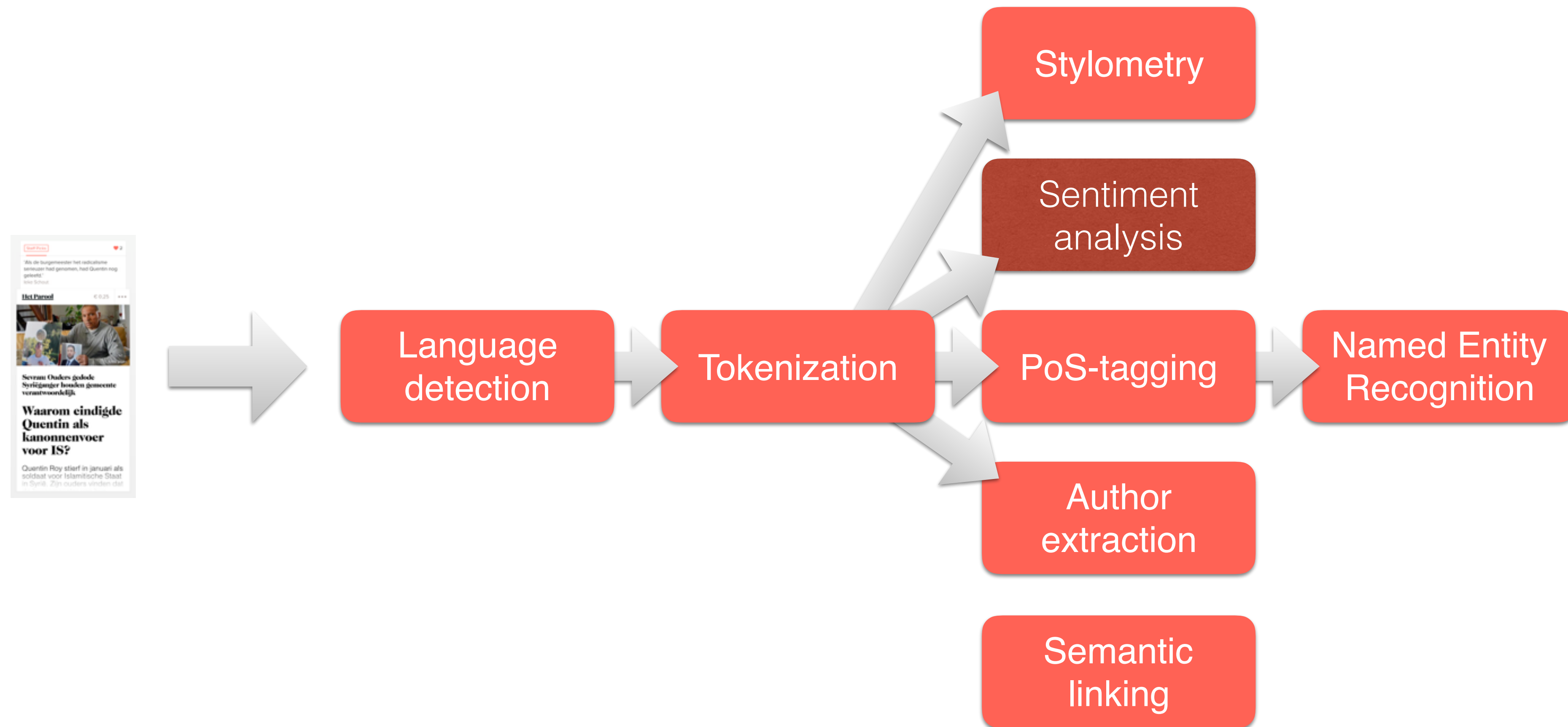
- **Single** consumption
- Less niche than music
- 75.000.000 users
- **Long** shelf life
- Feedback: explicit ratings



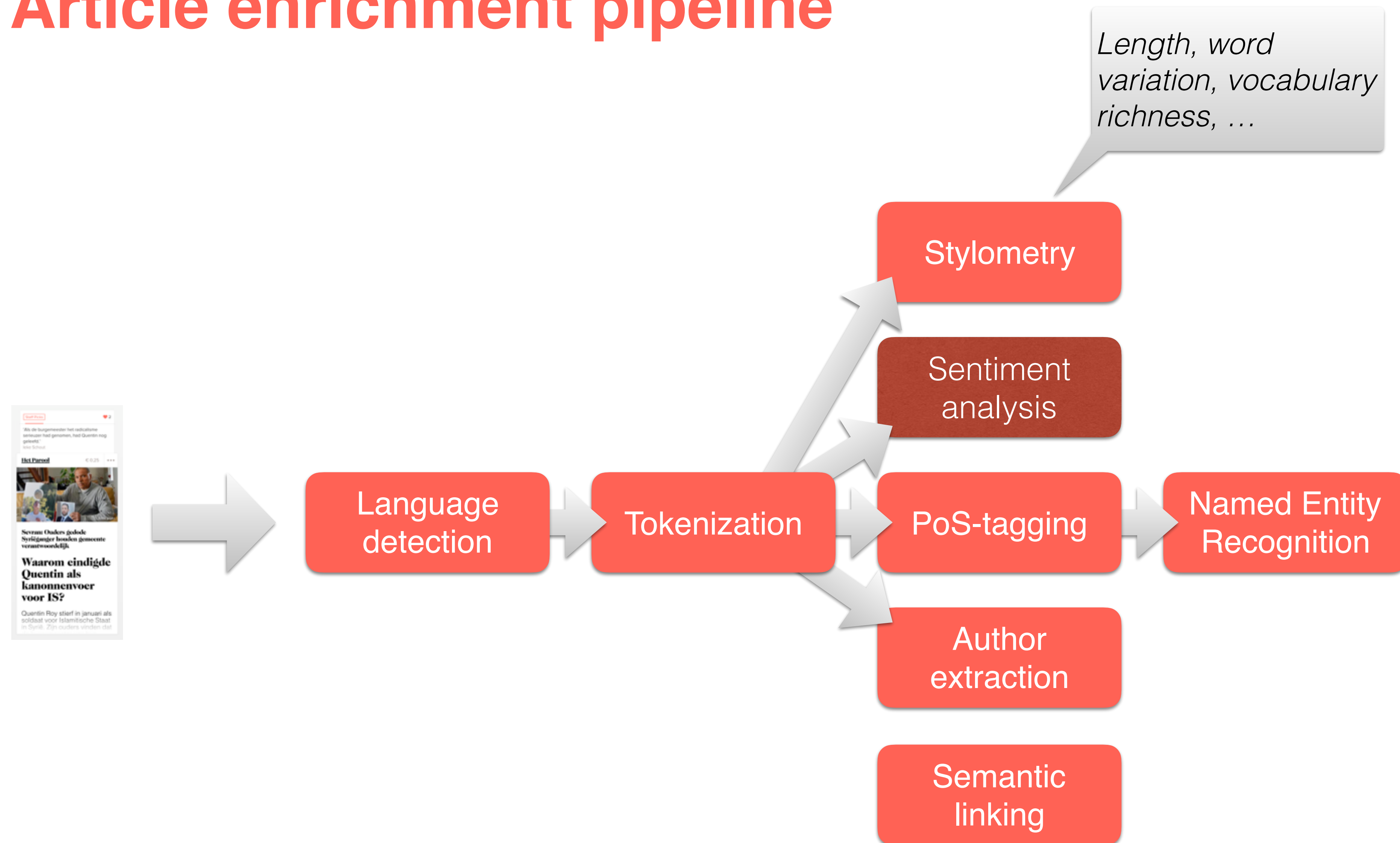
Articles

- **Single** consumption
- Very **short** shelf-life → cold start problem, Collaborative Filtering won't work
- **Diversity** of suggestions very important → Content Based Recommendation won't work
- A lot of information **overlap**
- Feedback: implicit through reading behaviour, refunds, also 'like' option
- Changing relevance (niche vs popular)

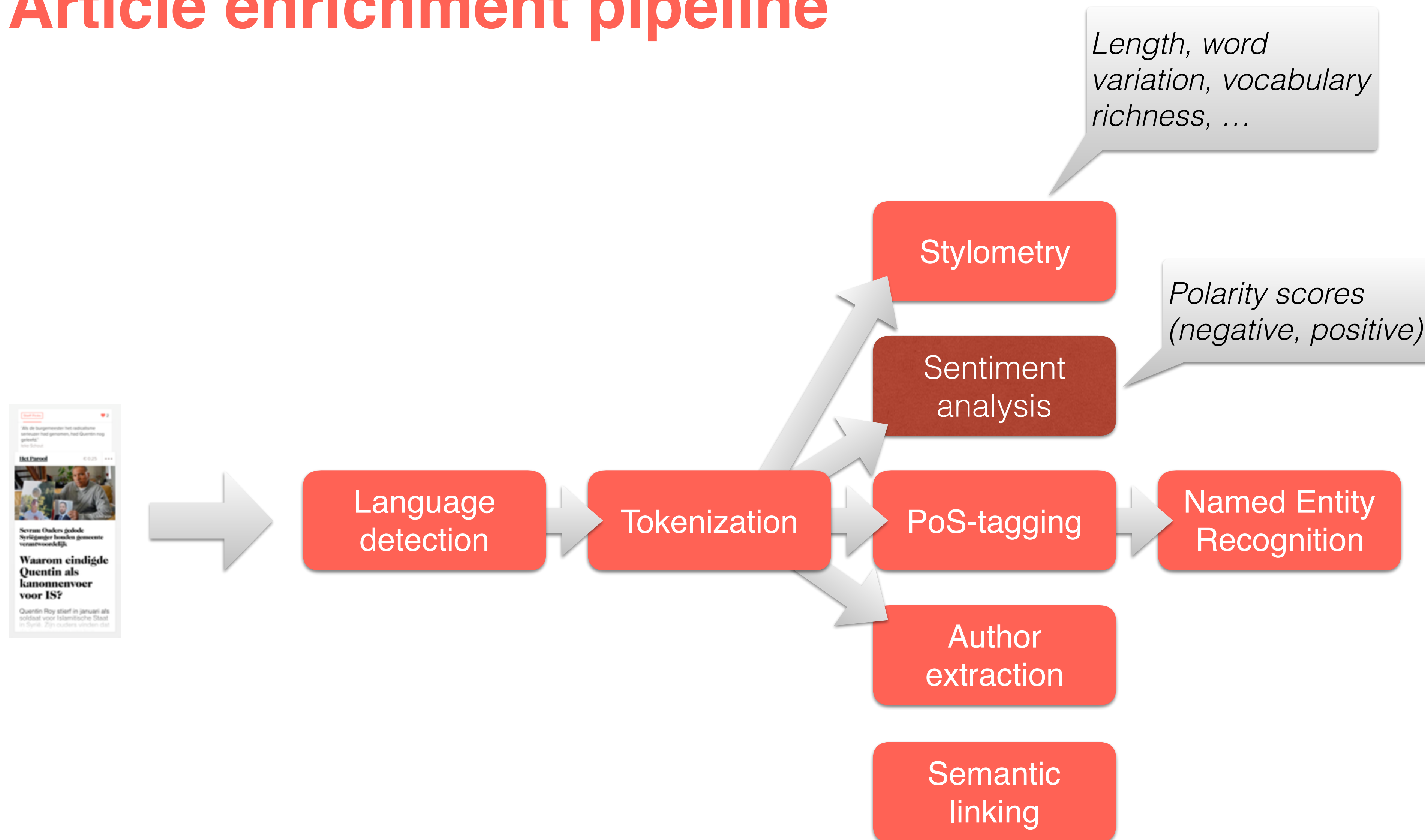
Article enrichment pipeline



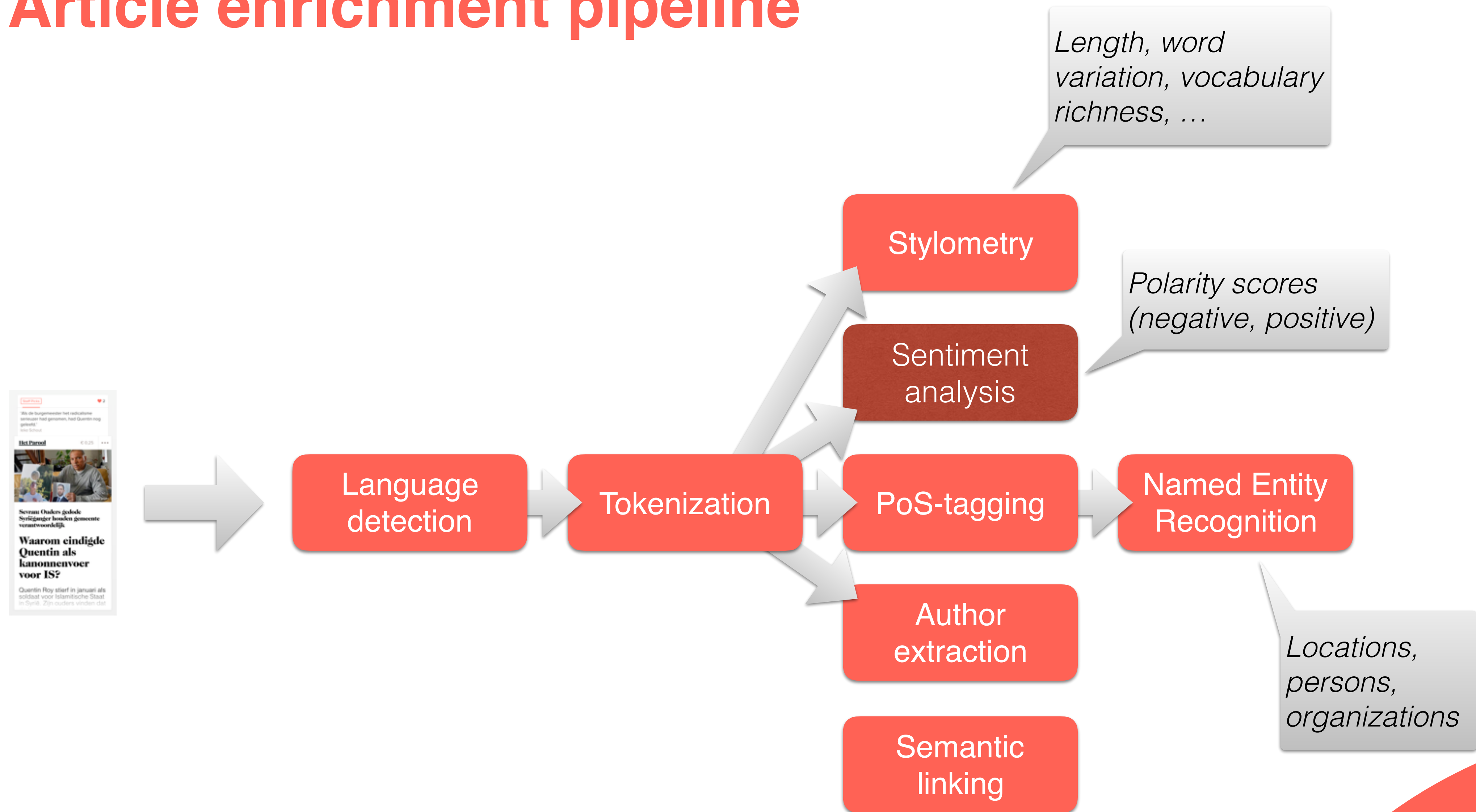
Article enrichment pipeline



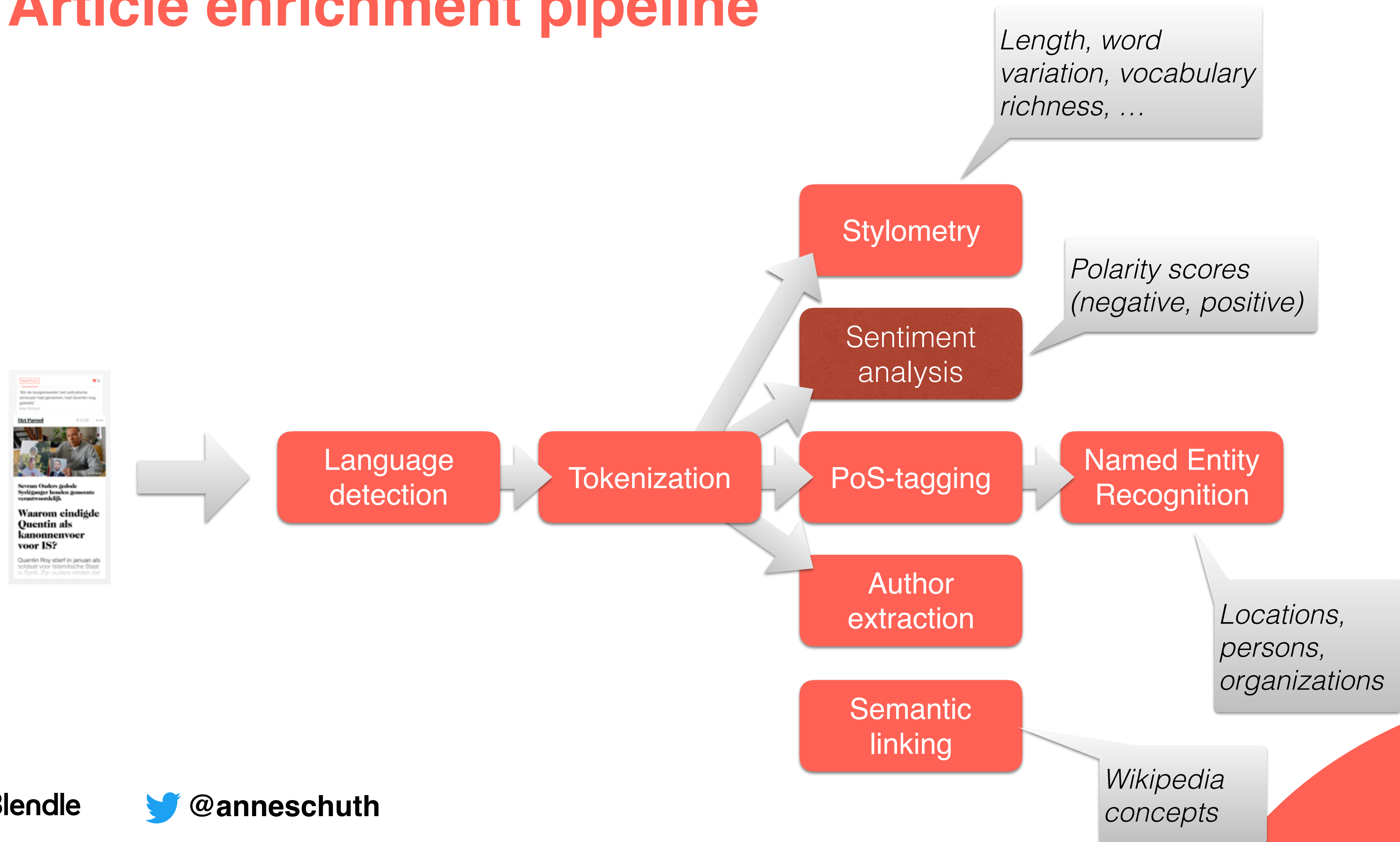
Article enrichment pipeline



Article enrichment pipeline



Article enrichment pipeline



Article enrichment pipeline

Feature: Enrich item pipeline

Scenario: Enrich an item

Given I have the following item data

```
"""json
{"_links":{"self":{"href":"s3://static.blendle.nl/publication/par/2015/01/27/item/3895280/version/1/private/content.json"}},
"body":[{"content":"Smulpapen","type":"kicker"},
{"content":"Tom Kellerhuis stoomt uit allerhande keukens bedrieglijk simpele gerechten klaar.","type":"intro"},
{"content":"Dronken vegetariër","type":"hl1"}, {"content":"TOM KELLERHUIS","type":"byline"},
{"content":"Ik ken een flink aantal restaurateurs dat gruwet bij het woord vegetariër. Ik ken er zelfs een paar die
hem bewust uit hun zaak weren. \"U eet geen vlees of vis? Wat doet u hier dan?\" hoorde ik ooit als pissig antwoord
op de vraag of er soms ook gerechten zónder vis of vlees op de kaart stonden.","type":"p"}, {"content":"Een goeie
kok wil nu eenmaal laten zien wat hij in huis heeft. Dat zit hem vooral in de vlees- en visbereidingen, of in de
sauzen, waarin vaak ook vlees- en visproducten zijn verwerkt. Vanuit dat oogpunt snap ik de kok wel. Maar steeds
meer mensen kiezen ervoor om minder vlees of vis te consumeren, om hen moverende redenen. De één denkt aan het
milieu en de consumptiemaatschappij, de ander aan het dierenwelzijn, weer een ander louter aan zijn eigen
gezondheid.","type":"p"},
{"content":"Hoewel er aan het woord vegetariër nog steeds een sektarisch randje kleeft, mogen we toch wel stellen
dat hij tegenwoordig lang zo suf niet meer is als het aloude stereotype beeld: het op sandalen – waarin platvoeten
in geitenwollensokken – sjokkende en van elke humor gespeende tiepje, met linnen draagtas op zoek naar gedroogde
sojabonen en tempé voor de avondmaaltijd.","type":"p"}, {"content":"Ik kreeg laatst het receptenboek De dronken
vegetariër van Floor Overgoor, Hanny Roskamp en Yannick Slagter toegezonden, dat gehakt maakt van dat oude cliché.
Allemaal vegagerechten met een tic (en soms flink wat, zoals in bijgaand recept. Roskamp schrijft in haar voorwoord:
'Ve zijn gaan beseffen dat die dronken vegetariër een betere ambassadeur is voor een diervriendelijke wereld dan een
vegetariër die zichzelf opsluit achter een muur van morele superioriteit.'","type":"p"}, {"content":"Welaan, het
beeld is dus aan het kantelen. Ik hoop snel eens zo'n dronken vegetariër tegen te komen!","type":"p"},
{"content":"Reageren? t.kellerhuis@parool.nl","type":"p"}, {"content":"Wijnfondue","type":"ph"},
{"content":"(voor 2 personen)","type":"p"}, {"content":"1/2 liter vega-bouillon, 1/2 fles pinot noir, 1/2 aubergine
in plakken, 200 g oesterzwammen, 200 g bospaddenstoelen, 1 courgette in plakken, 200 g tofoe in repen.","type":"p"},
{"content":"Simpeler en romantischer dan dit wordt het niet. Maak de bouillon, breng aan de kook, giet de wijn erbij
en breng opnieuw aan de kook (maar laat niet doorkoken: zonde van de wijn!).","type":"p"}, {"content":"Zet dit
heerlijke vocht op een fonduestel, warmhoudplaatje of theelicht. Gooi een paar flinke handen groentes en tofoe in
het hete vocht en haal de stukken er na een minuut of 5 weer uit; liefst met zo'n Chinees fonduemandje, maar een
vork werkt ook prima.","type":"p"}, {"content":"Laat de volgende portie groente en tofoe garen in het vocht, terwijl
je de eerste portie opeet.","type":"p"}, {"content":"Eet de stukjes groente met brood en knoflookboter. Dippen in
knoflooksaus kan uiteraard ook.","type":"p"}],
"date":"2015-01-27T00:00:00+00:00",
"format_version":4,
"id":"bnl-par-20150127-3895280",
"issue":{"id":"bnl-par-20150127"},
"item_index":70,
"provider":{"id":"par"}}
"""
```


When I send it through the enrichment pipeline language|tokenizer|postagger|ner|stylometry|author|semanticizer|sentiment

Then I should have the following enrichments in the result

enrichment	value	
language	nl	
author	TOM KELLERHUIS	

And I should have the following stylometry enrichments in the result

enrichment	value	
number_of_paragraphs	12	
number_of_words	432	
article_length	2542	
vocab_richness	40.5376820773	
word_variation	0.585648148148	
avg_word_length	4.98611111111	
word_size_short	0.553240740741	
word_size_medium	0.391203703704	
word_size_long	0.0555555555556	
hapax_legomena	0.4375	
hapax_dislegomena	0.0763888888889	
number_of_images	1	

And I should have the following semantic links in the result

url	title	text	senseProbability	count	norm	
http://nl.wikipedia.org/wiki/Avondmaaltijd	Avondmaaltijd	avondmaaltijd	0.36666666666666664	1	http://dbpedia.org/resource/Supper	
http://nl.wikipedia.org/wiki/Tic	Tic	tic	1.0416666666666667	1	http://dbpedia.org/resource/Tic	
http://nl.wikipedia.org/wiki/Platvoeten	Platvoeten	platvoeten	0.47058823529411764	1	http://dbpedia.org/resource/Flat_feet	
http://nl.wikipedia.org/wiki/Groente	Groente	groente	0.41304347826086957	2	http://dbpedia.org/resource/Vegetable	
http://nl.wikipedia.org/wiki/Sojaboon	Sojaboon	sojabonen	0.6206896551724138	1	http://dbpedia.org/resource/Soybean	
http://nl.wikipedia.org/wiki/Vegetarisme	Vegetarisme	vegetariër	2.6956521739130435	6	http://dbpedia.org/resource/Vegetarianism	
http://nl.wikipedia.org/wiki/Pinot%20noir	Pinot noir	pinot noir	0.8827586206896552	1	http://dbpedia.org/resource/Pinot_noir	
http://nl.wikipedia.org/wiki/Knoflooksaus	Knoflooksaus	knoflooksaus	0.35294117647058826	1		
http://nl.wikipedia.org/wiki/Consumptiemaatschappij	Consumptiemaatschappij	consumptiemaatschappij	0.5121951219512195	1		
http://nl.wikipedia.org/wiki/Courgette	Courgette	courgette	0.8421052631578947	1	http://dbpedia.org/resource/Zucchini	
http://nl.wikipedia.org/wiki/Tofoe	Tofoe	tofoe	0.8275862068965517	3	http://dbpedia.org/resource/Tofu	
http://nl.wikipedia.org/wiki/Bouillon%20%28voedsel%29	Bouillon (voedsel)	– bouillon	0.5752212389380531	1	http://dbpedia.org/resource/Broth	
http://nl.wikipedia.org/wiki/Bouillon%20%28voedsel%29	Bouillon (voedsel)	bouillon	0.5752212389380531	2	http://dbpedia.org/resource/Broth	
http://nl.wikipedia.org/wiki/Aubergine%20%28plant%29	Aubergine (plant)	aubergine	0.8666666666666667	1	http://dbpedia.org/resource/Eggplant	
http://nl.wikipedia.org/wiki/Kensoort	Kensoort	ken	0.8787878787878788	2		

And I should have the following named entities in the result

entity	tag	count	
TOM KELLERHUIS	PER	1	
FLOOR OVERGOOR	PER	1	
HANNY ROSKAMP	PER	2	
YANNICK SLAGTER	PER	1	
WIJNFONDUE	LOC	1	
SIMPELER	PER	1	

User enrichment pipeline

Feature: Combine enriched items into a user profile

Scenario combine two enriched items

Given "item1" with these enrichments

enrichment	value	
language	nl	
author	TOM KELLERHUIS	

And these stylometry enrichments

enrichment	value	
number_of_paragraphs	12	
number_of_words	432	
article_length	2542	
vocab_richness	40.5376820773	
word_variation	0.585648148148	
avg_word_length	4.98611111111	

And these named entities

entity	tag	count	
YANNICK SLAGTER	PER	1	
WIJNFONDUE	LOC	1	
SIMPELER	PER	1	

Given "item2" with these enrichments

enrichment	value	
language	de	
author	PETER LANDERS	

And these stylometry enrichments

enrichment	value	
number_of_paragraphs	7	
number_of_words	600	
article_length	4560	
vocab_richness	61.62260711030082	
word_variation	0.6265060240963856	
avg_word_length	5.578313253012048	

And these named entities

entity	tag	count	
ANDRÉ SCHÜRRLE	PER	3	
VFL WOLFSBURG	ORG	4	
KAA	ORG	5	

Feature: Combine enriched items into a user profile

Scenario combine two enriched items

Given "item1" with these enrichments

enrichment	value
language	nl
author	TOM KELLERHUIS

And these stylometry enrichments

enrichment	value
number_of_paragraphs	12
number_of_words	432
article_length	2542
vocab_richness	40.5376820773
word_variation	0.585648148148
avg_word_length	4.98611111111

And these named entities

entity	tag	count
YANNICK SLAGTER	PER	1
WIJNFONDUE	LOC	1
SIMPELER	PER	1

Given "item2" with these enrichments

enrichment	value
language	de
author	PETER LANDERS

And these stylometry enrichments

enrichment	value
number_of_paragraphs	7
number_of_words	600
article_length	4560
vocab_richness	61.62260711030082
word_variation	0.6265060240963856
avg_word_length	5.578313253012048

And these named entities

entity	tag	count
ANDRÉ SCHÜRRLE	PER	3
VFL WOLFSBURG	ORG	4
KAA	ORG	5

When I run the `UserEnrichmentTask` task and it's dependencies

Then the user "anne" should have the following profile:

```
"""json
{
  "count": 2,
  "enrichments": {
    "language": [
      {
        "count": 1,
        "name": "nl"
      },
      {
        "count": 1,
        "name": "en"
      }
    ],
    "author": [
      {
        "count": 1,
        "name": "TOM KELLERHUIS"
      },
      {
        "count": 1,
        "name": "PETER LANDERS"
      }
    ]
  },
  "stylometry": [
    {
      "count": 2,
      "mean": 2177.5,
      "stdev": 515.4808434849931,
      "name": "article_length"
    },
    {
      "count": 2,
      "mean": 5.165410504361017,
      "stdev": 0.2535676336592832,
      "name": "avg_word_length"
    }
  ]
}
"last_100_purchased_items": ["item1", "item2"],
"tracking_uid": "anne"
"""
```


Machine Learning Features

When I compute features

Then I should have the following features and values:

feature	value
user_nr_purchased_items	5
user_nr_channels	1
user_nr_subscriptions	2
user_nr_providers	1
user_vmoney_almost_spent	1
user_alerts	1
user_new_edition	1
user_digest	1
user_magazine_digest	1
user_image_mean	0.5
user_likes_it_long	0
purchased_channel_intersect	1
purchased_channel_intersect_size	1
newsletter_purchased_channel_score	0.05
newsletter_purchased_provider_score	0.025
item_provider_followed_by_user	0
item_channel_followed_by_user	0
item_rating_score	0.9
item_number_of_images	1
item_length_class	0
user_item_language_score	0.5
user_item_author_in_top3	1
user_item_semlink_in_top10	1
user_item_semantic_overlap	0.097777777777777
user_item_named_entity_overlap	0.04081632653061
user_item_length_pdf	0.00316108841344

Learning to rank: preference learning



Learning to rank: preference learning



Learning to rank: preference learning

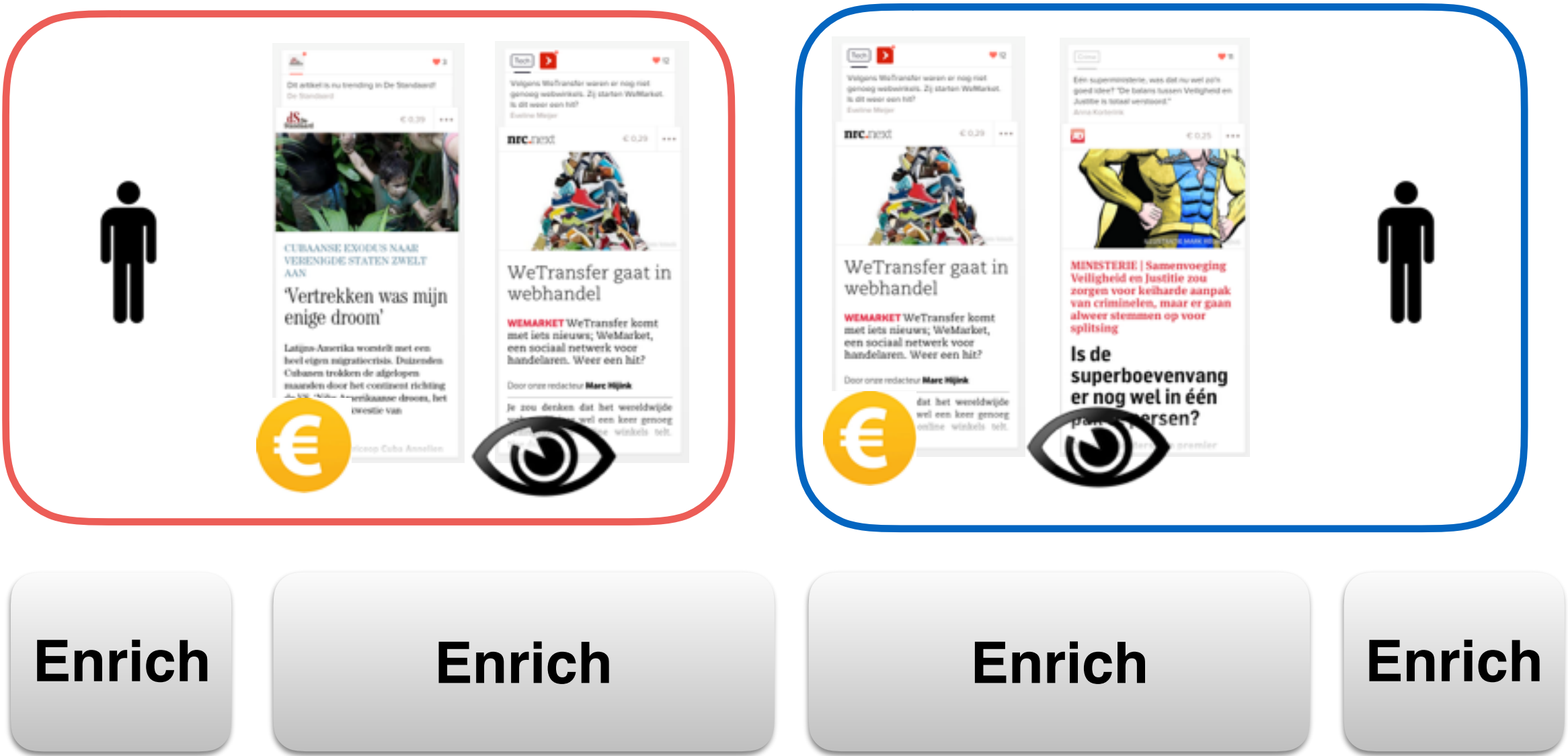




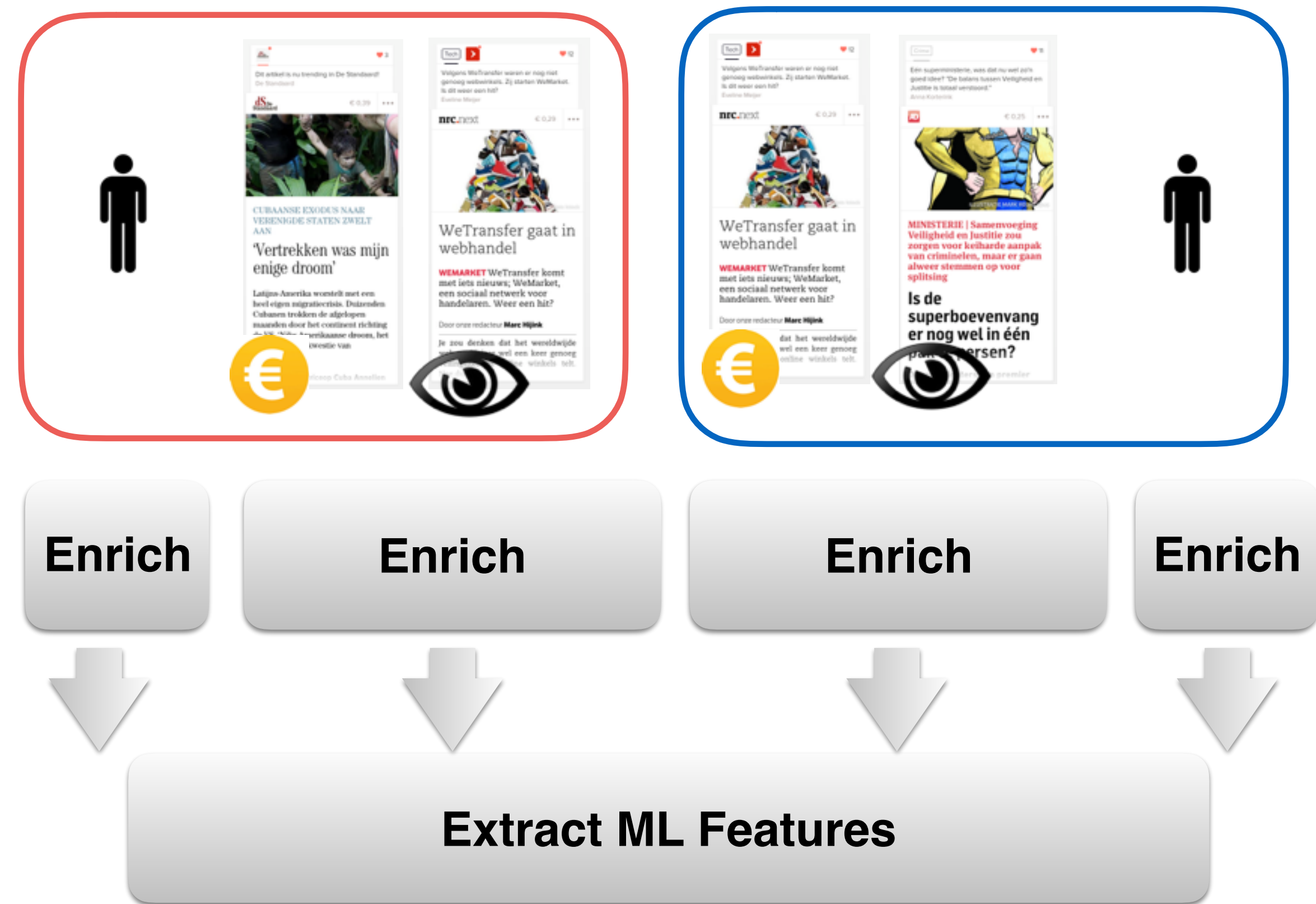




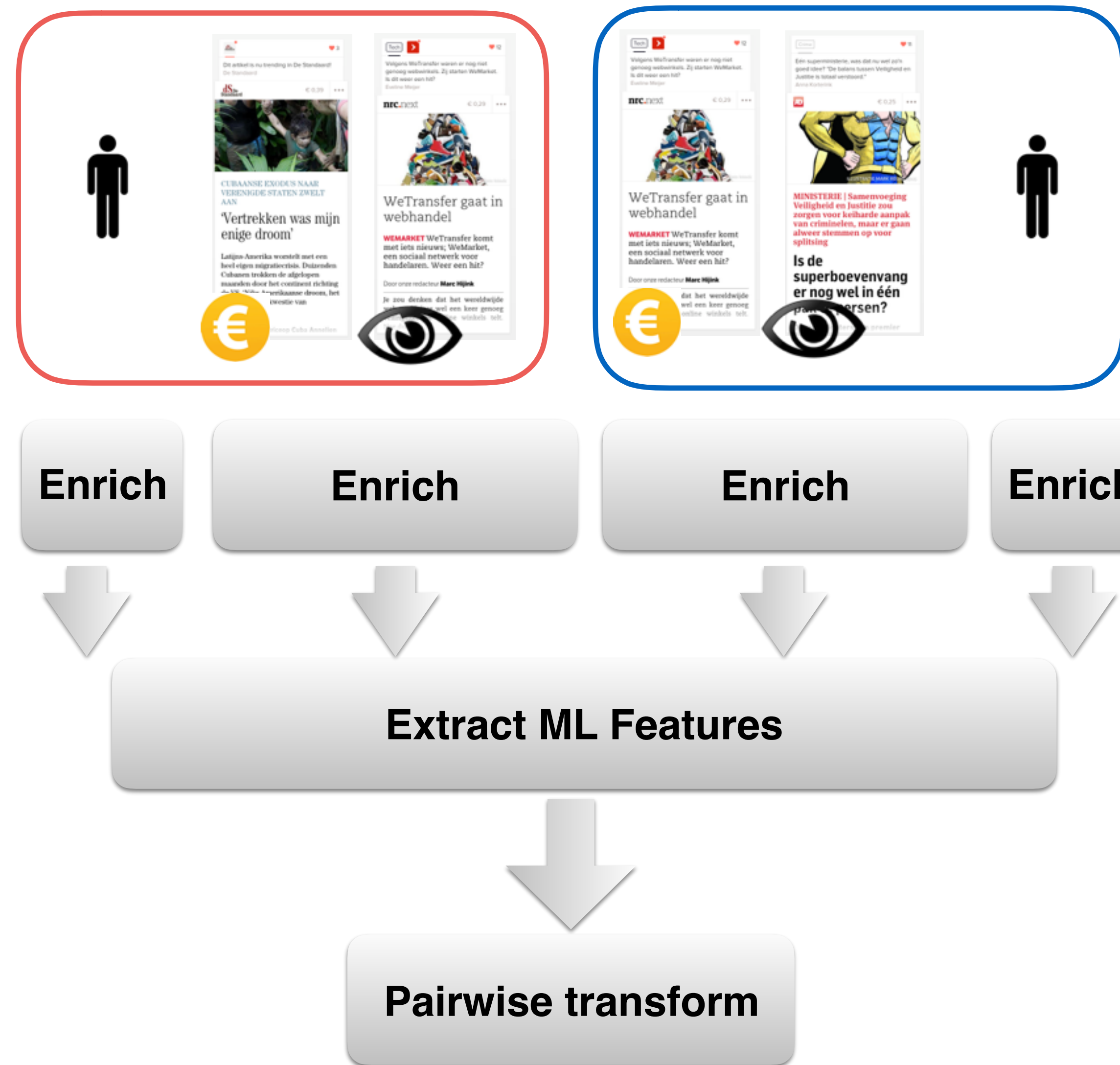
Learning to rank: preference learning



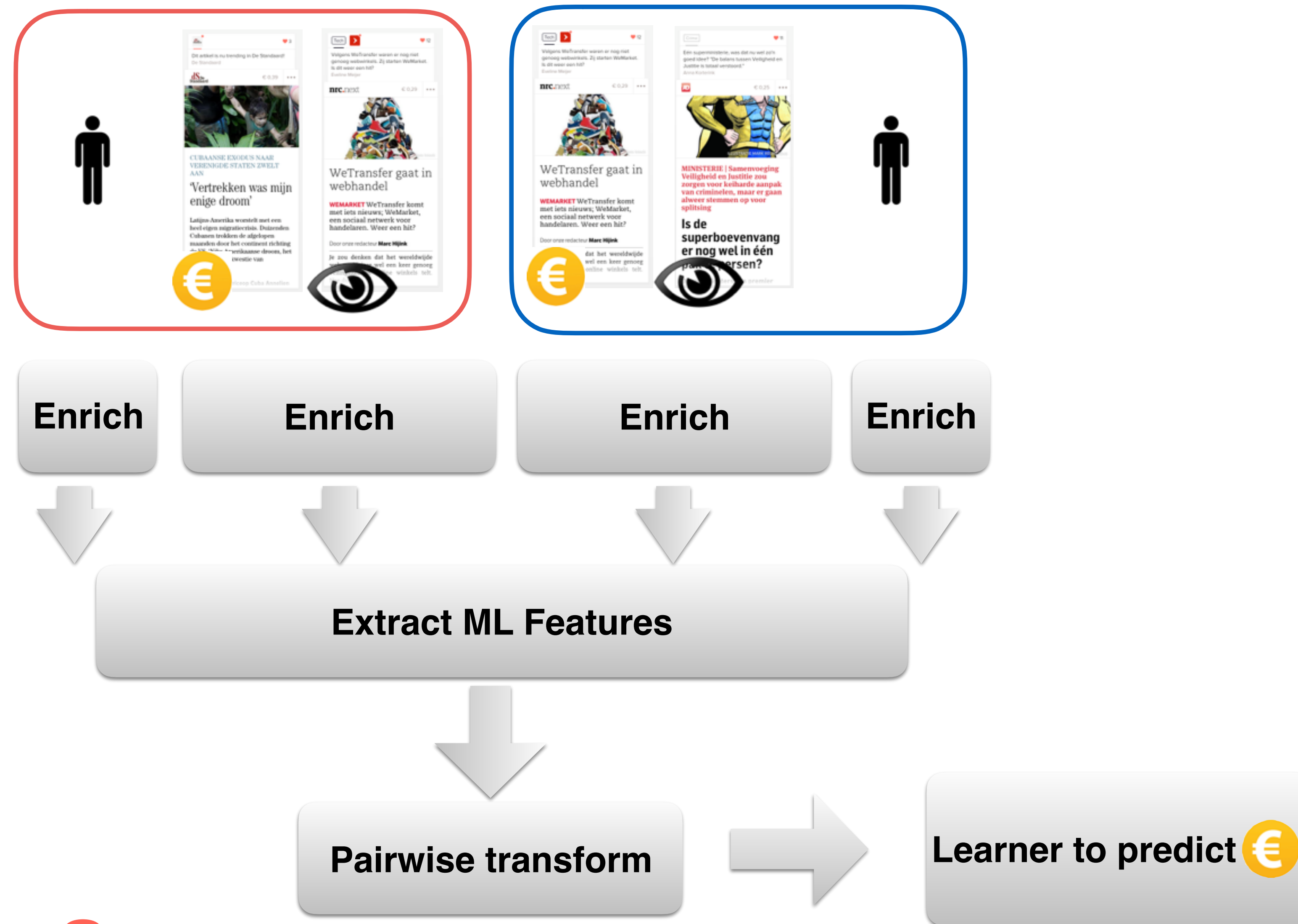
Learning to rank: preference learning



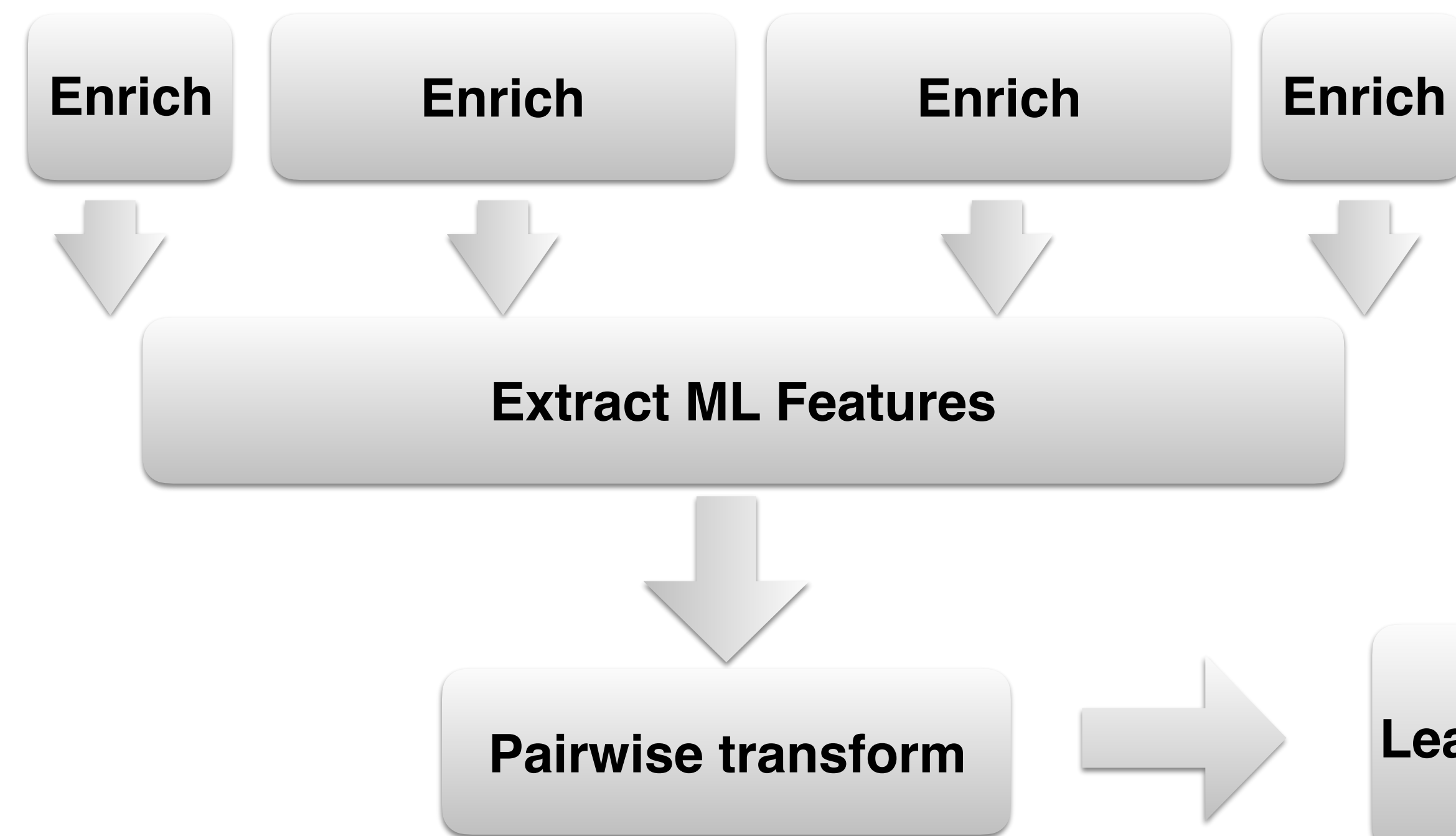
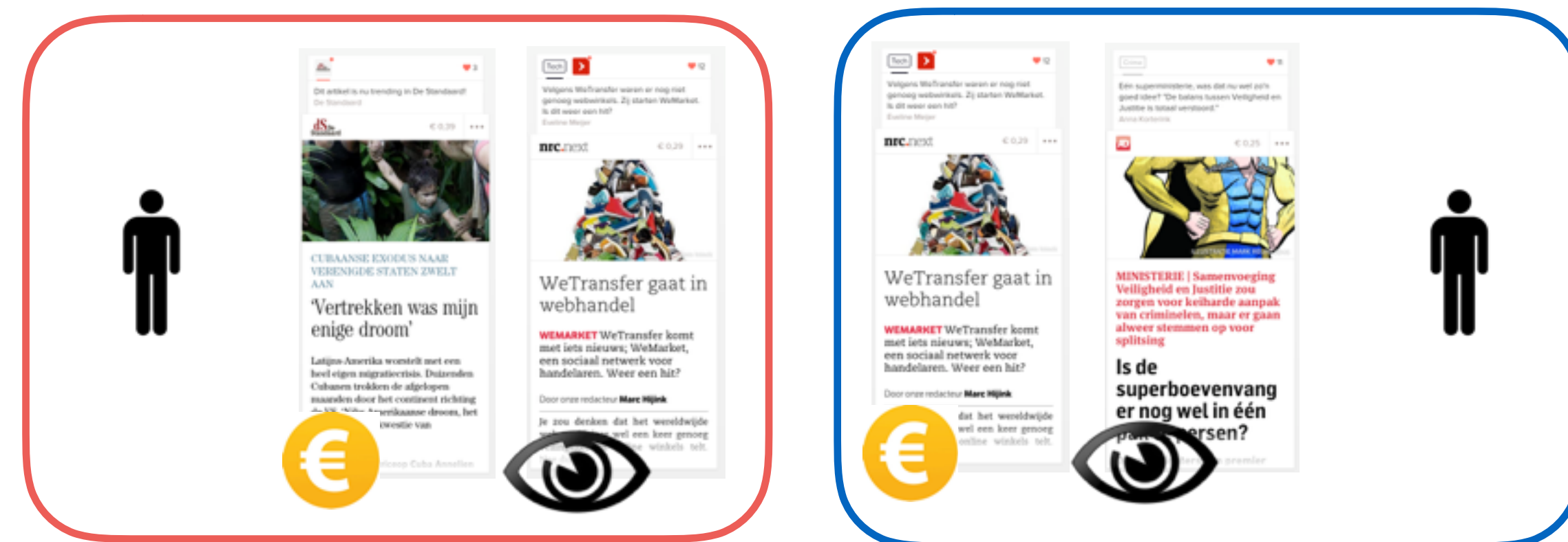
Learning to rank: preference learning



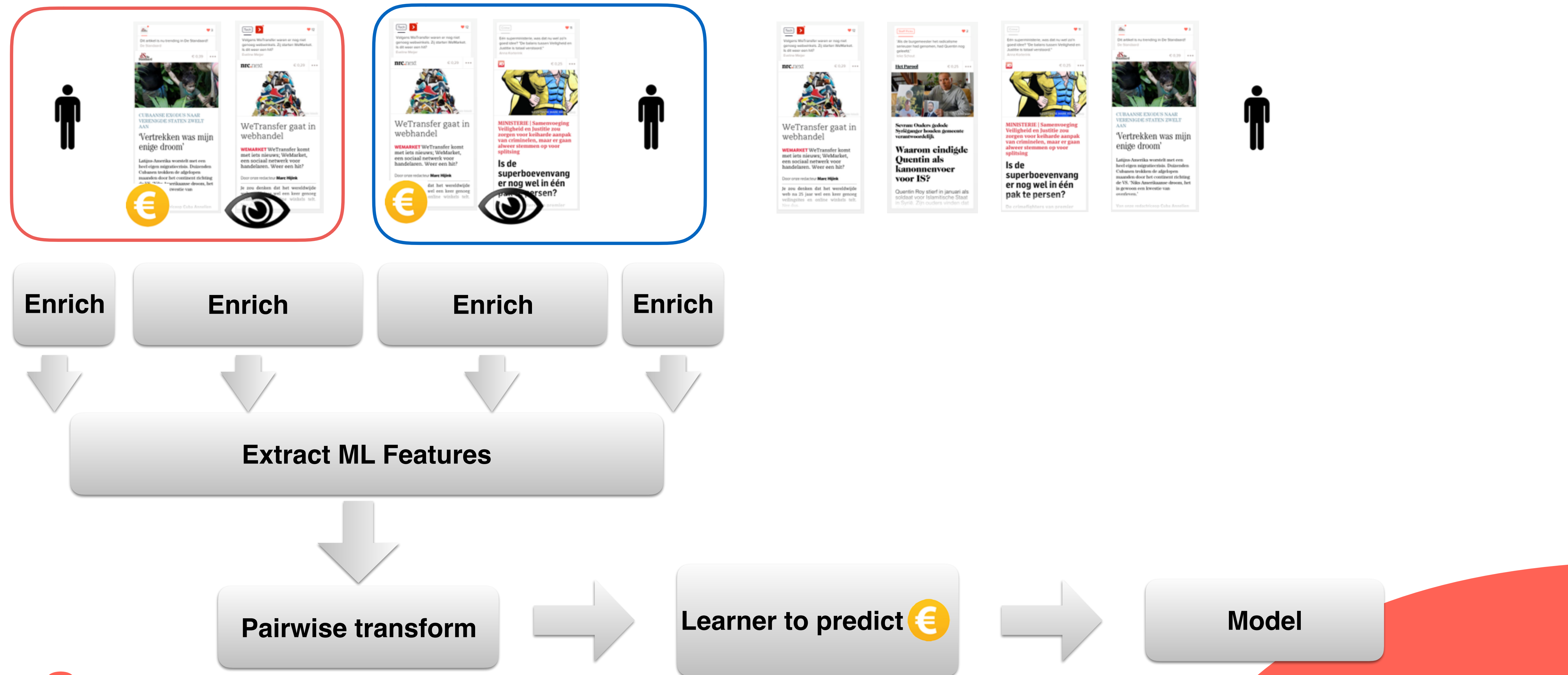
Learning to rank: preference learning



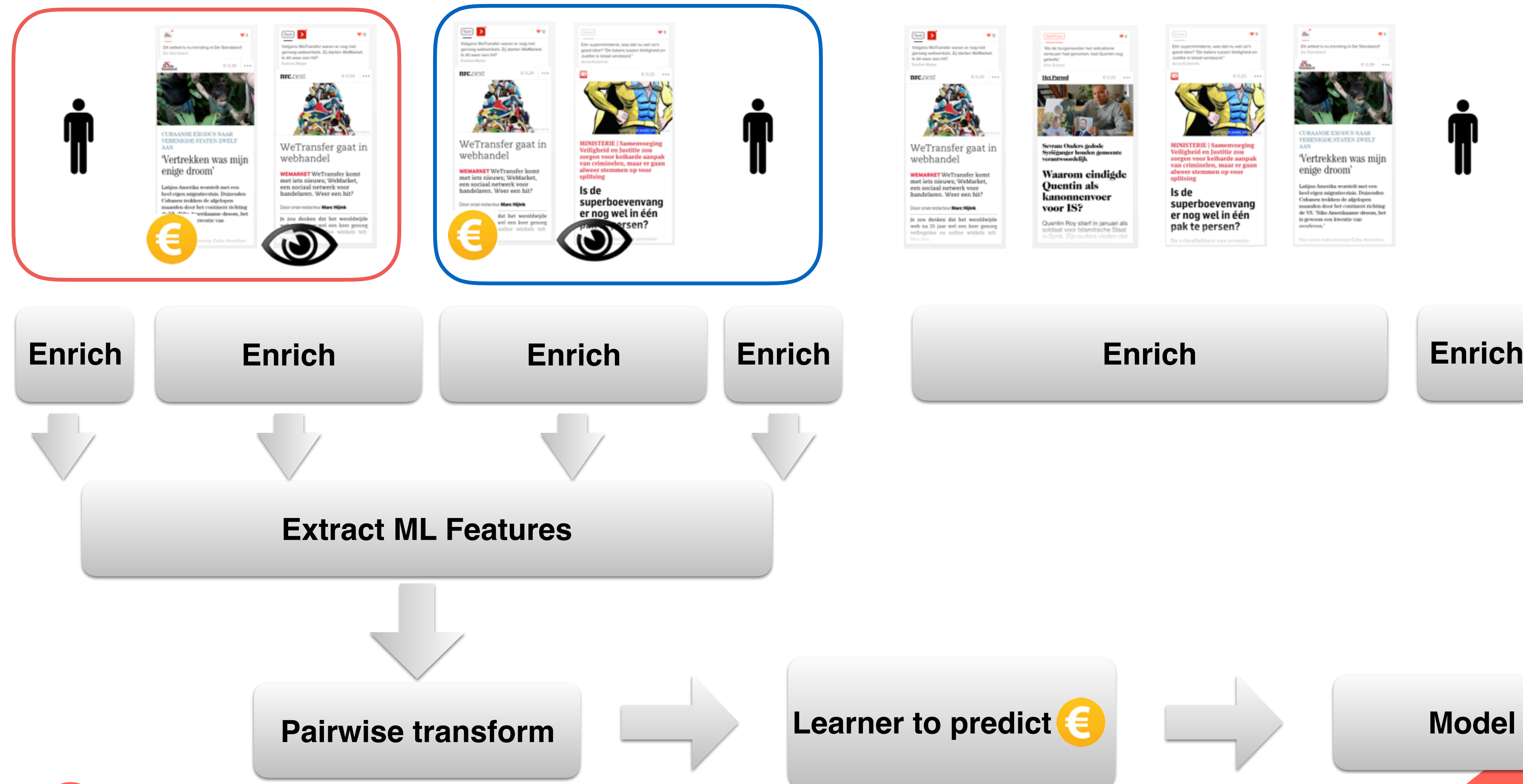
Learning to rank: preference learning



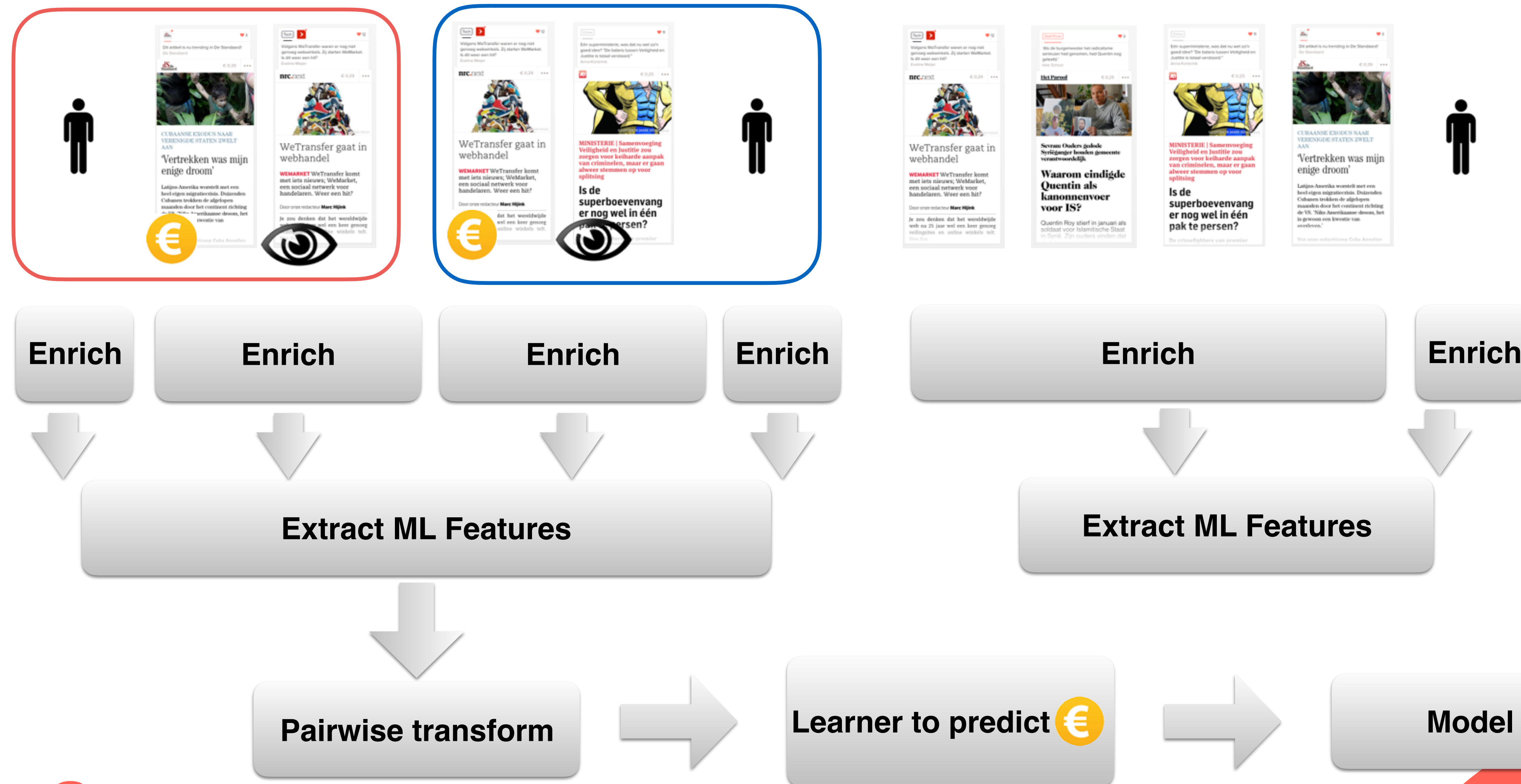
Learning to rank: preference learning



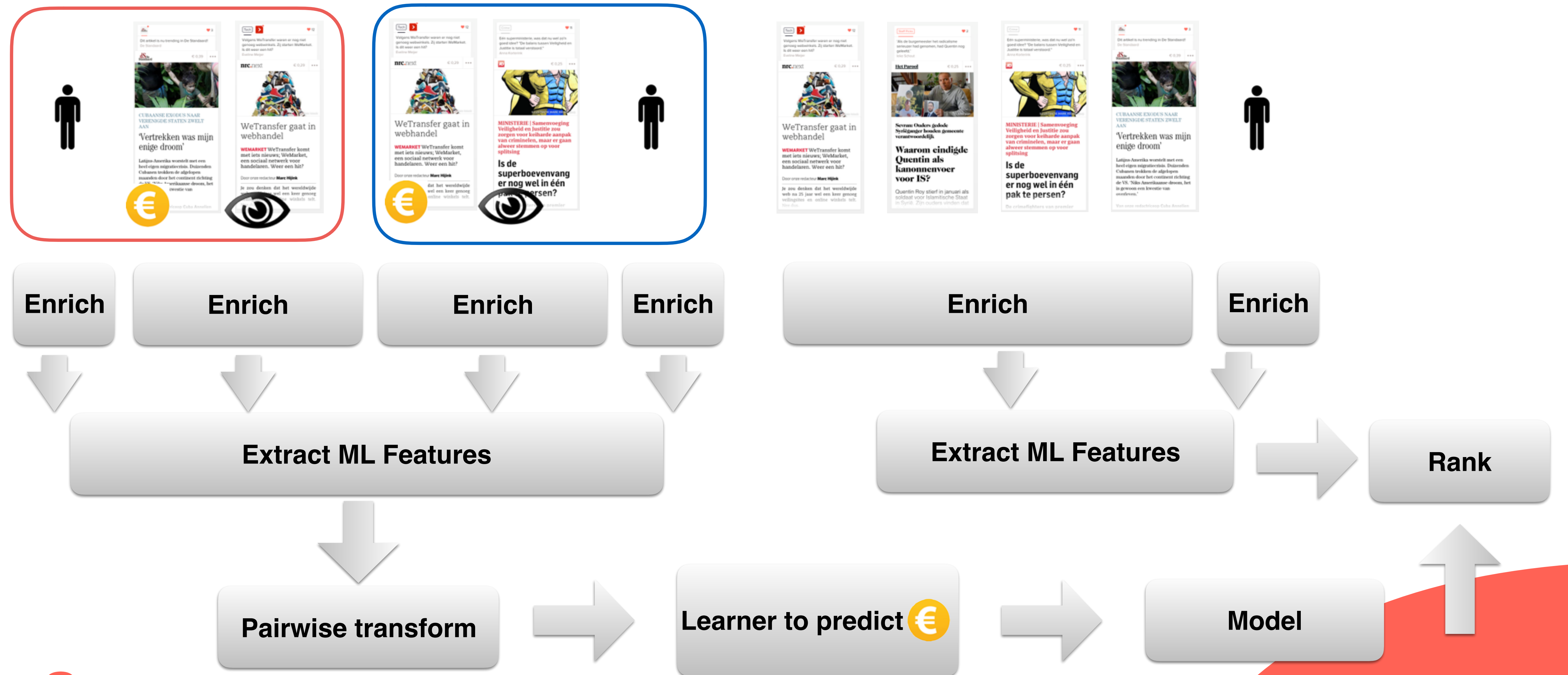
Learning to rank: preference learning



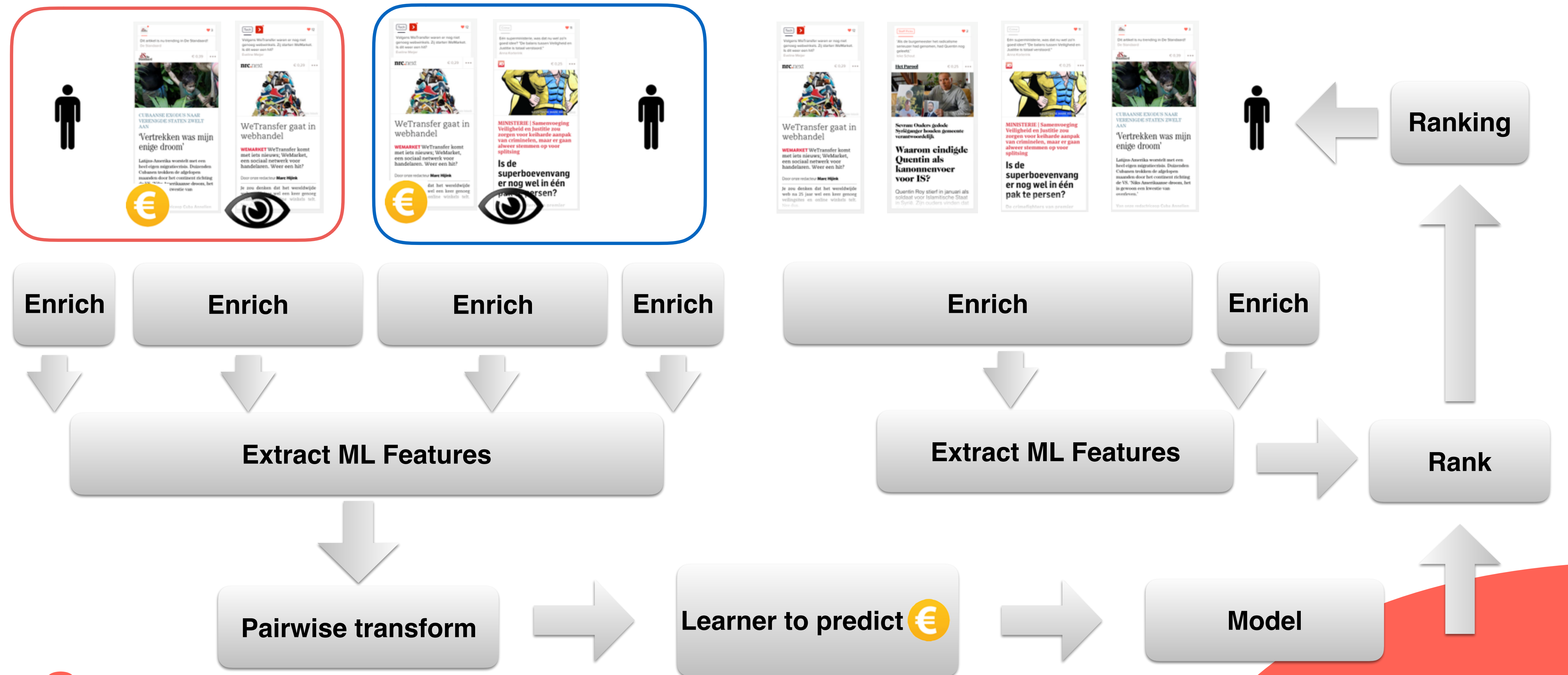
Learning to rank: preference learning



Learning to rank: preference learning



Learning to rank: preference learning



Next Steps

- First experiments are already running in the newsletter

Next Steps

- First experiments are already running in the newsletter
- Bring this to production all over Blendle

Next Steps

- First experiments are already running in the newsletter
- Bring this to production all over Blendle
 - Coming to “My Blendle” very soon

Next Steps

- First experiments are already running in the newsletter
- Bring this to production all over Blendle
 - Coming to “My Blendle” very soon
- Recommend subscriptions

Next Steps

- First experiments are already running in the newsletter
- Bring this to production all over Blendle
 - Coming to “My Blendle” very soon
- Recommend subscriptions
- Bring in social signals

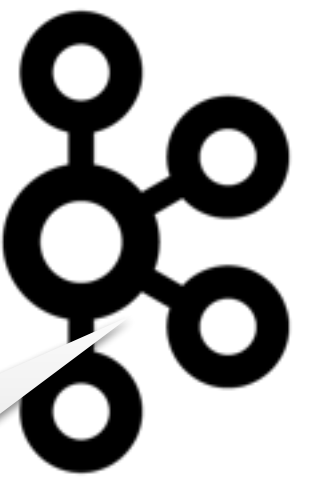
Next Steps

- First experiments are already running in the newsletter
- Bring this to production all over Blendle
 - Coming to “My Blendle” very soon
- Recommend subscriptions
- Bring in social signals

Components we use for Machine Learning



Components we use for Machine Learning



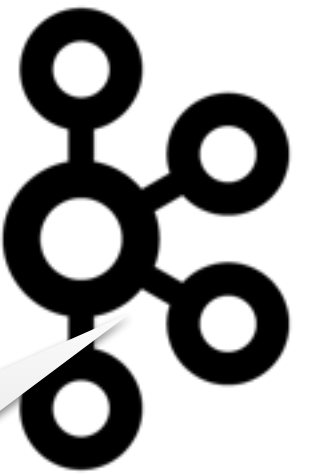
Kafka
- *events*



Components we use for Machine Learning



*Database
- data warehouse*



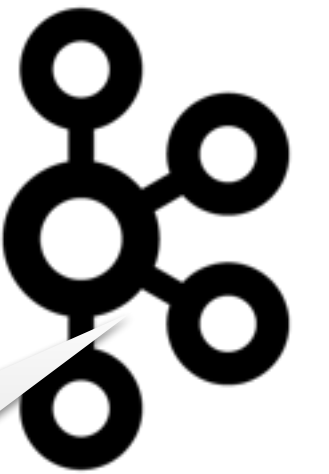
*Kafka
- events*



Components we use for Machine Learning



*Database
- data warehouse*



*Kafka
- events*



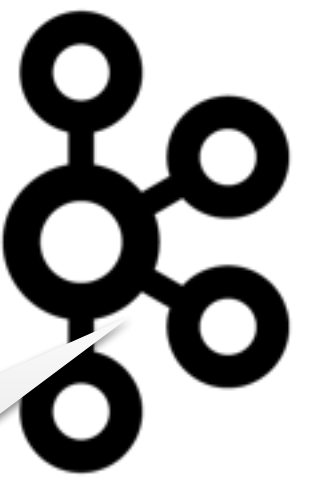
*Spark
- processing large data*



Components we use for Machine Learning



*Database
- data warehouse*



*Kafka
- events*



cucumber

*Spark
- processing large data*



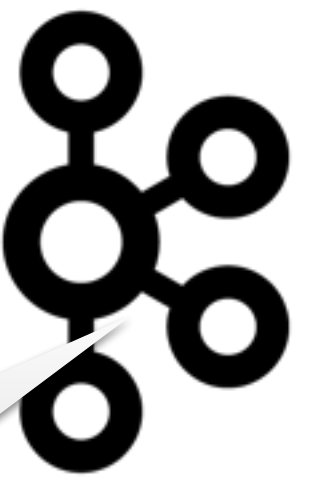
*Elastic Search
- fast search and access*



Components we use for Machine Learning



*Database
- data warehouse*



*Kafka
- events*



cucumber

*Spark
- processing large data*

*Scikitlearn
- experimental ML*



*Elastic Search
- fast search and access*



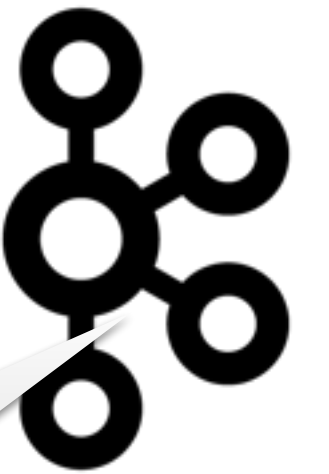
Components we use for Machine Learning



*Database
- data warehouse*



*Kafka
- events*



cucumber

*Spark
- processing large data*

*Scikitlearn
- experimental ML*



*Elastic Search
- fast search and access*



*MLib
- production ML*



Components we use for Machine Learning

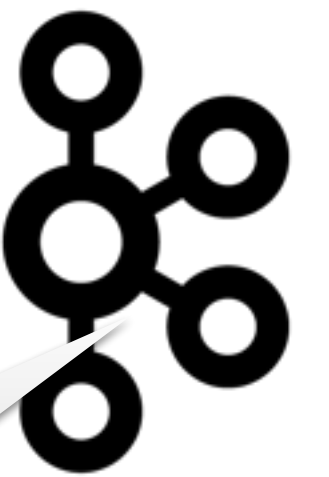


*Database
- data warehouse*



redis

*Redis
- fast access*



*Kafka
- events*



cucumber

*Spark
- processing large data*

*Scikitlearn
- experimental ML*



elastic

*Elastic Search
- fast search and access*



*MLib
- production ML*

Components we use for Machine Learning

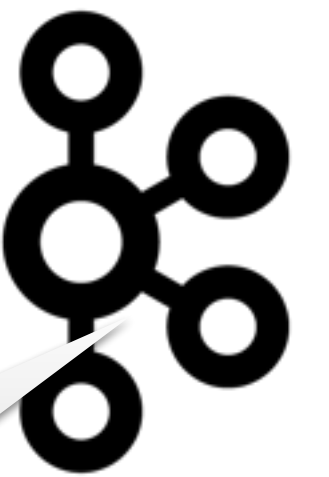


*Database
- data warehouse*



redis

*Redis
- fast access*



*Kafka
- events*



*S3
- raw content*

*MLib
- production ML*

cucumber

*Spark
- processing large data*

*Scikitlearn
- experimental ML*



*Elastic Search
- fast search and access*



Components we use for Machine Learning

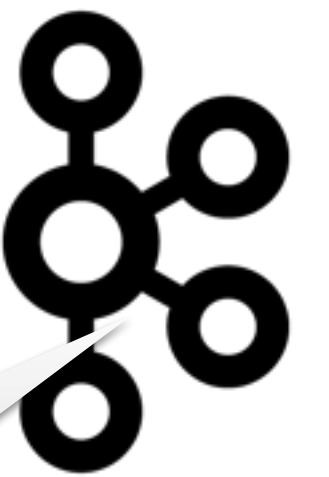


*Database
- data warehouse*



redis

*Redis
- fast access*



*Kafka
- events*



*Cucumber
- BDD development*

cucumber



*Spark
- processing large data*



*S3
- raw content*

*MLib
- production ML*

*Scikitlearn
- experimental ML*



*Elastic Search
- fast search and access*



Thank You