big data - big science

Lars Kai Hansen
www.imm.dtu.dk/~lkh
themes

big data hope or hype?
my vision
the mechanisms
the science and the algorithms
a glimpse of the future
What is big data?
Digital footprints of people, “things”, transactions, ...

The pursuit of meaning, value, monetization in same...

Information asymmetry is a challenge to democracy
but Google is getting smarter:

“... Google will buy London-based artificial intelligence company DeepMind. The Information reports that the acquisition price was more than $500 million, and that Facebook was also in talks to buy the startup late last year. DeepMind confirmed the acquisition to us, but couldn’t disclose deal terms...” TechCrunch Jan 26, 2014
Gartner Group’s
Hype Cycle for Emerging Technologies, 2013

"Area9 blev i februar købt af ... McGraw-Hill Education for et hemmeligt beløb, der ifølge Børsen ligger på den pæne side af en milliard danske kroner."

http://www.gartner.com/newsroom/id/2575515
The information you need when you need it.

Support productivity and well-being.
Respect democratic living and privacy.
The mechanisms: Search engines

Search engine = Google?

PROOF #8

According to Google trends, the term "Google" is searched for more than the terms "God", "Jesus", "Allah", "Buddha", "Christianity", "Islam", "Buddhism" and "Judaism" combined.
Our brains are search engines

Primary cognitive function is to query food and sex

Brains search by activating muscles

Brains don’t see - they look! (Bajchy)
Social intelligence? search too...

Social abilities allow division of work and specialization

...if we can find somebody to share with

Specialization requires communication of recipes

Facebook
LinkedIn
Dating sites
The internet hosts numerous specialized big data indexes for search

- Wikipedia (facts)
- Facebook (friends)
- Youtube (video)
- Last.fm (music)
- Dating.dk (dates)
- Twitter (news)
Big data mechanisms: Shallow indexing

The Unreasonable Effectiveness of Data

Alon Halevy, Peter Norvig, and Fernando Pereira, Google

The same meaning can be expressed in many different ways, and the same expression can express many different meanings.

For many tasks, words and word combinations provide all the representational machinery we need to learn from text.

Because of a huge shared cognitive and cultural context, linguistic expression can be highly ambiguous and still often be understood correctly.
Back to Silicon valley -98
Larry Page

Sergey Brin

Gates Building, Dept Computer Science
Stanford University, CA, USA
Larry Page og Sergey Brin founded Google in 1998 based on their recent scientific discovery

This equation explains random walk traffic patterns

$$P(j) = \sum_{k \in I(j)} \frac{P(k)}{|O(k)|}$$

Your PageRank is the average number of random walkers visiting you

[www.google.com/corporate/tech.html](http://www.google.com/corporate/tech.html)
What happens when you query Google?

Query term is located in the Google index

Web pages with your words are relevant-

Pages ranked by Global PageRank are served

Everything happens in less than a second!
Jaguar

Jaguar.dk ☆
Salg af nye og brugte biler. Værksted, reservedele og tilbehør. Sitet indeholder oplysning om modeller, biler og nyheder. Virtuelt showroom med butik og ...
Tilkorte biler - Nye biler på lager - Priser & udstyr
www.jaguar.dk/ - Cached - Lignende

Tiger - Det er for vildt! ☆
Tiger har altid noget du kan bruge - hele året rundt. Masser af sjove, nyttige ting inden for bolig, pleje, legetøj, dyrefoder, kontor, krydderier og meget ...
Tigers katalog - Adresser og åningstider - Job i Tiger - Hobby & fritid
www.tiger.dk/ - Cached - Lignende

Jaguars.com - 2010 Season Tickets On Sale Now! ☆ - [Oversæt denne side]
The official team site with scores, news items, game schedule, and roster.
2010 Draft - Tickets - Message Board - TEAM
www.jaguars.com/ - Cached - Lignende

Billeder af ~jaguar - Rapportør billeder

Jaguar International - Jaguar International ☆ - [Oversæt denne side]
Our mission at Jaguar has been to create and build beautiful fast cars. The XK, XF, X-TYPE and now All New XJ bring the exhilaration of driving to life.
www.jaguar.com/ - Cached - Lignende

Jaguar
Lækre designerbiler kun kr. 1495
Upload dit foto og prøv dem online
www.louisielsen.dk

Jaguar
Find ny brugt bil - ALD CarMarket!
kvalitetsbiler m. fejler FDM-tests
www.aldcarmarket.dk

Jaguar på bilguide
Test og omtaler af Jaguar.
Alt i guides, sport, lukker & brugt
Bilguide.dk/Jaguar

Frisersaks.dk
Danmarks største udvalg af kvalitets frisørsaks.
www.frisersaks.dk

Jaguar på Biltorvet
35.000 brugte biler fra hele landet
find din Jaguar på Biltorvet
www.biltorvet.dk

Nissan service
Auto service til Nissan,
...future of big data
Mechanisms: Human predictability

A new picture of human predictability is emerging

• Barabasi group’s study of mobility (70–93% /hour)
• Individuals are predictable (intersubject variability is huge ... long tail ... “power law”)
• DTU mobile phone context tracking (80–90% /4 min)

Why is this important?

• Curiosity: Understanding the human brain
• Engineering: New services/products are based on human predictability (the Google paradigm)

C. Song, Z. Qu, N. Blumm, A.-L. Barabási
Limits of Predictability in Human Mobility
Science 327, 1018-1021 (2010).
Algorithms

Approximate inference

Text modeling

World record (speed, non-zeros) topic model (MLSP 2010)

AFINN is now SAS Analytics’ emotion lexicon

Streaming signal processing

The Smartphone Brain Scanner: A Portable Real-Time Neuroimaging System

Arkadiusz Stopczynski, Carsten Stahlhut, Jakob Eg Larsen, Michael Kai Petersen, Lars Kai Hansen

Section for Cognitive Systems, DTU Compute, Technical University of Denmark, Kgs. Langeland, Denmark
Deep indexing
what about subjective data?
Quantifying subjectivity

“Affective computing”

Affective computing is research in systems that can recognize, interpret, process, and simulate human emotion.

Emotions are omnipresent and extremely important to communication / opinion formation / intent reading etc.

Psychology of emotion is well developed but still far from complete...

7. The law of hedonic asymmetry: “Negative emotions last longer, positive emotions fade”
Sentiment detection

A step towards understanding subjectivity, opinion

Important to many services
  Recommender (Amazon reviews)
  Information navigation, e.g. navigating music

Oasis "Wonderwall"
Emotional content in the song lyrics through time

Proactive search ... Mind reading?

Behavioral modeling

_recommender system_

Technology that anticipates your needs

Lifelogging

Store everything you do, think and communicate

Quantified self movement

Gordon Bell MyLifeBits
Search Apps

Three audio engines (MIRocket, CastSearch, Muzeeker)

Brede Search neuroinformatics engine

FindZebra, diagnostic queries
How deep can you go?
Web 3.0 — A mind reading internet
EEG — brain monitoring

24/7 state monitoring
Unconscious queries
Privacy model

Smartphones as pocketable labs: Visions for mobile brain imaging and neurofeedback
Arkadiusz Stopczynski *, Carsten Stahlhut, Michael Kai Petersen, Jakob Eg Larsen, Camilla Falk Jensen, Marieta Georgieva Ivanova, Tobias S. Andersen, Lars Kai Hansen
