

Intelligent søgning efter information (og viden?)



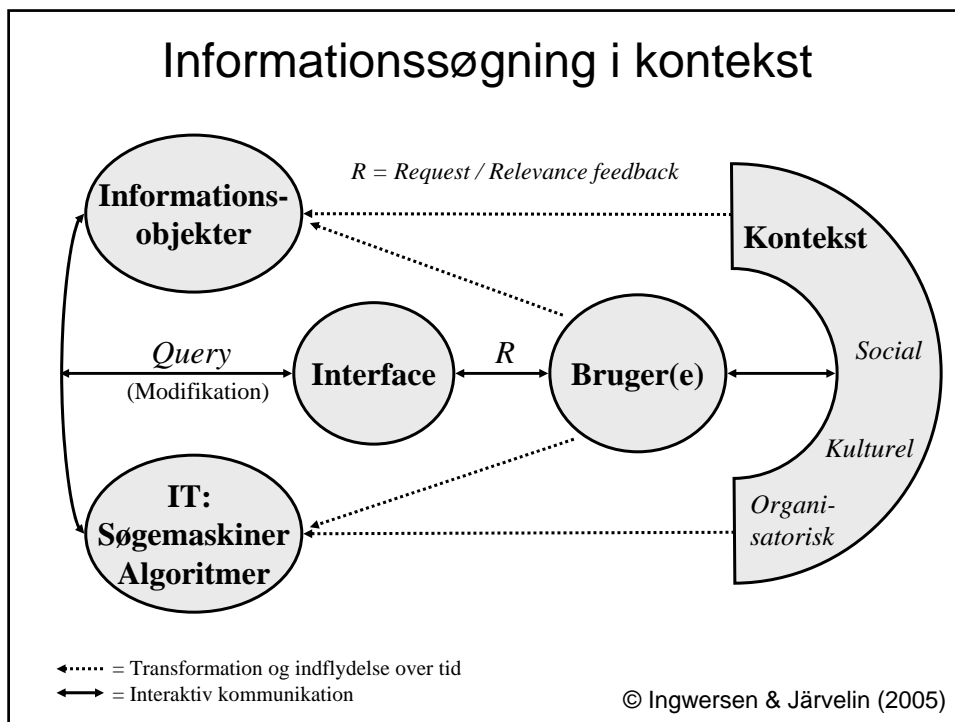
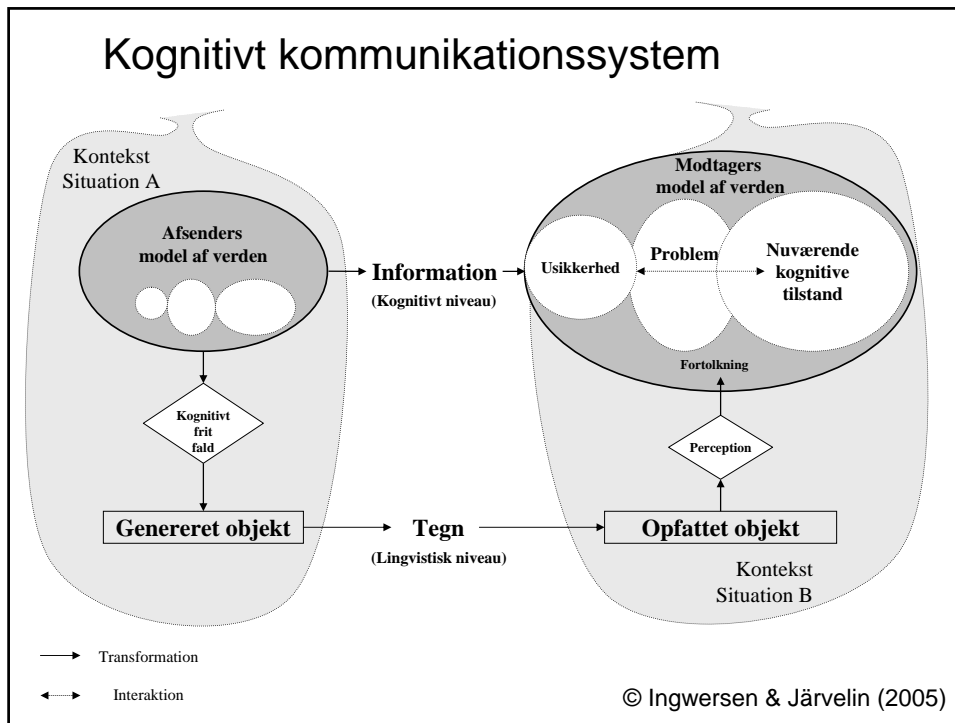
Birger Larsen, ph.d.
Institut for Informationsstudier
Danmarks Biblioteksskole
København
email: blar@db.dk

Program



- Information? Viden? Intelligent søgning?
 - ... set fra et kognitivt synspunkt
- Kunstig intelligens og søgning
- Status og fremtiden
- Links og referencer

2



Intelligent søgning



- Hvor skal intelligensen ligge?
 - Hos brugeren
 - Kræver uddannelse af brugere
 - ...og god dokumentation af systemerne
 - I søgemaskinen
 - Stor skala = kunstig intelligens
 - Mindre skala: diverse former for brugersupport under søgning

5

Kunstig intelligens og søgning



- En lang række prototyper blev testet fra midten af 1980'erne og frem
- De prøvede alle at simulere intelligens ifht. søgning på en eller flere måder, bl.a. vha.:
 - Brugermødelles (f.eks. søgeviden, mål med søgningen, type af behov, stereotyper)
 - System- og domænemødelles
 - Mødelles af informationen i dokumenterne
- De forskellige datastrukturer blev kombineret via regelsæt (f.eks. hvis-så regler)

Se f.eks. Ingwersen (1992, kap. 7) eller Mizzaro (1996)

6

I³R – *Intelligent Information Retrieval*

I3R Data Structures		
Model	Conceptual Structure	Implementation
Request Model	Probabilistic request	Hash Table of Structures
Domain Knowledge Model	Semantic Net	Hash Table of Structures
User Model	Stereotypes and Expectations	Association List
Document Representation	Network of Documents and Terms	Relational Database
Browsing Model	Semantic Net fused with Net of Docs. & terms	Hash Table of Structures

The screenshot shows a graphical user interface for the I3R system. At the top right is a red square with a white crown icon. Below it, the title 'I3R – Intelligent Information Retrieval' is displayed. The main interface consists of several windows:

- Browse Map:** A window showing a semantic network diagram. It features a central node labeled 'Doc' connected to other nodes: 'Auth', 'Term', and 'Doc'. There are also nodes for '25 Docs' and 'Doc?'. Relationships are indicated by lines with labels like 'N N', 'N N Cite', and 'Bib'.
- Document:** A window displaying the text of a document titled 'The Structure of the "THE"-Multiprogramming System'. The text is partially obscured by a black box.
- Content:** A window showing a list of content items: 'Relevant', 'Author's Docs', 'Journal Issue', 'Nearest Neighbors', 'Bibliography', 'Citations', 'Done', and 'Help'.
- Window:** A small window with buttons for 'Window', 'Suspend', and 'Help'.

Se Croft & Thomson (1987)
og Ingwersen (1992, kap. 7)

Kunstig intelligens og søgning

- Prototyperne med kunstig intelligens til søgning er i dag stort set opgivet
 - Omkostningerne ved at etablere og vedligeholde vidensbasen var for stor
 - Det lykkedes aldrig rigtigt at få gode resultater i andet end små samlinger og små domæner
 - For megen indblanding fra systemet kan føles begrænsende (Mr. Clips-effekten)
- En styrke var at man forsøgte at få bedre og mere omfattende forespørgsler fra brugerne

Status



- Meget store dokumentmængder i elektronisk form i mange forskellige formater
- Mange leverandører med omfattende og sammentænkte løsninger
- Forskningen i informationssøgning (*Eng. = Information Retrieval*) styrkes og bredes i mange retninger
 - Multi-Agent Information Retrieval, Recommender Systems, Personalization, Distributed Searching, Cross-language IR

9

Fremtiden 1



- Større standardisering af dokumentformater?
 - Bedre værktøjer til skabelse, f.eks. skabeloner
 - Tvungne metadata / automatisk generering
- ... vs. endnu større diversitet
- OpenSource løsninger?
 - MG, Lemur, Terrier
- Bedre udnyttelse af **kontekst**
 - Mange typer af kontekst...
(Se Ingwersen, van Rijsbergen & Belkin, 2004)

10

Fremtiden 2



- Relevans feedback
 - Information fra relevante dokumenter (markeret af brugeren) anvendes til at forbedre forespørgslen
 - Dynamisk tilpasning over tid (Campbell, 2000)
 - Implicit feedback (White et al., 2004)
- Brugertilpasning; Brug af profiler
- Bedre præsentation af resultater
 - *Query biased summarisation* (White et al., 2003)
 - Dele af dokumenter i stedet for hele dokumenter
 - Søgning i XML dokumenter/elementer (Se INEX)

11

Links



- OpenSource systemer
 - MG (<http://www.cs.mu.oz.au/mg/>)
 - Lemur toolkit (<http://www-2.cs.cmu.edu/~lemur/>)
 - Terrier (<http://ir.dcs.gla.ac.uk/terrier/>)
- INEX (INitiative for the Evaluation of XML IR)
 - <http://inex.is.informatik.uni-duisburg.de/2005/>



12

Referencer



- Campbell I (2000): *The ostensive model of developing information needs*. University of Glasgow, 207 p. (Ph.d.-afhandling - <http://www.dcs.gla.ac.uk/~iain/papers/>)
- Croft WB, Thompson R (1987): I³R: a new approach to the design of document retrieval systems. *Journal of the American Society for Information Science*, 38(6), 389-404.
- Ingwersen P (1992): *Information Retrieval Interaction*. London: Taylor Graham, 246 p. (<http://www.db.dk/pi/iri/>)
- Ingwersen P & Järvelin K (2005): *The Next Turn – Integration of Information Seeking and Retrieval in Context*, Springer Verlag. (under udarbejdelse)
- Ingwersen P, van Rijsbergen K & Belkin N (2004): *ACM SIGIR Workshop on "Information retrieval in Context"*. (IRiX workshop notes - <http://ir.dcs.gla.ac.uk/context/>)
- Mizzaro S (1996): *Intelligent interfaces for information retrieval: A review*. Technical Report UDMI/18/96/RR, University of Udine. (<http://www.dimi.uniud.it/~mizzaro/research/pub.html>)
- White RW, Jose JM & Ruthven I (2003): A task-oriented study on the influencing effects of query-biased summarisation in web searching. *Information Processing and Management*. 39(5), 707-733. (Præprint: <http://www.umiacs.umd.edu/~ryen/publications.html>)
- White RW et al. (2004): A Simulated Study of Implicit Feedback Models. *Proceedings of European Conference on Information Retrieval (ECIR 2004)*, p. 311-32. (Præprint: <http://www.umiacs.umd.edu/~ryen/publications.html>)

13