

“Du är vad du läser”

Tidningen som lär sig vad du gillar

Hur maskininlärning och sök hjälper oss att skapa smartare digitala produkter

Exempel du känner igen från dn.se och di.se

Några resultat vi lärt oss av

Också en del prototyper, inte alltid förverkligade på sajt

Vi har en stund för frågor efter dragningen

Alla deltagare får bilderna



“En jordbävning i långsamt tempo”

— — —

Ett medielandskap i förändring

Papper är dyrt, men också långsamt att distribuera

Konkurrerar om uppmärksamhet, tid och annonser med internets jättar

Mindre ytor, andra format, nya plattformar

Digitala produkter är “något annat”

Organisationer som bygger om sig för att publicera *hela tiden*, inte en gång om dagen

“Läste du tidningen i morse?” -> vilken tidning då? Inte samma för alla

Tidningen ändras i takt med att nytt material publiceras

Tidningen ändras i takt med att du läser

Du *deltar aktivt* i att bygga dina nyheter

Förväntningarna ställs högre idag

— — —

“Alexa, spela något från den där spellistan på Spotify, som jag gillar så mycket!”

En bra digital produkt tolkar ditt språk, gör antaganden om vad som kommer närmast, drar slutsatser

Vi tycker om att få hjälp med tips om resor, medicinska råd, val av utbildning, shopping, navigering, hitta kunskap och underhållning

Nyheter är annorlunda, eller är de det?

— — —

Nyheter ska vara “det som är viktigt” att veta just nu

Men de innehåller också material som man själv väljer och vill leta rätt på (kultur, sport, ekonomi)

Vi har tidningar *också* för att förkovra oss, bli underhållna, söka information

En del av kundlöftet är att skriva om det du gillar och ta reda vad det är

Beteendedata och textanalys

— — —

Vilka mönster kan vi se i användares vandring genom sajten, som talar om vad de gillar?

Collaborative filtering -> Oftast metoder som är “agnostiska” till innehållet

Eller se på “metadata”, markörer på intresse som är mer långlivade än läsningen av en viss artikel

Saker vi känner igen från DN och Di

Rekommenderat för dig



Expert om maktkuppen: "Nu äger Xi allt"



Följ måndagens börsöppning



Gammal fritureolja kan bli nästa flygbränsle

INNEHÅLL FRÅN RESIA FÖRETAG



Anders Hågerstrand: Geelys brist på respekt är en stor risk

"Andra läser just nu" (IBCF)

"Utvalt för dig" Hybrid mellan metadata och beteende

"Rekommenderat för dig",
Innehållsbaserat med diversifiering

Samtliga är *individuella* urval av artiklar.
Du är ensam i din målgrupp.

Andra läser.



Nyheter om Sverige får



NRA slår tillbaka mot



Totalstopp på E4 norr om

Utvalt för dig

Vad är det här?



Sport
Johan Esk: Med bättre skidor hade OS-succén blivit ännu större



Sport
Sagnia tangerade Klüfts rekord: "Herregud"



Världen
Hundratals svenskar strandade på Teneriffa



Sport
OS-minnena vi aldrig glömmer



Kultu
I be
alla

Hur vet man att det är bra?

Vi gör kontrollerade experiment online där vi tar med en viss procent användare

Mest läst är ofta en bra baslinje, det bästa generella urvalet som kunnat ske

Personalisering artikelsida, +13% fler användare som klickar (beteendedata)

Personalisering förstasidan, +4% fler användare (hybrid)

Personalisering på metadata, +7%

Men det är inte säkert att klick på yta är den bästa metriken, för vad man vill åstadkomma

Totalt antal sidvisningar, sidvisningar per besök, bounce rate, lästid är andra viktiga mått

Språktjänster och metadata

Ledare

Ledare: Skattekverulanter vinner inga poäng



Foto: Erik Ardelius

UPPDATERAD 2017-11-22 PUBLICERAD 2017-11-06

DN 7/11 2017. Att betala skatt är faktiskt inte sextigt. Men Leif Östling, veteran i industrin och ordförande för Svenskt Näringsliv, borde ha vetat bättre än att häva ur sig "Vad fan får jag för pengarna?" när SVT i helgen ställde frågor om hans knep för att undvika skatt.

```
1 pd.DataFrame(tag_suggestions, columns=['categories', 'score', 'doc_count'])
```

	categories	score	doc_count
0	Skatter	20.347203	361
1	Skatter i Sverige	4.962850	182
2	Pension	4.600227	79
3	Skatterätt	2.984135	96
4	Svenska allmänna pensionssystemet	2.854506	104
5	Svensk tjänstepension	2.768436	70
6	Privatekonomi	1.551057	114
7	Försäkringsrätt	1.488115	91
8	Arbetsmarknad	1.270030	163
9	Lön	1.223855	50

**Skatter, Pension,
Skatterätt**

```
1 pd.DataFrame(entities, columns=['entities', 'score', 'doc_count'])
```

	entities	score	doc_count
0	Magdalena Andersson (landshövding)	17.889530	157
1	Svenskt Näringsliv	17.268648	303
2	Leif Östling	9.973324	91
3	Carola Lemne	4.949739	45
4	Landsorganisationen i Sverige	3.136994	163

**Finansministern, Svenskt
Näringsliv, Leif Östling**

```
1 pd.DataFrame(conceptName, columns=['sector', 'score', 'doc_count'])
```

	sector	score	doc_count
0	labour	6.472052	1018
1	politics	0.498033	516
2	society	0.046916	311
3	economy, business and finance	0.014728	261

**Arbetsmarknad, Samhälle,
Politik**

```
1 pd.DataFrame(topicnames, columns=['topic', 'score', 'doc_count'])
```

	topic	score	doc_count
0	retirement	12.304516	591
1	employment	12.079739	769
2	government policy	2.631085	380
3	labour market	1.141238	230
4	labour relations	0.939793	207

Named entity recognition

Maskininlärning för att känna igen
benämningar av personer, organisationer
och platser i text

För varje ord (token), vad är
sannolikheten att det är början (B), på
insidan (I) eller utanför (O) en
benämning?

Prefix, suffix, ordklasser -> En mycket
stor mängd features

Entity linking

Det finns många “Michael Jackson” att välja på i den verkliga världen

Vi gör en sökning mot Wikipedia, lägger till ämneslikhet, popularitet, förekomst av ankartexter

En rankingmodell (LambdaMART) sorterar resultaten, en annan modell kollar kvaliteten

Ut kommer WikidataID, universella identifierare

Koppla till en kunskapsdatabas

Om jag vet att personen som nämns, är riksdagsledamot, så “handlar artikeln också om” dennes parti

Vi gör inferenser, drar slutsatser, om det som skrivits

Riksdagens öppna data, Premier League, NHL, Allsvenskan, från bolag till bransch

Branschroll

Di:s branschspecifika nyhetsbrev täcker Sveriges tyngsta bolag och verksamhetsområden. Varje vecka får du en summering av det viktigaste som har hänt i din bransch – underlaget du behöver för att ta välinformerade och effektiva affärsbeslut.



FÖR PRENUMERANTER

Bank & Finans

Branschroll Bank & Finans

Måndag, morgon

Nyheter inom ämnet Bank och Finans. Missa inget av Di:s rapportering av storbankerna, konjunkturen och den globala ekonomin.

Anmäl dig



FÖR PRENUMERANTER

Industri

Branschroll Industri

Måndag, morgon

Nyheter inom ämnet Industri och bolag såsom Sandvik, SKF, Volvo, Assa Abloy och Atlas Copco samlade i ett utskick.

Anmäl dig



FÖR PRENUMERANTER

IT & Telekom

Branschroll IT & Telekom

Måndag, morgon

Nyheter inom ämnet IT och Telekom. Få alla nyheter om den stenhårda och konkurrenstylda sektorn i ett utskick.

Anmäl dig



FÖR PRENUMERANTER

Energi

Branschroll Energi

Måndag, morgon

Nyheter inom ämnet Energi – allt om vindkraft, Sveriges kärnkraftverk, elpriser och dramatiken på oljemarknaden.

Anmäl dig



FÖR PRENUMERANTER

Råvaror

Branschroll Råvaror

Måndag, morgon

Nyheter inom ämnet Råvaror – det vi skriver om guld, olja och naturresurser och bolag inom sektorn samlade i ett utskick.

Anmäl dig



FÖR PRENUMERANTER

Fastighet

Branschroll Fastighet

Måndag, morgon

Nyheter inom ämnet Fastighet. Få allt om den överhettade bomarknaden och byggbolagen som släss om varje kvadratmeter.

Anmäl dig



FÖR PRENUMERANTER

Konsumentvaror

Branschroll Konsumentvaror

Måndag, morgon

Nyheter inom ämnet Konsumentvaror – allt om sektorn där börsjättar som SCA, Swedish Match och Electrolux figurerar.

Anmäl dig



FÖR PRENUMERANTER

Konsumenttjänster

Branschroll Konsumenttjänster

Måndag, morgon

Nyheter inom ämnet Konsumenttjänster. Från flygbolag till spelsajter – allt om bolagen som passar upp på konsumenterna.

Anmäl dig

Klassa till en taxonomi

IPTC har en väl utformad taxonomi över typer av nyheter

Kategorierna är det vi kanske skulle hittat som sektioner i en tidning (politik, konflikter, kultur etc.)

Vi tar definitionen av varje kategori och gör en semantisk beräkning med hjälp av ordvektorer (word2vec)

Kategorierna rangordnas efter hur stor andel av nyckelorden texten innehåller

Namngivna ämnestaggar

Wikipedias kategorier, efter litet
kurering

38000 ämnesord som kan tilldelas
texten som analyseras

Mycket bra kategorier, fria från
värderingar, överraskande träffande

Ex: “Sportevenemang i Norge”, “Rysk
kvinnohistoria”, “Politiska affärer”,
“Protestsånger”

“Jag är det jag läser”

— — —

Den följd av artiklar jag läst, kan ses
som en samling av alla metadatapunkter

Användarens preferenser, uttrycks på
samma sätt som ett dokument, en
artikel

De “finns i samma värld” och “talar
samma språk”

En *fiktiv* användarprofil

— — —

Engelskspråkiga filmer, Filmer i regi av Ingmar Bergman, Brittiska animerade filmer, HBTQ-relaterad film, Amerikanska animerade filmer, Europeisk film, Franska filmer baserade på böcker, Amerikanska familjefilmer, Tyska komedifilmer, Filmer från Universal Studios, Svenska filmer baserade på romaner, Amerikanska manusförfattare, Svenska dramafilmer, Amerikanska äventyrsfilmer

Robert Pattinson, Anna Carlsten, Greta Gerwig, Ingmar Bergman, Magnus Krepper, Alicia Vikander, Liev Schreiber, Oscar II, Roman Coppola, Gaspard Bauhin, Isabell Huppert, Astrid Lindgren, Tom Stoppard, Yoko Ono, Kickstarter, Michael Fassbender, Buenos Aires, Melinda Kinnaman, Scarlett Johansson, Hugo Weaving

“Environment”, “Human Interest”, “Arts, culture and entertainment”, “Society”

“Accomplishment”, “People”, “Nature”, “Culture”, “Values”, “Mass media”, “Arts and entertainment”

En språkmässig bro

Mellan användare och innehåll,
användaren *är* en sökfråga

Mellan innehåll och användare, “ge mig
de 1000 som är mest intresserade”

Mellan sajter och innehåll. Jag har läst
det här på Dagens Industri -> Vad på
Dagens Nyheter kan vara intressant

Redaktionella taggar, saknar betydelse
utanför huset, men en tagg som är
förankrad i Wikipedia, behåller den

<https://www.elastic.co/blog/looking-at-content-recommendation-through-a-search-lens>

En innehålls- rekommendation

En läshistorik, från två olika tidningar

Vi matchar med en individuell
rekommendation av redaktionellt
material

Så relevant som möjligt

Min läshistorik

```
In [42]: 1 pd.DataFrame(titles, columns=['epi_id', 'title'])
```

	epi_id	title
0	dn.epi.1959769	Marianne Björklund: Goda nyheter för alla med ...
1	di.epi.1958518	Johan Wendel: Bostadsfrossan är ett Stockholms...
2	dn.epi.1885065	Längre tid att sälja en bostad i Stockholm

Mina rekommendationer

```
In [43]: 1 pd.DataFrame(recommendation, columns=['epi_id', 'title'])
```

	epi_id	title
0	dn.epi.1863923	Riksbanken: Räntehöjning i mitten av 2018
1	dn.epi.1921589	Riksbanken splittrad om räntehöjning
2	dn.epi.1737775	Tre i Riksbanken emot ökade stödköp
3	dn.epi.1728027	Riksbanken skjuter upp räntehöjningar
4	dn.epi.1845362	Ingves: Vi är granne med en elefant
5	dn.epi.1923431	Riksbankschef vill se lägre bostadspriser
6	dn.epi.1841158	Sitter Ingves kvar? Striden om vem som blir ri...
7	dn.epi.1841780	I dag väntas besked om riksbankschef
8	dn.epi.1836018	Tidigare chef vill banta Riksbankens ledning
9	dn.epi.1908641	Riksbanken lämnar räntan oförändrad

En rekommendation av nativeannonser

Jag har läst tre artiklar, från olika tidningar

Vi gör en individuell matchning med annonser

Prototyp: idag gör vi bara en kontextuell matchning

Min läshistorik

```
In [56]: 1 pd.DataFrame(titles, columns=['epi_id', 'title'])
```

	epi_id	title
0	dn.epi.1959769	Marianne Björklund: Goda nyheter för alla med ...
1	dn.epi.1934009	Maria Crofts: Tänk på att amortering är bra my...
2	dn.epi.1885065	Längre tid att sälja en bostad i Stockholm

Mina rekommendationer

```
In [57]: 1 pd.DataFrame(recommendation, columns=['epi_id', 'title'])
```

	epi_id	title
0	di.epi.1959484	Stigande räntor – en risk för börserna 2018
1	di.epi.1866791	Så får du bättre villkor på ditt lån utan banken
2	di.epi.1722308	Sista intrycket är avgörande för framgång
3	di.epi.1749385	Lendify tar in 200 miljoner kronor från profes...
4	di.epi.1743413	Globala trenden för att investera pengar – nu ...
5	di.epi.1874778	Investera i lån – så funkar det
6	di.epi.1865434	”Bankerna förstår inte hur pass allvarlig utma...
7	di.epi.1744659	Nya samarbeten mellan fintech och etablerade a...
8	di.epi.1739041	”Bankerna förstår inte hur pass allvarlig utma...
9	di.epi.1881834	Därför är bankdagen död

Personaliserat sök

— — —

Beroende på vad jag läst precis innan, blir mina sökresultat annorlunda

Tar hänsyn till intressen, begrepp och kontext, inte bara det sökord du skriver in

När jag läst om kärnvapen, blir min sökning på “Trump” annorlunda och resultaten något annat

```
In [168]: 1 # Jag gör en sökning på "Trump"
          2 pd.DataFrame(simple_search(es_metadata, "Trump"), columns=cols)
```

epi_id	title
0 di.epi.1788246	Här är Trumps favorituttryck
1 di.epi.1784427	Trump-o-metern: Trump har negativ påverkan på ...
2 di.epi.1827644	Bannon: Det är Trumps värsta snedsteg
3 di.epi.1837451	Trump: De kommer inte finnas kvar
4 di.epi.1791671	Trump-granskning sänker börsen
5 di.epi.1815948	Trump om Östersjön: "Vi är alltid redo"
6 di.epi.1790751	Trump-utredningen sväller
7 di.epi.1790097	Trump pressar senatorer om sjukvårdsreform
8 di.epi.1824202	Nytt bakslag för Trumps inreseförbud
9 di.epi.1801928	Trump: Tuffare tag mot Nordkorea

```
In [260]: 1 ch = ['di.epi.1836136', 'di.epi.1796038'] ## Jag har läst två artiklar om kärnvapen
          2 pd.DataFrame(fetch_by_ids(es_metadata, ch), columns=cols)
```

epi_id	title
0 di.epi.1836136	Kina manar till besinning om Nordkorea
1 di.epi.1796038	Analytiker: Nordkoreas robot kan nå USA

```
In [261]: 1 pd.DataFrame(personalized_search(es_metadata, "Trump", click_history=ch), columns=cols)
```

epi_id	title
0 di.epi.1801928	Trump: Tuffare tag mot Nordkorea
1 di.epi.1788698	Trump: Iran följer kärnavtal
2 di.epi.1804165	Trump: "Militära lösningar är nu på plats"
3 di.epi.1824208	Trump: Militära åtgärder mot Nordkorea inte ou...
4 di.epi.1802853	USA: Trump improviserade hotet mot Nordkorea
5 di.epi.1827944	Trump varnar Nordkorea för tuffare åtgärder
6 di.epi.1823207	Trump om Nordkorea: "Vi får se vad som händer"
7 di.epi.1822034	Trump erbjuder Japan och Sydkorea fler vapen
8 di.epi.1798608	Trumps sanktioner mot Ryssland är motvilliga
9 di.epi.1829826	Trump planerar Asienresa

Rekommendera användare till en artikel

— — —

När vi publicerat en ny artikel, har vi alla metadata inom några sekunder

Vi använder artikeln som sin egen målgruppsdefinition och hämtar valfritt antal användare

Smala ämnen bör ha mindre målgrupper, bredare kan ha större

Ingen aktiv bevakning behövs från användaren

```
In [156]: 1 pd.DataFrame(get_top_users_for_article(es_metadata, 'di.epi.1875832', 20)
          2              , columns=['splus_id', 'score'])
```

	splus_id	score
0		3.332739
1	id:n på	3.332739
2	användare	3.332739
3		3.332739
4		3.332739
5		3.332739
6		3.332739
7		3.332739
8		3.332739
9		2.645609
10		2.645609
11		2.645609
12		2.645609
13		2.579213
14		2.579213
15		2.579213
16		2.579213
17		2.578362
18		2.578362
19		2.578362

Concept search:

- => Sök efter telecom och få Ericsson B
- => Enklare att hitta våra teman och taggar

Automatiska temasidor (sektioner?):

- => Allt vi skrivit om "hållbarhet" eller "valet 2018"
- => Ingen handpåläggning, alltid uppdaterat

Användaranalys

- => Vad är speciellt med de som läst den här artikeln?

Cross device recs:

- => En "utvalt för dig" baserad på intresse för bolag/aktier

Personaliserad nativeplacering:

- => Steg 1: Kontextuell relevans
- => Steg 2: Individuell relevans - de mest intresserade

Företagsspecifik landningssida:

- => Sorterar nyheter efter vad som är relevant för just dem (kunder, branscher, de anställdas läsmönster)

Premiumrekommendation:

- => Betalande användare får det mest relevanta från Weekend och Digital

Vidareutveckla merläsningen:

- => "Läs mer" fokuserar på aktier eller taggar
- => Fiskpinnen blir fokuserad och ger mer klick

Riktade erbjudanden för konvertering:

- => En låst artikel genererar erbjudanden för specifika individer

Personalized search:

- => Sök som lär sig av vad du läst
- => Du har läst om banker
- => Din sökning rangordnar bank högre

Trimmat sök:

- => Vi lär oss vad som ger bäst relevans
- => Samlar in data på söknet, viktat resultat

Rekommendationer för delad trafik

- => Din klickhistorik på en annan sajt ger personliga rekommendationer hos oss
- => Inget cold-start problem

Innehållsanalys

- => Vad är utmärkande för den här artikeln som gett så många sidvisningar?

Användningsfall



**DAGENS
NYHETER.**



ola.gustafsson@bonniernews.se

www.linkedin.com/in/ogustafsson