

Den demokratiske søgefunktion

DBC

Christian (udvikler)

Bo (chefstrateg)



Hvorfor er vi her?

Vi har brug for input

- Hvad er målene for ”den demokratiske søgefunktion”?
- Hvordan får vi det biblioteksfaglige med?
- Hvilke ønsker er der fra bibliotekerne og andre?
- Hvilke udfordringer er der politisk, teknisk, ...?

bow@dbc.dk

cbo@dbc.dk

DISCLAIMER

Demokratibegrebet bruges i dette oplæg og workshop for at få os til at reflektere over bibliotekets rolle og mulighed i den digitale virkelighed

Det er ikke udtryk for en bestemt demokratiopfattelse eller en forestilling om at bibliotekerne skal gå ind i en helt ny rolle - folke- og skolebibliotekerne ER allerede en af hjørnestenene det moderne samfund

Wonchi, Ethiopia, 2012



Tablets, telefoner, web...

- Være til stede hvor brugerne er
- Give brugerne værdi

Data science kan hjælpe

Bibliotekernes brugere og borgerne i almindelighed ønsker services og løsninger som de dominerende media giganter leverer gennem deres grænseflader der har en kognitiv forståelse for brugeren og dennes kontekst - men bibliotekerne skal understøtte borgeren ud fra et ikke-kommercielt perspektiv



ER I KLAR
TIL
EN LILLE ØVELSE
- eller quiz?

HVAD ER DETTE – HVEM ER HAN?



DAN WAGNER





OG HAMOG DET?





Aidan Macallan - Data Scientist for Frank Underwood's election campaign in Netflix - House of Cards.

HVAD ER MØNSTRET HER?



DE INVESTERER I OG LAVER
LØSNINGER TIL BRUGERNE BASERET
PÅ BRUGERNES MØNSTRE gennem
anvendelse af data science, mv.

A couple of headlines... [all November '14]

Baidu's Andrew Ng on Deep Learning and Innovation in Silicon Valley

Nervana Systems raises \$3.3M to build hardware designed for deep learning

by Derrick Harris Aug. 21, 2014 - 5:48 AM PST

Artificially Intelligent Robot Scientists Could Be Next Project for Google's AI Firm

Deep learning might help you get an ultrasound at Walgreens

A Googler's Quest to Teach Machines How to Understand Emotions

Google, Spotify, & Pandora bet a computer could generate a better playlist than you can

Butterfly Network Hopes to Bring Deep Learning AI to Medicine

Enlitic picks up \$2M to help diagnose diseases with deep learning

OG HVAD ER SÅ DETTE?



IBM's Milepæle

1997 **Deep Blue** vs. Garri Kasparov : 3,5 – 2,5

2011 **IBM Watson** vs. Brad Rutter & Ken Jennings: 2 – 0

2013 Medical Watson supporterer sygeplejesker på et amr. Hospital og et cancer forskningscenter med faglig beslutningsstøtte og diagnosticerings- og behandlingsstøtte i relation til lungecancer

2015 Watson's Cognitive Computer anvendes til cancerdiagnose, Sundhedsøkonomi, Logistics, Serviceudvikling,.....

2016

Google - DeepMind slår legendarisk Go Spiller og verdensmester Lee Se-dol



2016

THE BIGGEST DISRUPTOR



Google

"The biggest disruptor that we are sure about is the arrival of **big data and machine intelligence**. This disruption will not only change every business globally, it will also have an important impact on the consumer."

Google Chairman Eric Schmidt

Et aktuelt studie fra McKinsey Global Institute konkluderer, "at mangel på det analytiske og ledelsesmæssige talent nødvendigt for at få mest ud af Big data er en betydelig og presserende udfordring (for USA) .

"Rapporten anslår, at der vil være **fire til fem millioner arbejdspladser i USA , der kræver data analyse færdigheder i 2018** , og at forfatterne også identificerer et behov for **1,5 millioner flere ledere og analytikere med dybe analytiske og tekniske færdigheder** ", som kan stille de rigtige spørgsmål og forbruge resultaterne af analysen af store data effektivt.

I REMEMBER WHEN ONLY A DEEP LEARNING SUPER COMPUTER COULD BEAT ME IN A DATA SCIENCE COMPETITION!



"BIG DATA og
datascience /
machine learning

"også for biblioteker



Data Science

“Data Science is the extraction of knowledge from large volumes of data that are structured or unstructured”

- Wikipedia

Data Science

- **Data ⇒ Viden**
- **... understøttet af statistik**

**Family; Father; Mother; Gender;
Height;**

1; 78.5; 67; M; 73.2;

1; 78.5; 67; F; 69.2;

1; 78.5; 67; F; 69;

1; 78.5; 67; F; 69;

2; 75.5; 66.5; M; 73.5;

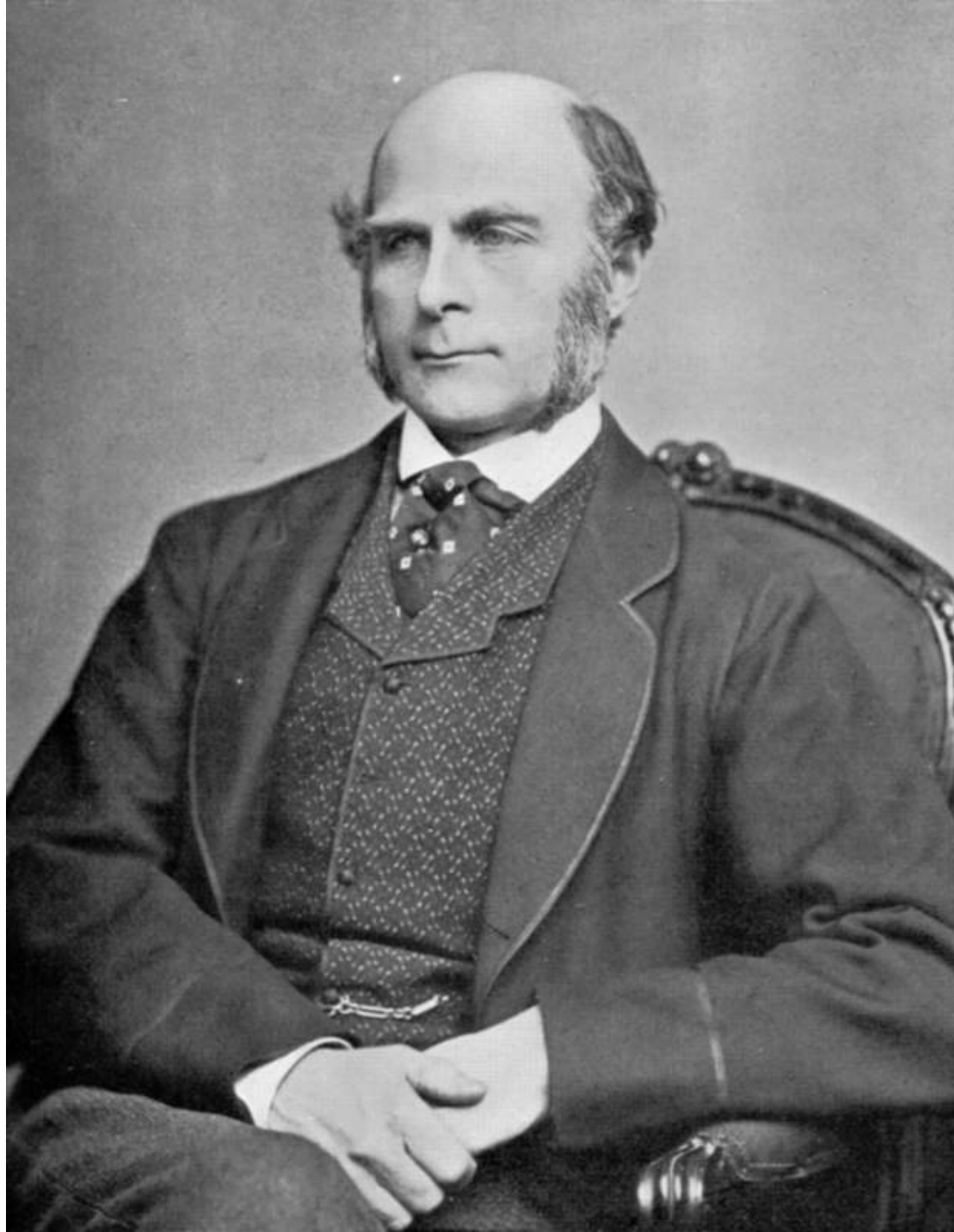
2; 75.5; 66.5; M; 72.5;

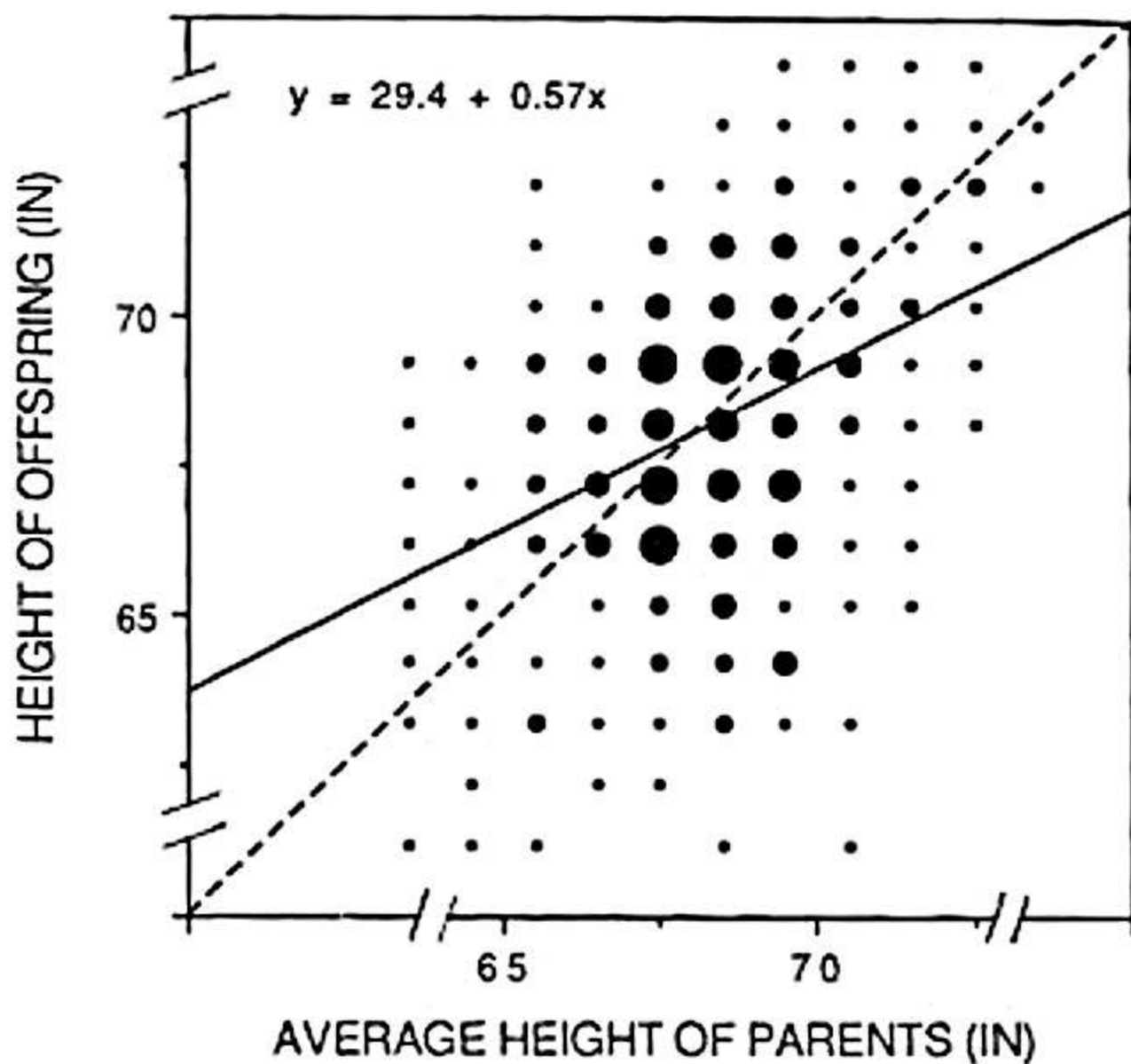
2; 75.5; 66.5; F; 65.5;

2; 75.5; 66.5; F; 65.5;

3; 75; 64; M; 71;

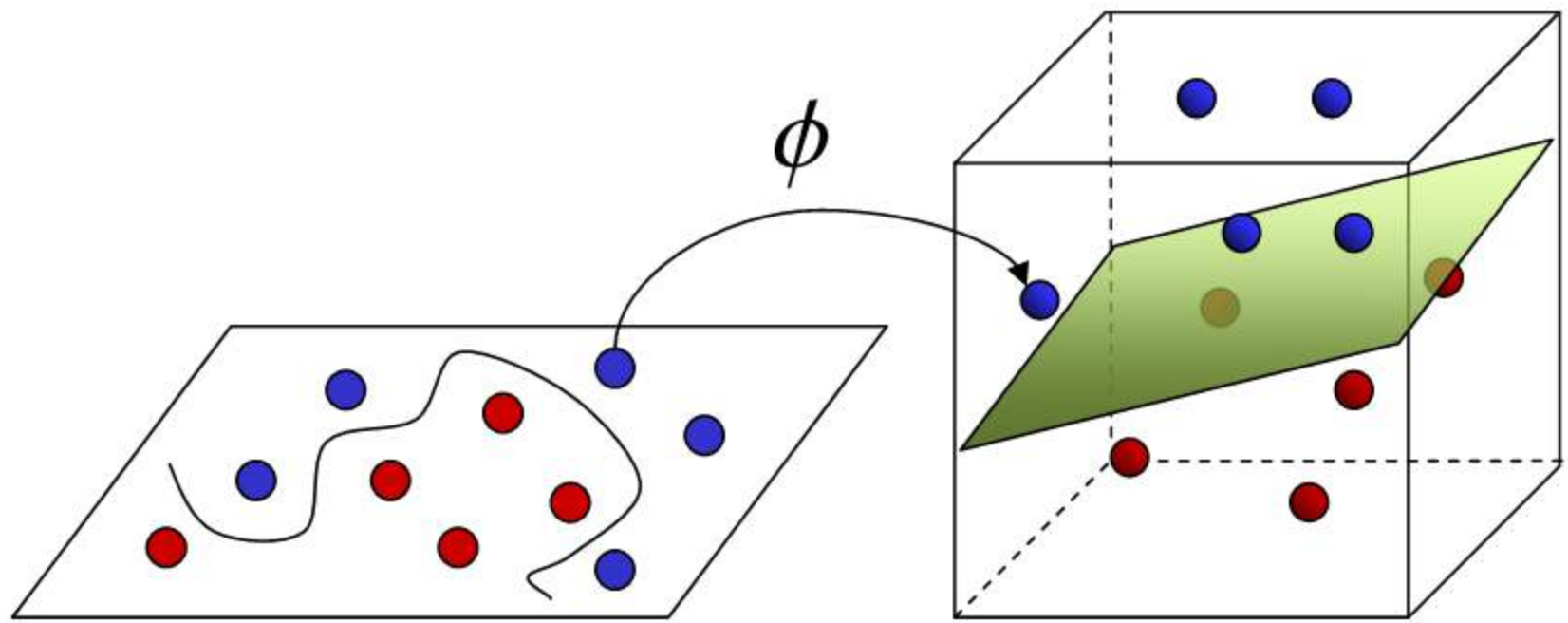
3; 75; 64; F; 68;





Machine learning

- **mere data \Rightarrow bedre resultat**
- **= læring**



Input Space

Feature Space

Data

- **udlånsdata**
- **bibliotekarernes lister**
- **materialevurderinger**
- **fuldtekster**
- **metadata**
- **søgestreng, visninger etc...**
- **brugerens lån, favoritter, profiler...**
- **artikler og sociale medier**
- **årstiden, nyheder...**

Hvad har vi nu

Anbefalinger af materialer gennem forskellige typer af recommendere

Completion-suggest (titler, emner, forfattere, biblioteker etc.)

“Rankering” – forbedret rankering gennem anvendelse af nye typer af metadata (popularitet)

Automagisk indeksering

“Hvad er på vej”

Søgning på ”eget ord” - mapning til metadata-hierarki: Bedre resultater og mulighed for navigering , “Browsning” (metadata navigation)

DDB aktiviteter der udnytter de mange data i videnssystemet til analyser, der kan understøtte serviceudvikling og systemudvikling

”TECH – TANGENT”

Similarity

- **Sim(a,b):** hvor meget ligner værk a og b hinanden målt på lånere (A og B)?

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$

Moon Palace : roman :: Paul Auster

[1] Mand i mørke : roman :: Paul Auster

[2] Brooklyn dårskab : roman :: Paul Auster

[3] Rejser i scriptoriet : roman :: Paul Auster

[4] Leviathan :: Paul Auster

[5] Opfindelsen af ensomhed :: Paul Auster

[6] Mr Vertigo :: Paul Auster

[7] Tilfældets musik : roman :: Paul Auster

[8] Orakelnat :: Paul Auster

[9] Illusionernes bog : roman :: Paul Auster

[10] I det sidstes land : roman :: Paul Auster

- Hverdagen med Aspergers syndrom :: H. W. Gade
- [1] Aspergers syndrom - og hvad så? :: Gunilla
- [2] Hvad er Aspergers syndrom, og hvordan vil det
- [3] Overlevelsesguide - for mennesker med
Aspergers
- [4] Det er godt at spørge - : om Aspergers syndrom
- [5] Aspergers syndrom : fra diagnose til behandling
- [6] Én fod ude - én fod inde : Aspergers syndrom :
- [7] Mellem autisme og normalitet : Aspergers
- [8] Aspergers syndrom :: Tony Attwood
- [9] Børn med Aspergers syndrom :: Matt Winter
- [10] I periferiens centrum : voksne med Aspergers

- LEGO minifigurer : år for år :: Gregory Farshtey
- [1] Den store bog om LEGO DC Universe super heroes & Batman
 - [2] LEGO Star wars - jedi-mesteren Yoda :: Daniel Lipkowitz
 - [3] LEGO legebog : ideer der bringer dine klodser til live
 - [4] LEGO Star Wars figurleksikon :: Hannah Dolan
 - [5] LEGO legends of Chima - bogen om CHI :: Louise Urth Olsen
 - [6] Lego Filmen - et klodset eventyr
 - [7] Star wars i 100 scener :: Jason Fry
 - [8] Den store bog om LEGO Star Wars :: Simon Beecroft
 - [9] Byg med LEGO-klodser :: Warren Elsmore
 - [10] LEGO ideer : slip fantasien løs :: Daniel Lipkowitz

DEMOKRATI TANGENT

Nogle
demokratiske
perspektiver
og parametre



- Hele hovedvægten skal lægges på MØNSTRE ikke individer
- Borgerne skal anmode om og godkende eksplicit hvis bibliotekets digitale services skal gemme viden og information om borgeren som individ, ligesom i "BOGEN KOMMER" - ikke som Facebook, og de evindelige "cookies Acceptance"....
- Hvad end bibliotekerne kan, bibliotekarerne er gode til - skal de digitale løsninger kunne replikere det (delvist 😊)

Litteratur"professoren" på biblioteket skal ikke kun vejlede, formidle litteratur ud fra egen eller andres læsning af teksterne.....

Men ud fra opsamling, aggregering, analyse og bearbejdning af store mængder af data kan vi lave noget der simulerer "den samlede viden"

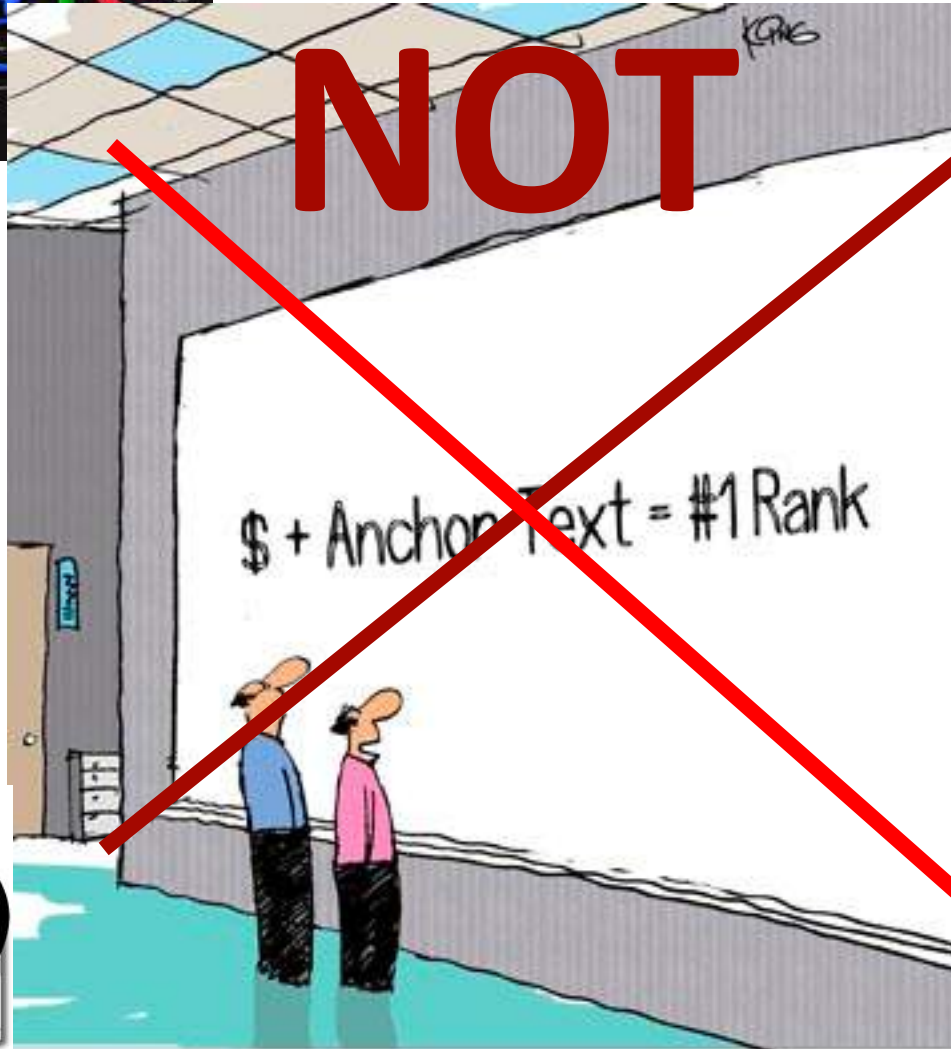
- Give borgeren mulighed for at vælge til og fra
 - Selv give sin stemme til kende via Lister mv.
- Biblioteket skal aldrig kun anbefale det populære, - men det populære kan også være godt

- En anbefaling eller en rankering af en søgning skal give værdi for borgeren
- Den skal ledsages af mulighed for indsigt i hvorfor og hvordan



“The machine learning algorithm wants to know if we’d like a dozen wireless mice to feed the Python book we just bought.”

- Bibliotekernes "BIG DATA" skal ses som et aktiv for den offentlige sektor og skal danne grundlag for deling, kvalitetsforbedring, tilgængeliggørelse,.....




$$PR(A) = (1-d) + \frac{d}{C(T1)} (PR(T1) + \dots + PR(Tn))$$

"This Google Algorithm always works for me!"

HVILKEN FAGLIGHED VI VIL GØRE
TILGÆNGELIG DIGITALT - TANGENT

Bibliotekarernes viden og kunnen
er brugbare i mange sammenhænge
- problemet er at der er så få af
dem



Datascience kan ses som et strategisk værktøj
til bibliotekerne for at kompensere for dette
og kan måske endda bringe bibliotekarernes
kompetencer og viden ind i situationer og på
måder, der kan være med til at innovere
bibliotekerne

I bibliotekssektoren producerer vi en masse metadata - MEN

At skabe og aggregere metadata i de mængder som borgerne behøver i mødet med bibliotekets digitale flader udelukkende gennem

intellektuelle faglige processer, "by hand"

- er et NO GO



Så vi følte en nødvendighed

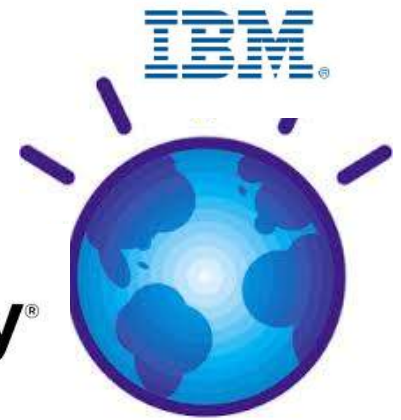
Akademisk netværking, kurser og eksperimenter i: Machine learning, datascience

Inspiration fra kommercielle



amazon.com[®]

NETFLIX



A Smarter Planet



Informationsspecialister kan strukturere viden og indhold & navigere i store mængder af DATA

Biblioteker kan anbefale det relevante for en bruger i en kontekst og formidle sit engagement



Informations-specialisten



The Librarian

I bibliotekssektoren producerer vi en masse metadata - MEN

At skabe og aggregere metadata i de mængder som borgerne behøver i mødet med bibliotekets digitale flader udelukkende gennem

intellektuelle faglige processer, "by hand"

- er et NO GO



Kompetencer simuleret ved hjælp af datascience

Skabe og aggregere metadata og metadatapraksis

Etablering af en ny taxonomi fra et eller flere data set

Kan finde og udvælge nye relevante data set

Kan connecte taxonomier

.....

Librarian intermediary



FILMSTABEN

Bibliotekernes Filmtilbud

Forældre Film Anvendt teater Den filmatiserede 30 nye film

Nye film



11 Forbandelsen

Der brænder en forbandelse over Johanna og hendes far, som bor sammen i et lille hus i en landsby.

Hver nat tjemsæges Johannes far af mærkeligtvisse dyr. For at beskytte ham...

55 min



For sønnen, der har mistet sin lille kærlighed, fortæller den gamle mand, at han...

11 45 min

Forældre



15 The future

Parret Jason og Sophie har besluttet sig for at adoptere en femårig...



15 Trust

14-årige Anna har kun et ønske, da hendes online-chat...



Kompetencer simuleret ved hjælp af datascience

Sætter sig ind i brugerbehov og -ønsker

Har Kritisk vinkel

Fortæller hvorfor hun anbefaler noget til dig

Cognitive "søgemaskine" og recommender-system baseret på Machine Learning og datascience, eksisterende web services, bruger feedback, brugeradfærd, taxonomier, metadata, data sæt fra sociale media, etc. giver brugerne det bedst mulige content/vej til content afhængigt af kontekst....med transparens

En kombineret vision-
BIG **META** DATA og et forsøg på at
styrke bibliotekerne
baseret på Machine Learning,
Datascience, cognitive
computing,.....
og biblioteksfaglighed - i et
demokratisk og lærende
perspektiv

Lad os for et øjeblik kalde det
for "Digitale Demokrati"

Lidt definitioner

- **Big data** er et begreb der bredt dækker over indsamling, opbevaring, analyse processering og fortolkning af enorme mængder og typer af data
- **Datascience** er generelt set etablering af viden ud fra data. Det indbefatter metoder fra mange discipliner, matematik, statistik, datalogi, filosofi,....
- **Machine Learning og Deep Learning** er en disciplin, der arbejder med at undersøge og etablere læring fra data gennem algoritmer
- ***Cognitive computing*** er et sammensat begreb for store investeringer og forskning i Big Data, datascience, machine learning og high scale computing, der skal flytte grænserne for fladerne, grænserne og skaleringerne mellem mennesket og teknologier