

#dbc



LÆS BLANDT ANDET OM:

Bibliotekernes big data

Computeren foreslår emneord

Ny relevans-rankering på bibliotek.dk

INDHOLD

- 3 Bibliotekernes big data
- 6 Det dynamiske vidensystem
- 7 Filmstriben nu også på tablet og smartphone
- 7 Udviklingsplan og årsberetning for Danbib
- 8 Computeren foreslår emneord
- 10 bibliotek.dk gætter, hvad du leder efter
- 12 Bits&Bytes

LEDER

Demokratiske data

Data vokser eksplosivt, og maskinerne er menneskets bedste ven, når overblikket skal genetableres.

Vi ved, at dataindsamling, dataanalyse og statistik kan forbedre vores liv og gøre teknologien mere smart. Det kan effektivisere, spare både klimaet og pengepungen og gøre hverdagen nemmere og sjovere.

Så godt som alt, hvad vi foretager os, efterlader spor. Det er et dilemma, fordi den gode brugeroplevelse i stigende grad er afhængig af, at systemet ved, hvor vi befinder os, hvad vi er optaget af, og hvem vi er venner med. På den anden side er det i mange tilfælde ikke transparent, hvem der har adgang til disse data, og hvad de kan (mis)bruges til.

Bibliotekerne har en lang tradition for at anvende statistik om brugere og produkter til materialevalg og i formidlingen. Som demokratisk videns-, lærings- og kulturinstitution skal biblioteket være en offentlig arena til samtale og udveksling – også i fremtiden. Det gøres bedst ved at tilbyde attraktive og konkurrencedygtige rammer, både fysisk og digitalt, hvor denne udveksling kan foregå.

Nøgleordene er åbenhed og transparens: Borgernes data skal forblive borgernes, og brugerskabte data skal komme tilbage til fællesskabet i form af merværdi og kollektiv videnskabelse.

I dette nummer af #dbc sætter vi fokus på big data og data science i en bibliotekssammenhæng. Både på visionerne for, hvordan de kan komme brugerne til gode, og på, hvordan DBC bruger computeren til at hjælpe mennesket, der katalogiserer.



Mogens Brabrand Jensen

Administrerende direktør
DBC

Læs mere på dbc.dk
og følg os på de sociale medier



social.dbc.dk

Design: BGRAPHIC

Forsidefoto: Lone Rasmussen

Tryk: PrinfoHHK.dk

Oplag: 2.000 stk.

ISSN 2445-4796 (trykt udgave)
ISSN 2445-980X (online)

#dbc udkommer 4-6 gange årligt.

Redaktion i dette nummer:
Madeleine Saunte og
Camilla Bødker Thomsen

Ansvarshavende redaktør:
Mogens Brabrand Jensen

Marked og Produkter
DBC A/S
Tempovej 7-11
2750 Ballerup
Telefon: 44 86 77 77
dbc@dbc.dk



Bibliotekernes

BIG data

Med data science kan den digitale biblioteksoplevelse skræddersys til den enkelte bruger.

TEKST: MADELEINE SAUNTE / FOTO: LONE RASMUSSEN

Data produceres med en hast og en volumen, som skaber helt nye vilkår og stiller nye krav til bibliotekerne. Data science er den videnskab, med hvilken man trækker viden ud af data.

I DBC har et team af udviklere efter en periode med flere mindre eksperimenter nu for alvor taget fat på arbejdet med

en platform, der med data science kan berige brugeroplevelsen.

Systemet lærer brugeren at kende

For slutbrugerne er det første synlige skridt på vejen webservicen Melvil, der stiller forslag til materialer (recommendations) og sorterer søgeresultater

(ranking) baseret på udlånsdata. Dette projekt vil kunne spille sammen med en mobilvenlig side, som DBC sideløbende udvikler for at kunne demonstrere webservicen i brug. Der er i første omgang tale om prototyper, som kan ses og afprøves på Danmarks Biblioteksforenings årsmøde i Aarhus den 16.-17. april.



I prototypen opbygger du først din egen personlige profil ved at vælge en række materialer til og fra. I takt med, at din profil opbygges, vil systemet kunne komme med forslag til andre materialer, der kan have din interesse. Det vil også kunne rankere dine søgeresultater ud fra profilen, så de materialer, du typisk interesserer dig for, står øverst i søgeresultatet. Systemet lærer dig altså at kende og tilpasser sig. Prototyperne synliggør mulighederne og skal bruges til at afklare, i hvilket omfang der skal arbejdes videre i dette spor.

Mange flere data science-projekter står nemlig på ønskelisten for biblioteksudvikling, og mulighederne er mange. Det glæder data science-teamet i DBC:

”For os som udviklere er det særligt motiverende at arbejde med bibliotekernes metadata, fordi de er så bundsolide, at man kan gøre mange ting med dem,” siger Søren Møllerup.

Visioner om et dynamisk vidensystem

Data science-projekterne er en del af en strategisk vision i DBC for en platform, som aktivt bruger dynamiske events til at skabe en bedre og mere målrettet brugeroplevelse. Dynamiske events kan være brugsadfærd, brugerfeedback, handlinger på sociale medier, professionelle anbefalinger osv.

”Bibliotekernes position som kulturformidlere udfordres af kommercielle aktører, og der skal gøres en indsats for at fastholde og udvikle bibliotekets rolle på oplevelses- og læringsmarkedet,” siger strategisk konsulent i DBC Bo Weymann.

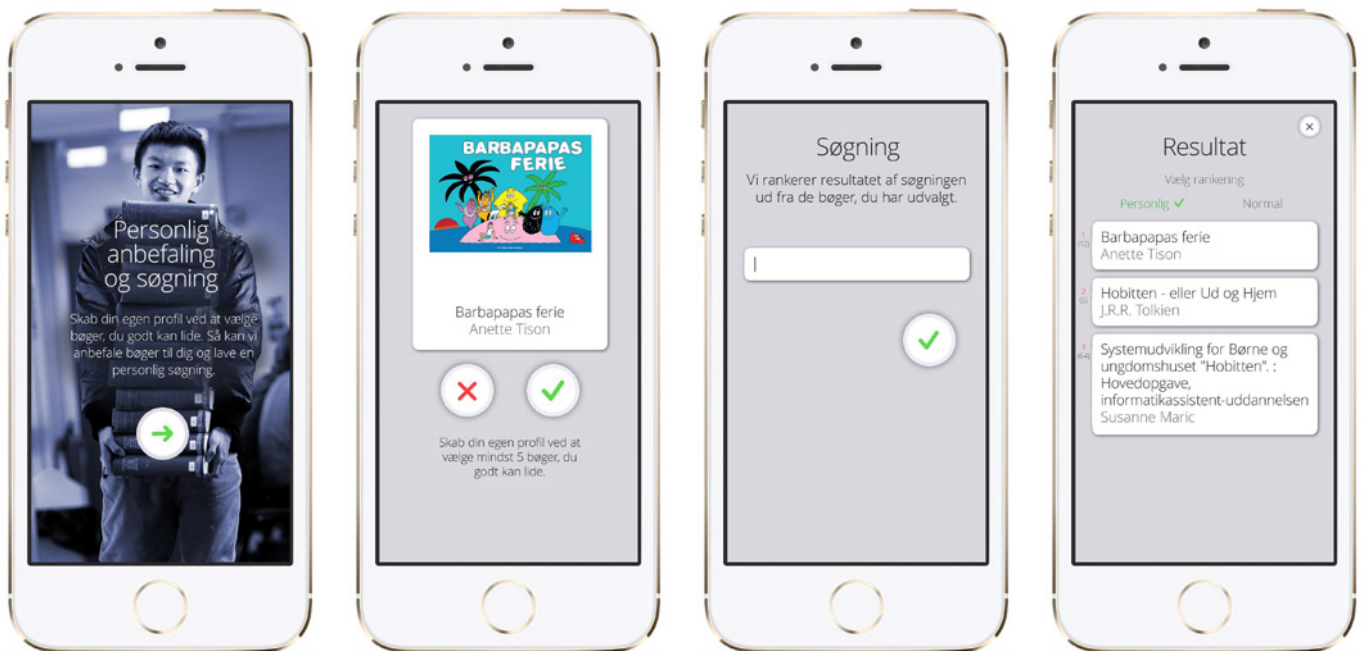
Formålet er at opsamle data, viden og trends og anvende denne viden til at gøre brugsoplevelsen bedre. Det kan være via forbedret filtrering i søgninger, forbedret ranking i svarsæt, og ved direkte at give anbefalinger. Derved bliver brugeroplevelsen mere målrettet

“ Det dynamiske vidensystem er den overordnede tanke, som vi arbejder os hen imod via mindre projekter. Vi opbygger viden og afprøver idéerne i praksis.

Strategisk konsulent Bo Weymann.



Prototypen Melvil kan prøves på melvil.demo.dbc.dk



Melvil. Prototypen Melvil stiller forslag til materialer og rankerer søgeresultaterne ud fra udlånsdata.



Data science-team. Jesper Damkjær, Adam Tulinius, Christian Boesgaard og Søren Møllerup arbejder på en webservice, som kan komme med anbefalinger og sortere søgeresultater baseret på brugerens udlånsdata.

og mere dynamisk. Samtidig ser Bo Weymann andre fordele i anvendelsen af et dynamisk system:

”Den opsamlede viden og data kan analyseres med henblik på serviceudvikling. Ved at vide mere om brugerne og deres adfærd kan man afdække behov og være på forkant. Derved får bibliotekerne et bedre beslutningsgrundlag for, hvad der skal tilpasses og videreudvikles.”

Eksisterende data i spil

Første skridt i den nye udvikling er at sætte allerede eksisterende data i spil, så de skaber værdi på nye måder.

”I første omgang vil vi eksperimentere med forbedring af digitale anbefalinger og rankinger af søgninger ved hjælp af indholdet i databrønden, webservicen Andre-Der-Har-Lånt og brugernes

adfærd på bibliotek.dk,” siger Bo Weymann.

Han er overbevist om, at det er denne vej, udviklingen går:

”Vi afgiver informationer om os selv hele tiden, fordi brugeroplevelsen bliver bedre af det, og fordi det er sådan, vi i dag socialiserer: Ved at vise omverdenen, hvad vi læser om, kigger på, lytter til, og hvad vi mener om dét.”

Brug af dynamiske events som en integreret del af oplevelsen i det digitale bibliotek skal skabe værdi for brugerne:

”Vi ser det som en nødvendig og spændende udvikling. Data science-projekterne i DBC handler både om at skabe værdi her og nu for brugerne, og om at opbygge viden og erfaring i dette felt,” afslutter han.

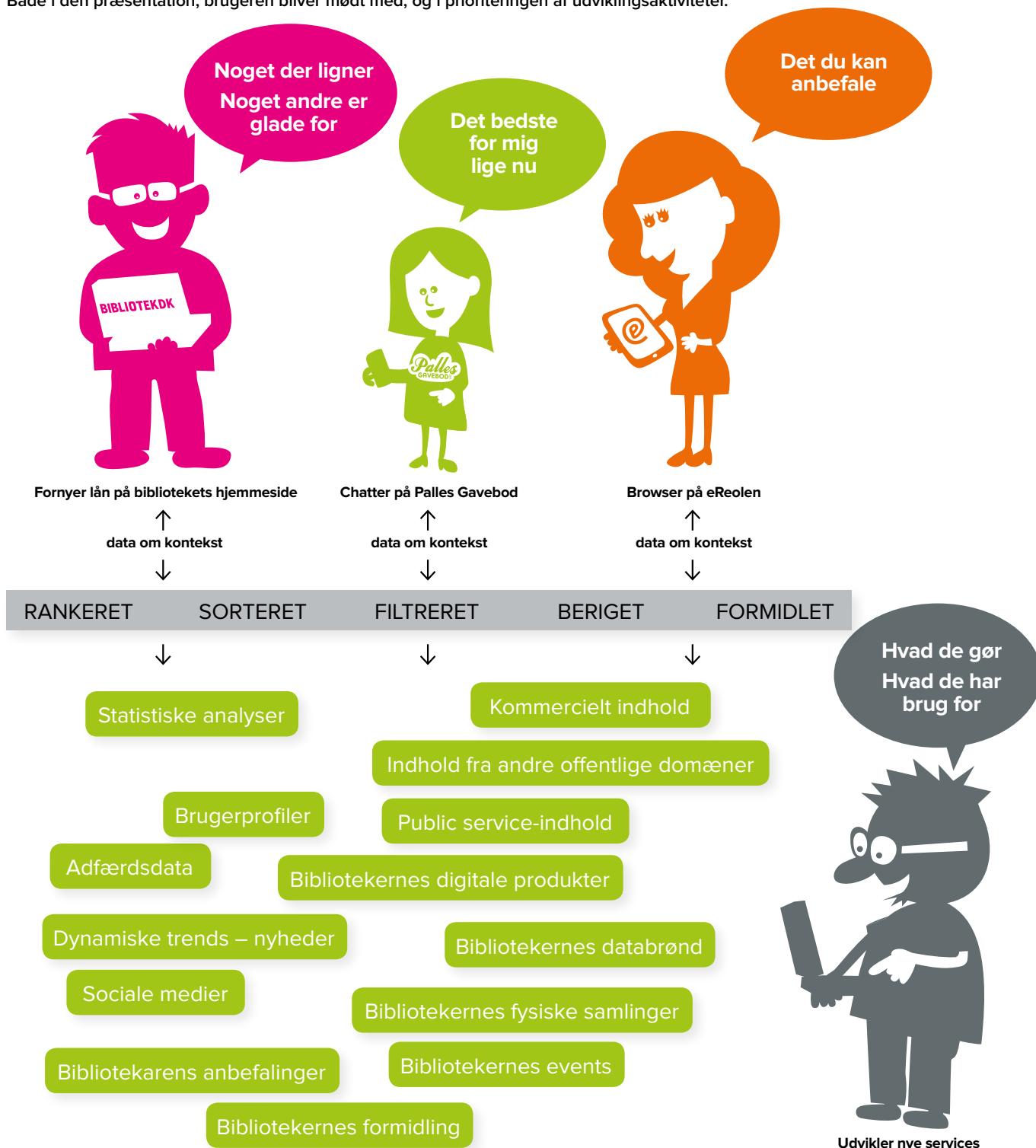


Store visioner og konkrete projekter. ”Vi ser store perspektiver i at anvende data science i en bibliotekssammenhæng,” siger strategisk konsulent Bo Weymann. ”Vi tager det et skridt ad gangen, i form af mindre projekter, som på sigt kan spille sammen i et dynamisk system.”

Det dynamiske vidensystem

Brugerens kontekst og behov er udgangspunktet i det dynamiske vidensystem.

Både i den præsentation, brugeren bliver mødt med, og i prioriteringen af udviklingsaktiviteter.



DBC-NYHED 09.03.2015

Filmstriben nu også på tablet og smartphone

Ny app til iPad og iPhone – og snart også til Android.

TEKST: CAMILLA BØDKER THOMSEN

Nu kan man se Filmstriben på iPad og iPhone – og snart også på Android. Det gælder både, når man vil se film hjemmefra og på biblioteket.

Den ny app, der gør det muligt at se Filmstribens film hjemmefra via tablet og smartphone, hedder 'Filmstriben – i lænestolen' og kan hentes gratis i App Store.

Den er stadig lidt "i beta", så hav tålmodighed i starten.

I forbindelse med filmarrangementer på biblioteket kan app'en 'Filmstriben – på biblioteket' bruges.

Den nye app til Filmstriben er blevet hentet 13.100 gange i løbet af marts.



DBC-NYHED 18.03.2015

Udviklingsplan og årsberetning for Danbib

Udviklingen af Danbib fokuserer også i 2015 på samspillet mellem den nationale og den lokale biblioteksinfrastruktur.

TEKST: CAMILLA BØDKER THOMSEN

Danbib er det sidste store system-kompleks, som DBC er i gang med at migrere til open source og service-orienteret arkitektur.

Flere af komponenterne blev udviklet i 2014, og i 2015 er turen blandt andet kommet til Danbib-basen og grænsefladen Netpunkt.

En række af de services, der er udviklet i 2014, vil i første omgang blive nyttiggjort i det fælles bibliotekssystem, Cicero. Udviklingen i 2015 vil fortsat have fokus på samspillet mellem den nationale og den lokale biblioteksinfrastruktur.

Læs mere i 'Danbib udviklingsplan 2015' og 'Danbib årsberetning 2014', som du finder på dbc.dk under Nationale opgaver > Danbib.

DBC udvikler, vedligeholder og driver Danbib på opdrag af Kulturministeriet, Kommunernes Landsforening og Københavns og Frederiksberg kommuner i henhold til Danbib-udviklings- og driftsaftalen.

De statslige og kommunale repræsentanter i DBC's bestyrelse skal være enige om alle væsentlige Danbib-dispositioner.

DBC-NYHEDER PÅ E-MAIL

På dbc.dk/news kan du tilmelde dig DBC-nyheder og få dem tilsendt på e-mail.

BIBLIOGRAFISK DATA SCIENCE:

Computeren foreslår emneord

Assisteret indeksering hjælper medarbejderen og giver mere konsistente emneord.

TEKST: CAMILLA BØDKER THOMSEN / FOTO: POUL BERGSTRØM HANSEN

”Det er en grundlæggende ændring i vores måde at arbejde på.”

Sådan siger Simon Ørregaard, der er teamleder for de medarbejdere i DBC, der registrerer artikler til Nationalbibliografien.

Ændringen går under betegnelsen ’assisteret indeksering’, der dækker over, at medarbejdernes daglige arbejdsredskab bruger data science til at hjælpe i katalogiseringsprocessen – ved at give medarbejderen automatiske forslag til emneord og klassemærker.

DBC har selv udviklet systemet, som er implementeret i en testversion i DBC’s katalogiseringsværktøj.

Foreløbig bruger to medarbejdere det nye system, når de katalogiserer artikler.

Stærkt vanedannende

”Det er stærkt vanedannende,” siger katalogisator Birgit Gøtterup med et smil. ”Jeg kommer i tanker om flere emneord end før og føler mig mere sikker på, at jeg ikke bare kommer til at gøre ’som jeg plejer’.”

Når Birgit og kollegaen Lasse skal katalogisere en artikel, hvor de har adgang til den fulde tekst, trykker de blot på en genvejstast, og frem på skærmen kommer en række forslag til emneord og DK5-klassemærke – med de mest relevante først. De kan nu let tilføje disse metadata til posten.

Eller de kan udforske andre muligheder, f.eks. ved at vælge et af de foreslåede emneord og så bede systemet om en oversigt over, hvilke andre emneord, der typisk bliver brugt sammen med dette.

Systemet kan også foreslå emneord ud fra konkrete klassemærker. Eller klassemærke på baggrund af emneord.

Røntgenblik i basen

At katalogisere er en ret omfattende, manuel proces, som trækker på den enkelte medarbejders overblik over emner og emneord. Med computerens hjælp kan medarbejderen få kvalificeret sit arbejde med forslag, som han eller hun måske ikke ellers ville være kommet i tanker om.

Mennesket og maskinen. ”Man sidder oven på en meget mere potent maskine end før, men det er stadig mennesket, der har kontrollen,” siger teamleder Simon Ørregaard.



DATA SCIENCE

Systemet regner selv en metode ud

Det system, som DBC tester på artikelområdet, bygger på en metode, hvor man ’fodrer’ et system med en række opgaver og disses løsninger, hvorudfra systemet selv generaliserer sig frem til en metode til at kunne løse nye opgaver af lignende slags.

Systemet er i dette tilfælde blevet forsynet med en omfattende mængde af tekster, som allerede – af mennesker – er blevet tildelt emneord og klassemærker. Det laver herudfra en algoritme, så det på baggrund af nye artikler, som det ikke tidligere har set, kan foreslå passende emneord og klassemærker.

”Det er som en slags røntgenblik ind i vores base,” siger Lasse Nordentoft, der sammen med Birgit Gøtterup har testet systemet siden januar.

”Når jeg før stod med en artikel, ville jeg normalt undersøge mellem ti og tredive lignende artikelposter for at tjekke, om jeg har alle relevante emneord med. Nu bliver hele basen gennemlyst på et øjeblik,” siger han.

Simon Ørregaard forklarer, at systemet også gør katalogiseringerne mere konsistente.

”I teamet laver vi med jævne mellemrum det, vi kalder et ’artikelfix’, hvor vi katalogiserer den samme artikel og så sammenligner, hvordan vi hver især har gjort. Her kan der være én, der foreslår emneord, som de andre ikke har tænkt på. Med det nye system kan man automatisere den proces og køre ’artikelfix’ op mod hele basen – ikke bare på et udvalg af artikler, men hver gang man katalogiserer.”

Mennesket har kontrollen

Nu kommer det måske til at lyde som om, DBC er på vej til at lade computeren helt overtage arbejdet med at katalogisere, men det er langt fra tilfældet, fortæller Simon Ørregaard:

”Det er kun fordi, der er mennesker, der har katalogiseret tusinder og atter tusinder af artikler, at computersystemet kan gøre det her. Hvis der ikke var mennesker, der havde sagt, at de her emneord passer til den slags artikler, kunne computeren ikke gøre det samme.”

Fordi systemet bygger på katalogiseringer, der allerede er foretaget, kommer det for eksempel også til kort, hvis der er behov for helt nye emneord. Her må mennesket på banen. Og det er i det hele taget stadig mennesket, der træffer beslutningerne – på baggrund af forslag fra computeren.

”Man kan sige, at man sidder oven på en meget mere potent maskine end før, men det er stadig mennesket, der har kontrollen,” siger Simon Ørregaard.

Det nye system kører som testversion i DBC’s interne katalogiseringsmodul og vil blive sat i egentlig drift senere på året. Til en start bliver det brugt ved katalogisering af artikler, men der vil formentlig også være perspektiv i at bruge dele af funktionaliteten til f.eks. fagbøger.

Arbejdet med computer-assisteret indeksering er en del af DBC’s løbende fokus på at optimere kvaliteten og produktionen af metadata.

ASSISTERET INDEKSERING

Sådan hjælper systemet katalogisatoren

Fuldtekst

Ud fra artikler i fuldtekst foreslår systemet emneord og DK5-klassemærker.

Emneord til emneord

Ud fra et (eller flere) emneord foreslår systemet andre, der plejer at optræde sammen.

Emneord til DK5

Ud fra emneord foreslår systemet DK5-klassemærker.

DK5 til emneord

Ud fra DK5-klassemærke foreslår systemet emneord.



De to testere. ”Har du prøvet også at se på de forslag, der står helt hernede?” Birgit Gøtterup og Lasse Nordentoft har siden januar brugt det nye system, når de katalogiserer artikler. Systemet kører i en testversion som en del af DBC’s katalogiseringsværktøj og bliver sat i egentlig drift senere på året.

RANKERING

bibliotek.dk gætter, hvad du leder efter

TEKST: CAMILLA BØDKER THOMSEN

Et par ord i søgefeltet. Et klik. Og frem kommer et søgeresultat, hvor det mest relevante står øverst. Det lyder såre simpelt men er det langt fra. For hvordan ved systemet, hvad du leder efter?

Nu har bibliotek.dk indført en ny måde at bestemme rækkefølgen af søgeresultaterne på. Kort fortalt er det nye, at bibliotek.dk først gætter på, om brugeren er i gang med at søge efter et emne, en titel eller en forfatter, og dernæst vælger en måde at rankere resultaterne på, som er god til netop dén type søgning.

Indeks giver fingerpeg

Funktionen, der gætter på, hvad brugeren søger efter, tager udgangspunkt i, hvilke indekser søgeordene optræder i. Helt forenklet er princippet, at hvis ordene optræder mange gange i forfatterindekset og kun få gange i

emneindekset, er det, der søges efter, nok en forfatter. I praksis er det dog nødvendigt at differentiere, så det f.eks. tæller mere, hvis et ord optræder i emneindekset end i forfatterindekset.

Tre algoritmer

Der er lavet tre forskellige rankeringsalgoritmer: én, der er egnet til titelsøgninger, én til forfattersøgninger og én til emnesøgninger.

Når rankeringsalgoritmen skal vælge, hvilke søgeresultater, der skal vises først, ser den blandt andet på, hvilke felter i posten, søgeordene optræder i. For eksempel giver Titel-algoritmen en højere placering til poster, hvor søgeordene optræder i feltet 'Hovedtitel' frem for i andre titelfelter. Og hvis søgeordene ligefrem står ved siden af hinanden i Hovedtitel-feltet, bliver posten endnu højere placeret.

Baggrunden for den ny rankering er dels en erkendelse af, at man med databrønden søger i betydeligt flere data end tidligere, og dels, at mange brugere udelukkende søger i fritekstfeltet.

Peter Pedal driller

I 80 % af tilfældene vælger bibliotek.dk nu automatisk den rigtige rankering. De søgninger, hvor det ikke er tilfældet, er typisk, når titelsøgninger består af navne – for eksempel, hvis man skriver 'Peter Pedal'. Heldigvis får man altid at vide, hvilken rankeringsform bibliotek.dk har valgt, så man kan vælge en anden, hvis det er nødvendigt.

Rankeringen vil løbende blive optimeret, blandt andet ved hjælp af input fra bibliotek.dk's brugerpanel.

Den ny rankering ligger i Search 4.0 og hedder rank_frequency.

1. Hvad søger brugeren efter?
(forfatter/emne/titel)



2. Hvilken rankerings-algoritme passer til?
(forfatter/emne/titel)

Der er to trin i den nye relevansrankering: Først gætter bibliotek.dk på, hvad man er i gang med at søge efter; dernæst vælger den en rankeringsalgoritme, der passer til netop den type søgning.

SÅDAN HAR VI GJORT

Både funktionen, der gætter, hvilken ranking, der er relevant, og selve rankingalgoritmerne er i høj grad blevet til ved at prøve sig frem, ændre en parameter og så prøve sig frem igen.

Gættefunktionen er lavet sådan

1. Vi laver en række fritekstsøgninger efter henholdsvis forfattere, emner og titler.
2. For hvert søgeresultat får vi en optælling af, hvor mange gange søgeordene optræder i henholdsvis forfatter-, emne- og titelindeks.
3. Vi ser, om antal forekomster i indekser alene kan fortælle, hvilken type søgning, der er tale om.
4. Det kan den ikke.
5. Vi eksperimenterer med, at det ikke skal tælle det samme, om ord står i det ene eller det andet indeks – f.eks. skal det tælle mere, hvis et ord optræder i titelindekset, og endnu mere, hvis det står i emneindekset.
6. Vi fortsætter med at eksperimenterer, indtil funktionen gætter rigtigt for flest testsøgninger.

Rankeringsfunktionen er lavet sådan

1. Vi laver en række fritekstsøgninger efter henholdsvis forfattere, emner og titler.
2. Et antal slutbrugere samt DBC-medarbejdere ser på søgeresultaterne og vurderer posternes relevans med point fra 0 til 9.
3. Vi eksperimenterer med sammensætning af rankingalgoritmen og gentager på den baggrund testsøgningerne og vurderer de nye søgeresultater – f.eks. skruer vi på, hvor meget det tæller, at ord står i bestemte felter eller i en bestemt rækkefølge.
4. Vi bliver ved med at skrue på algoritmen og vurdere relevansen af søgeresultaterne, indtil vi har flest muligt højt ratede poster øverst i søgeresultatet.

Alle materialer

inklusion

Søg

Er det mon et forfatternavn, en titel eller et emne, brugeren er i gang med at skrive i fritekstfeltet?

Nu gætter bibliotek.dk – og bruger sit gæt til at vælge, hvilken metode der skal bestemme rækkefølgen af søgeresultaterne.

Bits&Bytes

71.219

kontrollerede emneord er der i DBC's emneordsbase.

Danmark

er det mest anvendte emneord af dem alle i DBC's registreringer af bøger, musikoptagelser, film, artikler mm.

130.024

gange er 'Danmark' brugt. Til sammenligning er 'politik' brugt 20.424 gange og 'kærlighed' 18.324 gange.

MEST SETE FILM PÅ FILMSTRIBEN

Første kvartal 2015



Klumpfisken

Henrik Birch fik i år både Bodil- og Robert-pris for Bedste mandlige hovedrolle i 'Klumpfisken'.



12 years a slave

Ved den 86. Oscar-uddeling i 2014 fik '12 years a slave' prisen for Bedste film.



Otto er et næsehorn

I 2013 fik 'Otto er et næsehorn' Svend-prisen for årets danske børnefilm.

Billeder: Pressefotos

DBC i 2014 – Årsberetning

I 2014 fik bibliotekerne for alvor glæde af DBC's omfattende investeringer i migrering til serviceorienteret arkitektur og rationalisering af metadataproduktionen.

I DBC's årsberetning kan du læse mere om vores udvikling inden for det bibliografiske område, it og produkter og studere vores økonomiske nøgletal.

Læs mere på dbc.dk under Om DBC > Publikationer

DBC'S
ÅRSBERETNING

20
14



DBC