

20 års biblioteks- historie gennem REVY som prisme

□ **Max Odsbjerg Pedersen**, Informations-specialist, Det Kgl. Bibliotek, AUL Arts
Susanne Dalsgaard Kragh, Biblioteksleder, Det Kgl. Bibliotek, AU Library, ARTS / Aarhus BSS

Gennem textmining af 20 års digitale udgaver af REVY har vi udtrukket de mest signifikante ord fra hvert enkelt år og på den måde lavet en række nedslag i de seneste 20 års forskningsbibliotekshistorie. Visualiseringen på næste side er baggrund for denne artikel.

Helt konkret er ordene ud for hvert enkelt år de ord, som er særligt definerende for det enkelte år. "Særligt definerende" er i denne sammenhæng afgjort af ordenes *term frequency – inverted dokument frequency (tf-idf)*, som udregner et ords frekvens inden en række dokumenter, men samtidig "straffer" ord, som forekommer i alle årene ("den", "det", "der", osv.).

Disse særlige ord vil sikkert for mange give mening, men for størstedelen vil ordenes fremkomst inden for de enkelte år vække undren og ønske om kontekst – og det er netop hensigten i det følgende.

Max Odsbjerg Pedersen har stået for textminingen og Susanne Dalsgaard Kragh har stået for kontekstualisering. Hun er den helt rette til dette, da hun har været i branchen i hele den periode, som vores undersøgelse dækker. Hun startede sin karriere på Handelshøjskolen Bibliotek i Aarhus, hvorefter hun i 2002 blev leder af Det Teologiske Fakultetsbibliotek

på Aarhus Universitet. I 2009 stod hun for sammenlægningen af fire institutbiblioteker på Det Samfundsvidenskabelige Fakultet, og i begyndelsen af 2010'erne deltog hun i arbejdet med dannelsen af AU Library. De seneste år har hun været afdelingsleder på Det Kgl. Bibliotek med ansvar for bibliotekerne på Arts og Aarhus BSS.

Den 20-årige periode, vi her har gennemgået, har budt på sammenlægninger og fusioner, der har foregået på universitetsniveau, lokalt universitetsbiblioteksniveau såvel som nationalt biblioteksniveau. På baggrund af Susanne Dalsgaard Kraghs cv så bliver udvælgelsen af ord i denne artikel uvægerligt et århusiansk blik på de seneste tyve års Dansk Fag- Forsknings- og Uddannelsesbiblioteks-historie. Vi opfordrer alle, der har lyst til at give et andet geografisk blik på bibliotekshistorien at lave deres egne nedslag.

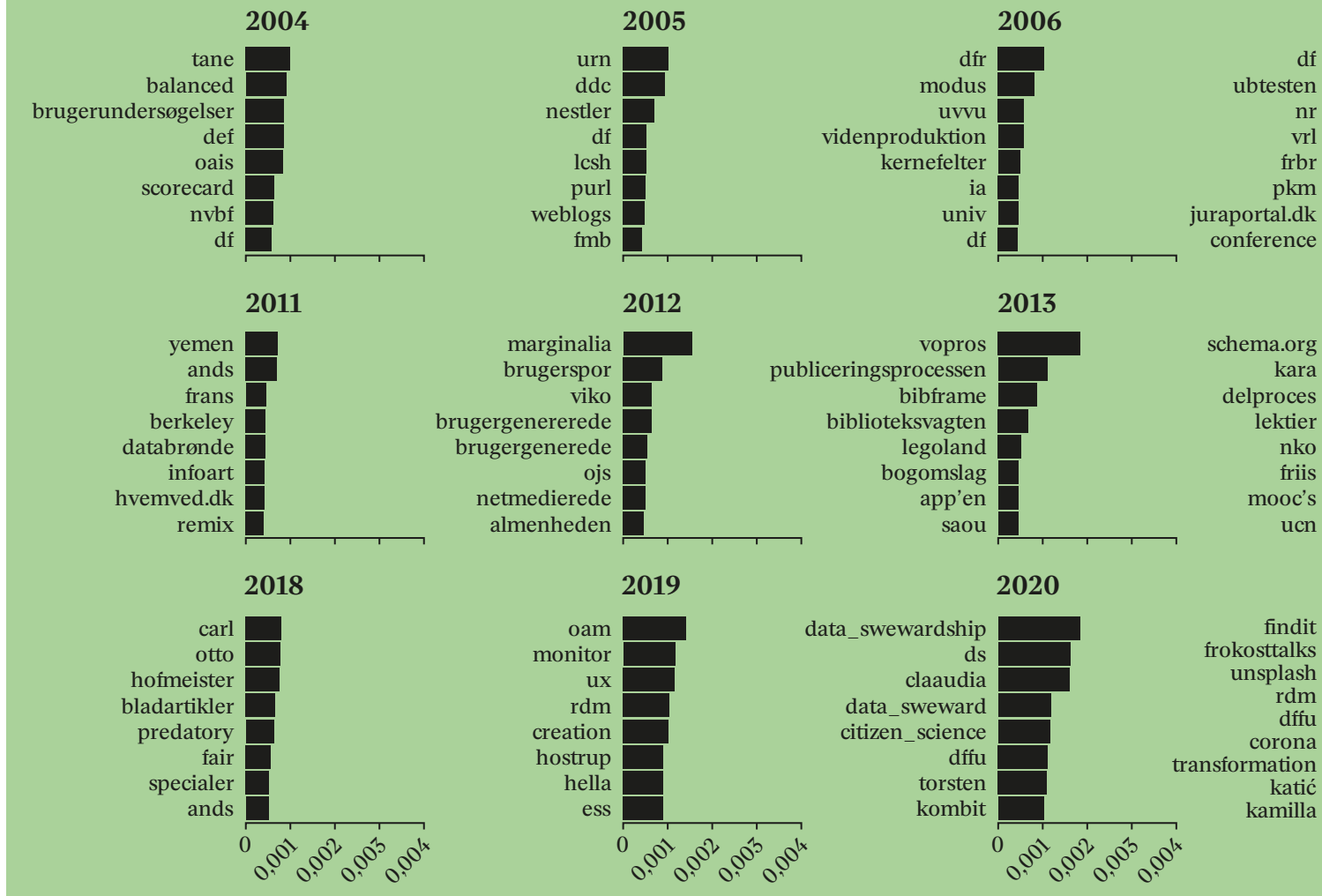
Det følgende kommer til at blive en række nedslag i enkelte års ord fra visualiseringen og en kontekstualisering af ordet og tiden. Af hensyn til omfanget af artiklen er mange ord skamløst blevet overset, ligesom nogle år er udeladt.

2007 – ubtesten.dk

I REVY nr. 2 fra 2007 kunne man læse om lanceringen af den nye e-læringsressource ubtesten.dk. Det var en portal, hvor studerende og forskere kunne teste deres egne kompetencer indenfor "akademisk informationshåndtering". Selvom det fremgik som en test, så lærte man altså informationshåndtering undervejs i testen, hvilket var et bevidst valg, da man vurderede, at test/quiz-formatet var velkendt

Term frequency – inversed document frequency (tf-idf)

Statistikken tf-idf er et mål for hvor vigtigt et ord er for et dokument inden for en samling af dokumenter. F. eks. en novelle inden for en samling af noveller eller en webside inden for en samling websider. I denne artikel er det vigtigheden af et ord for et år af REVY inden for 20 års udgivelser af REVY.



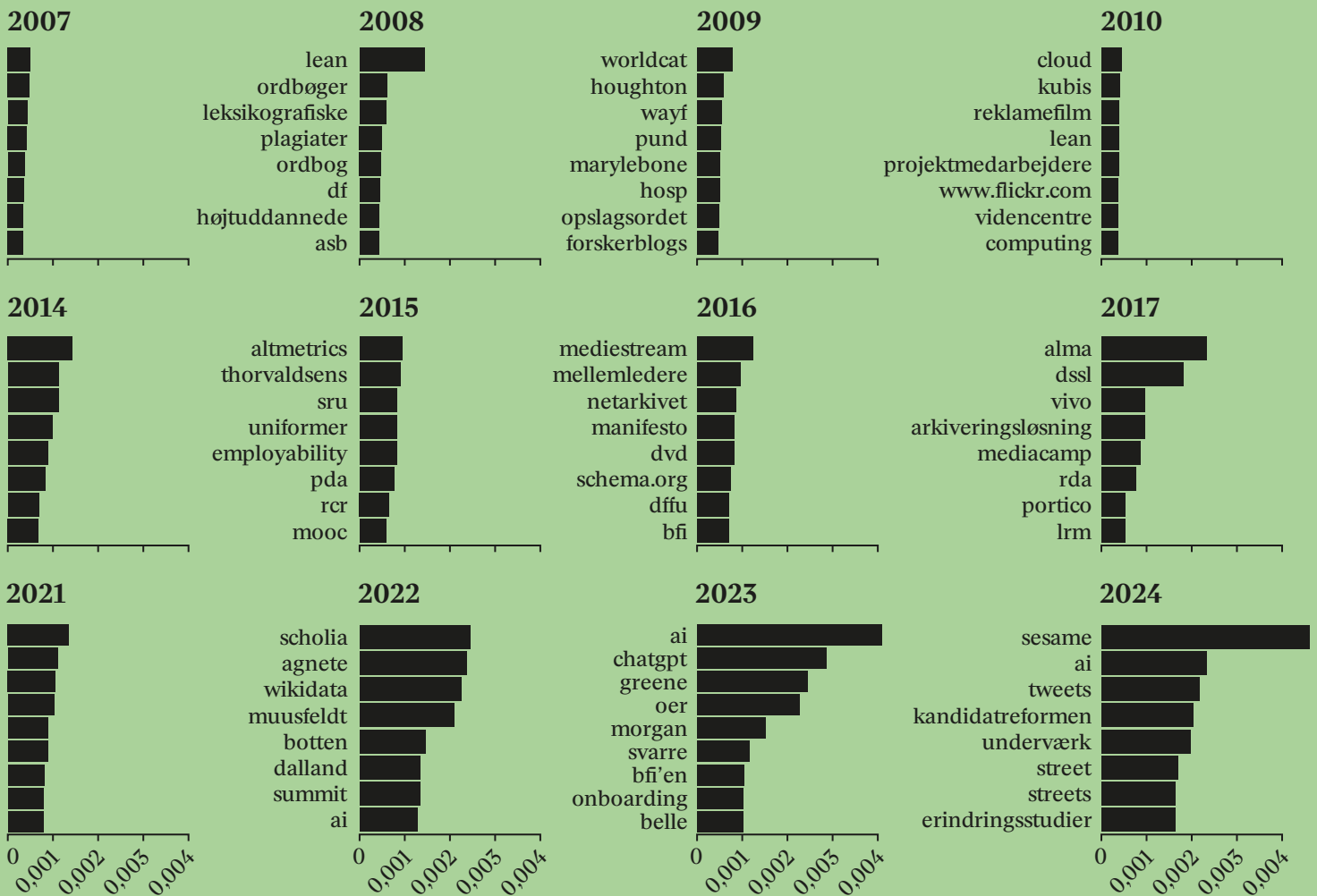
i målgruppen. Således var ubtesten.dk tænkt som et selvstændigt værktøj for studerende og forskere, men det kunne også inkluderes i biblioteksundervisningen. Projektet var delvis finansieret af DEFF (Danmarks Elektroniske Fag- og Forskningsbibliotek). En undersøgelse af ubtesten.dk i Internet Archives Wayback Machine viser, at sitet kørte videre til cirka 2020, hvorefter nogle andre har overtaget domænet fra i cirka 2021. Adspurgt om ub-testen fortæller Susanne Dalsgaard Kragh, at det var en af mange DEFF-finansierede projekter. ”Man kan have mange holdninger til DEFF og de projekter, der blev gennemført, og nogle projekter havde problemer med at leve videre, når de nåede implementeringsfasen. Men uanset hvilken mening man må have, så var DEFF med til at knytte de danske forskningsbibli-

oteker sammen i et udviklings- og arbejdsfællesskab, som ikke mindst på medarbejderniveau var med til at løfte samarbejdet på tværs af sektoren.”

2008 – Lean

Et af tidens hotteste modeord, fortæller Susanne Dalsgaard Kragh og mindes et væld af kurser og workshops, hvor man skulle lære at implementere lean-tilgangen. Det var jo ikke fordi, man ikke tidligere havde kigget på arbejdsgange og effektivisering, men nu skulle alle gennemgå processer med lean-briller. Sådan har der været forskellige bølger gennem tiden. I artiklen fra REVY nr. 1 fra 2008 kunne man læse, at ordet og det tilknyttede

Kilde: www.tidytextmining.com/tfidf



tankegods kom fra den japanske bilindustri og var tiltænkt produktion, hvor man ønskede at skære spild i produktionskæden fra til glæde for kunde og firma. Fra produktionsvirksomheder gik tilgangen sin sejrsgang i andre industrier og herfra til videnstunge institutioner og firmaer, såvel som til offentlige styrelser og kommuner. I den konkrete artikel handlede det om, hvordan Aalborg Universitetsbibliotek i en årrække havde arbejdet med lean-konceptet i deres Accessionsteam.

2009 - WAYF – Where Are You From

Etableringer af Where Are You From beskrives i REVY nr. 6 fra 2008, hvor det handler om arbejdet med at implementere middlewaret,

som alle i dag kender som WAYF, der håndterer login på tværs af en lang række services. Ifølge Susanne Dalsgaard Kragh så indvarsler dette en tid med bedre informationssikkerhed, da tiden før jo var en tid, hvor man ikke var så fokuseret på de sikkerhedsmæssige og rettmæssige udfordringer.

”Alle de nye digitale services gav uanede muligheder for videndeling, og vi var nok lidt naive i begyndelsen og kunne se flere muligheder end trusler.”

Bevidstheden om, at der både kunne være sikkerheds- og rettmæssige udfordringer indfandt sig hurtigt, og WAYF var et initiativ, der blev udviklet i 2008 til at styre og genbruge digitale brugeroplysninger fra forsknings- og uddannelsesinstitutioner og dermed øge sikkerheden både juridisk og teknisk.



2011 – Databrønde

I årene omkring 2011 kollapsede DEFF-projektet om at etablere en national databrønd, som ifølge en beskrivelse af projektet i REVY nr. 8 fra 2007 skulle have indeholdt: "[...] *datatyper som nationalbibliografien, danske bibliotekskataloger, danske forskningsdatabase, dansk kulturarv, globale e-bøger, globale e-prints og globale tidsskriftartikler.*" Adspurgt om databrøndprojektet fortæller Susanne Dalsgaard Kragh, at det var et ambitiøst projekt, og at ideen var tiltalende. "Man var stadig i den digitale opstart, og mængden af data og forskelligheden af produkterne gjorde det ikke lettere at søge og finde relevant materiale. Tanken om en stor databrønd, hvor man kunne søge på tværs var tiltalende, men også en drøm der var umulig at gennemføre både teknisk og økonomisk."

På daværende tidspunkt var udviklingen dog allerede løbet fra ideen om en stor national databrønd. Lignende betragtninger kan læses i REVY 2010 nr. 2 og 3.

Fremkomsten af databrønde i 2011 skyldes således flere artikler i løbet af året, som taler om databrønde i en bredere kontekst og leveret fra flere forskellige kanter.

2013 – Biblioteksvagten

Biblioteksvagten var et projekt, der tog sit udgangspunkt i folkebibliotekerne og som skulle besvare spørgsmål fra borgere på tværs af kommunegrænser via hjemmesiden biblioteksvagten.dk. Projektet udviklede sig fra et forsøgsprojekt med tre folkebiblioteker til en hel række folkebiblioteker, og fra år 2000 kom forskningsbibliotekerne med i samarbejdet.

Der opstod et ønske om at knytte forskningsbibliotekerne sammen med projektet. Susanne Dalsgaard Kragh blev udpeget til projektleder for den del af samarbejdet og skulle sammen med Vera Daggard fra Herning Centralbibliotek skabe en platform, hvor forskningsbibliotekerne og folkebibliotekerne i fællesskab skulle dække vagtbemanding og fordeling af spørgsmål. Susanne Dalsgaard Kraghs opgave var at overbevise forskningsbibliotekerne om det spændende i projektet og skaffe deltagere fra de forskellige biblioteker, der kunne indgå i vagtplanen.

"Der skulle samtidig undersøges muligheder for et nyt system, og allerede dengang blev der kigget på systemer med chatfunktioner og skærmdelinger. Det var et spændende arbejde, som for første gang virkelig prøvede at etablere et samarbejde på tværs af sektoren."

Susanne Dalsgaard Kragh stoppede som projektleder, da hun i 2002 blev leder af Det

“ Det var jo ikke fordi, man ikke tidligere havde kigget på arbejdsgange og effektivisering, men nu skulle alle gennemgå processer med lean-briller. Sådan har der været forskellige bølger gennem tiden.

Teologiske Fakultetsbibliotek på Aarhus Universitet.

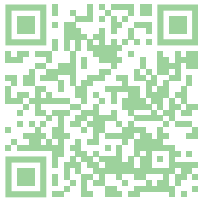
Når biblioteksvagten hopper frem i 2013 skyldes det to artikler i REVY nr. 4, der beskriver det som en ny mulighed for fag- og forskningsbiblioteker at deltage i samarbejdet, men uden at skulle deltage i bemanning og vagtplaner. Denne deltagelse uden om selve bemanningen blev muliggjort takket være et nyt system, Vopros, der blev udviklet primært af Biblioteksvagten.dk og DBC. Systemet muliggjorde udveksling af spørgsmål og svar mellem biblioteker, institutioner og fag ved at danne bindeled mellem forskellige spørgetjenester. Hvad Vopros betød, opfordrede den en til at spørge biblioteksvagten.dk om, men svaret hænger i det uvisse, da biblioteksvagten.dk lukkede den 1. december 2023.

i *Sådan har vi gjort:*

På baggrund af REVYs arkivside har vi indsamlet alle de digitalt tilgængelige udgaver af REVY, hvilket altså dækker perioden fra 2004 op til i dag. Til senere analyser kunne det være rigtig interessant at få digitaliseret de tidligere udgaver og lade dem indgå i en lignende analyse.

Se arkivside:

<https://rauli.cbs.dk/index.php/revy/issue/archive>



Indsamlingen har resulteret i 110 pdf-filer, hvilket altså svarer til en pdf pr. nummer pr. årgang (i 00'erne kunne der være op til otte numre pr. årgang). Disse pdf-filer var navngivet konsekvent og sirligt således, at de indeholdt information om årstal, årgang og nummer. Herefter var næste skridt at udtrække den rå tekst af pdf-filerne via statistikprogrammeringssproget R for at skabe et struktureret datasæt. Takket være navngivningen af filerne indeholdt dette datasæt rå tekst, såvel som information om årstal, årgang og nummer for den rå tekst.

Næste skridt burde være textmining-opgaven, hvor man ved at kvantificere teksten kan finde ordstatistikker af forskellig art fordelt over årene, men indledende analyser gav et benspænd. Det viste sig, at der under tiden er blevet udgivet artikler i REVY på svensk, norsk og engelsk, hvilket gav et mudret billede, når man forsøgte at kvantificere teksten.

Gode råd var dyre, da datas beskaffenhed var af en art, hvor der lå en stor bunke tekst pr. nummer pr. årgang. Med andre ord var det svært at fiske den ene svenske artikel ud af en bunke med alle artiklerne fra et nummer. Løsningen blev at udtrække den rå tekst på side-niveau i stedet for nummer-niveau. Data kom således til at have en form, hvor det var den rå tekst fra enkelte sider med information

om, hvilket årstal, årgang og nummer, en given side tekst kom fra. Håbet på denne baggrund var at kunne identificere de artikler på et andet sprog end dansk og sortere dem fra.

Første forsøg på at identificere sidernes sprog var naturligvis programmatisk. Med en algoritme til at identificere sprog blev de 2387 sider fra REVY 2004-2024 løbet igennem, og algoritmens bud på sproget blev noteret i datasættet. Desværre viste denne tilgang at sig at være massivt fejlbehæftet. Ganske vist identificerede den dansk langt hen ad vejen, men der var for mange fejlidentificerede sprog i blandt. REVY har stadig til gode at udgive artikler på gælisk.

Løsningen blev, beklageligvis, at der skulle menneskeøjne til. Takket være et excel-ark på Teams og studentermedhjælpere, der kedede sig i ydertimerne af bemanningstiden i Nobelparkens biblioteksskranke, blev data opmærket med sprog.

Herefter lykkedes det at få sorteret de ikke-danske artikler fra, således at analyserne viste et klarere billede. Med digital tekst er der et væld af muligheder for analyser inden for textmining, der udnytter, at teksten kan kvantificeres. Men den analyse, der er valgt til den nærværende analyse, er den såkaldt *term frequency – inversed document frequency*. Tanken er her, at man udregner et *terms* (altså et ords) frekvens inden for en række dokumenter – i vores tilfælde 20 års REVY-udgivelser. De højeste frekvenser inden for hvert år vil naturligvis være ord som "og", "den", "det", "der" og så videre. Derfor har man brug for et mål, der tager højde for termets frekvens på tværs af dokumenter – altså på tværs af årene. Dette mål er *inversed document frequency*, der "straffer" ord, som forekommer i alle årene ("den", "det", "der", osv.), således at de falder til bunden af undersøgelsen. Termer der kun forekommer i et enkelt eller nogle få dokumenter "belønnes", og hvis deres term-frekvens er høj, vil de boble op til overfladen i undersøgelsen. I vores undersøgelse har vi taget de otte ord med den højeste tf-idf-værdi inden for hvert enkelt. Det bliver således et udtryk for ord, der er særlige for de enkelte år.