

Amerikanske biblioteksdata-baser - forskning og udvikling

Nogle iagttagelser i anledning af en studierejse

af forskningsbibliotekar, cand.mag. Anne-Mette Kirkeby

I november 1988 deltog jeg i en studierejse til New York, Washington og Chicago. Formålet med besøget var bl.a. at udveksle erfaringer med amerikanske kolleger og at studere forholdene på nogle akademiske biblioteker. Da jeg arbejder på et humanistisk universitetsbibliotek, var jeg særlig interesseret i at iagttage, hvorledes de amerikanske studerende benyttede de forskellige tilbud inden for den nyeste informationsteknologi.

De mange valgmuligheder og den voldsomme ekspansion på området gør det spændende at foretage en sådan studierejse. Der er lige så mange forskellige onlinekataloger, som der er biblioteker, og de fleste steder havde de et stort udvalg af CD-ROM og videoplader til fri afbenyttelse for de studerende. Vi så, hvorledes de filmstuderende på New York University Library sad foran skærme og studerede filmene med teksterne ved siden af sig. På Columbia University Library i New York havde de klassiskstuderende en fuldtekstdatabase med tekster fra Homér til år 600 f.Kr. En database, som blev udviklet i begyndelsen af 70-erne med navnet *Thesaurus Linguae Græcæ*, og som efter sigende er meget nyttig, men vanskelig at arbejde med. Alle steder kappedes studenterne om at benytte de store bibliografiske databaser på CD-ROM. Især MLA bibliography og Dissertation Abstracts var populære hos humanisterne. - På New York Public Library havde man lagt store databaser som Psychological Abstracts, Medline m.m. ind i bibliotekets online katalog. Man fremhævede, at det, folk ønskede, var information - ikke bøger. Denne tendens til at overføre baser fra CD-ROM til online bibliotekskataloger tegnede sig på flere store forskningsbiblioteker.

På Lincoln Centrets bibliotek så vi de balletstuderende sidde ved video-skærme og iagttage de berømte pas de deux. Måske en idé til Erik Aschengreens dansecenter på Københavns Universitet på Amager. Kunsthistorikerne havde adgang til hele museers beholdning af billeder. Således foretager

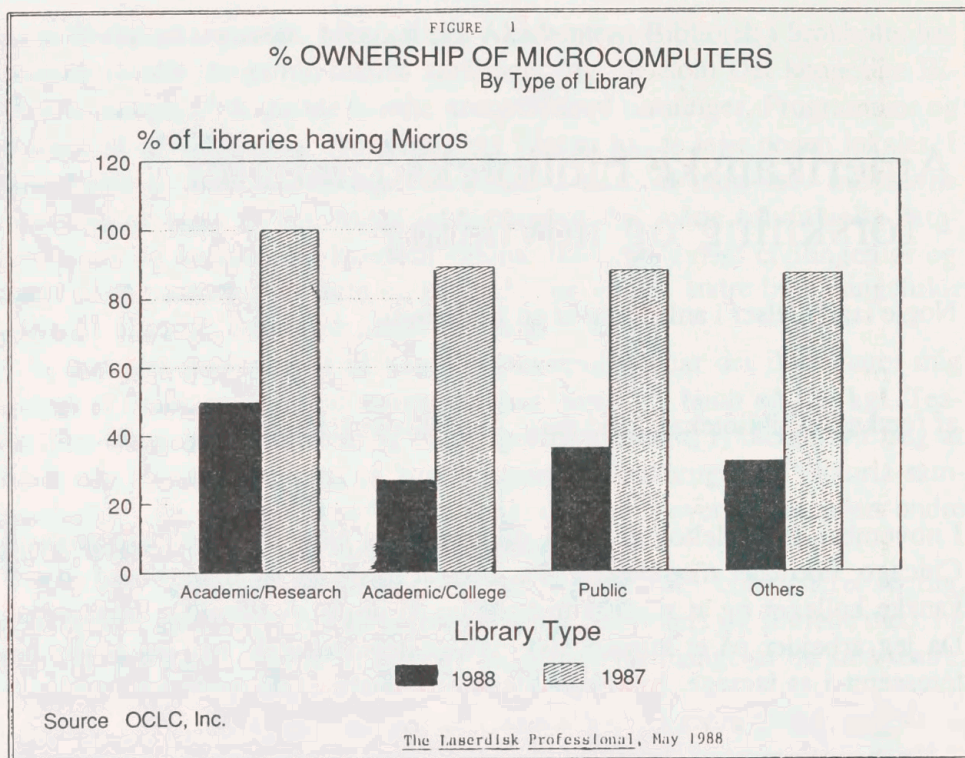


Fig. 1

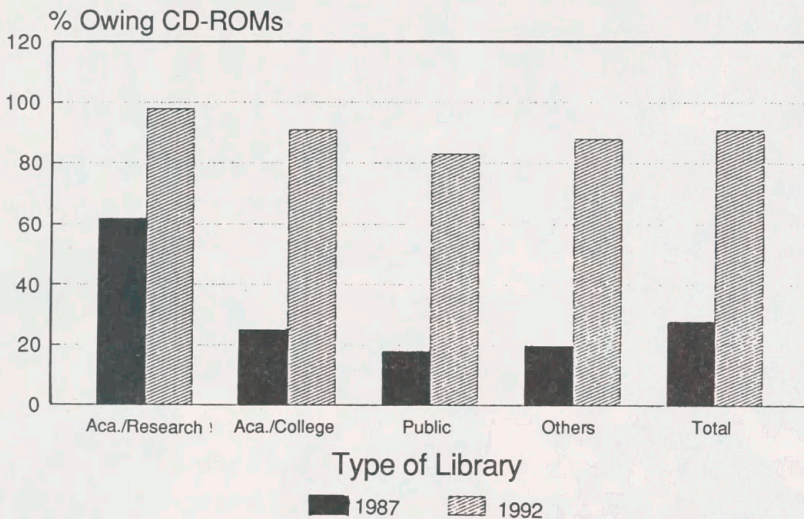
Smithsonian eksperimenter med CD-I (Compact-Disk-Interactive) for at fremstille *Treasures of the Smithsonian*. Et museumskatalog med såvel billeder i farver som tekst og spørgsmål/svar faciliteter.

I det hele taget viste tendenser en stigende brug af andre former end magnetbånd til produktion af databaser, nemlig floppy disc, video disc og CD-ROM. Det er især CD-ROM, som er dominerende på markedet for optiske informationssystemer, og de benyttes især inden for uddannelses- og biblioteksvæsenet. I 1988 blev antallet af CD-ROM titler fordoblet fra 189 til 390 (de fleste inden for naturvidenskab og medicin). Hele online databasemarkedet er støt stigende. De sidste uafhængige markedsundersøgelser viser en stadig vækst på 20% i USA og i Europa. Markedsanalyser tyder imidlertid også på, at CD-ROM markedet vil forblive en begrænset del af hele online databasemarkedet. Det menes ikke at ville fortrænge dette i fremtiden.¹⁾

En af hovedårsagerne til, at CD-ROM har vundet sådan indpas på bibliotekerne, er den næsten 100% udbredelse af mikrodatamater i USA. Bibliotekerne hører her til den mest automatiserede og teknologiforbrugende del af samfundet (se fig. 1 og 2). Desuden er telekommunikationsudgifterne steget meget i de sidste år.

FIGURE 2

% OWNERSHIP OF CD-ROMs By Type of Libraries, 1987-1992



Source OCLC, Inc.

The Laserdisk Professional May 1988

Fig. 2

De mest solgte CD-ROM produkter er online-baserede databaser, som f.eks. ERIC, Medline, PsycLIT og Dissertation Abstracts Int. Det er de større akademiske biblioteker, som har de fleste CD-ROM, og de mest succesfulde CD-ROM plade-fremstillere er de firmaer, som har tilknytning til online markedet, nemlig Information Access Company, Wilson og Dialog.

Amerikanske online bibliotekskataloger

Med hensyn til de amerikanske online bibliotekskataloger tegner der sig et broget billede. De første opstod for 15-20 år siden som non-profit sammenslutninger med tilbud om fælles edb-katalogisering, distribution af kartotekskort, microfiche o.s.v. Det var især 3 store databaseforetagender, som var dominerende, nemlig OCLC, RLG/RLIN og WLN.

OCLC

OCLC var blandt pionéerne. The Ohio Library Center blev oprettet i 1967 i Columbus, Ohio. Det var et fælles databasesystem, som oprindeligt var



Læsesalen på Harper Library, the University of Chicago.

udarbejdet til Ohio College bibliotek til katalogiseringsbrug, som accessionskatalog og som en del af det nationale databasesystem. I 70'erne fik ikke-akademiske biblioteker adgang til at blive medlemmer, og det samme fik biblioteker uden for staten Ohio. På dette tidspunkt talte mange biblioteksledere med optimisme om et kommende nationalt database-system, som skulle styres centralt. Library of Congress, der er kongressens bibliotek i Washington, havde i 60'erne udviklet MARC formatet, hvilket var en grundlæggende opdagelse for hele udviklingen af edb-katalogsystemet i verden. Library of Congress stod for distributionen af de såkaldte MARC bånd med maskinlæsbare katalogposter. Disse poster udgør grundstammen i databasesystemer som OCLC, og det var derfor på tale at placere et centralt nationalt databasesystem på Library of Congress, som er et slags uofficielt nationalbibliotek for USA. Herfra skulle de store forskningsbibliotekers databaser evt. kædes sammen med de store bibliografiske baser.

Disse planer blev aldrig realiserede, der var ikke den nødvendige økonomiske basis og heller ingen støtte fra regeringens side. Library of Congress valgte derfor at opgive rollen som nationalt databasecenter. Det repræsenterer dog USA med hensyn til standarder, og det har længe været den eneste

modtager af katalogposter fra andre lande. Dette har dog ændret sig i de sidste par år, hvor OCLC og RLG har fået deres egne aftaler med udlandet vedrørende katalogiseringsbidrag til databaserne. Så vidt jeg ved, er der således planer om, at det danske ALBASystem skal med i OCLC. Det er nu blevet en slags uofficiel national database med over 17 millioner Library of Congress MARC og British Library UKMARC poster. Den vokser med 30.000 poster om ugen og benyttes af mere end 6.000 biblioteker i USA og 18 andre lande.

OCLC er stadig en non-profit organisation. Den er selvfinansierende gennem betaling for produkter og service. Fra tid til anden har OCLC modtaget støtte til sine talrige udviklingsprodukter, f.eks. fra undervisningsministeriet i USA. Nyheden i 1988 var CD-ROM med 14 databaser. OCLC siger selv, at de bevæger sig væk fra bibliografisk information over mod elektroniske produkter i bredere forstand med fuldtekst i Macintosh format.²⁾ Det næste vil blive en musikbase, Hyper Bach, med både musik og tekst.

OCLC er et totalt databasesystem, hvilket vil sige, at der ydes service inden for alle biblioteksfunktioner, katalogisering, lån fra det ene bibliotek til det andet, tidsskriftkontrol og accession. Medlemmerne har mulighed for at overtage poster og benytte dem til eget brug og bidrage med nye, samt modtage magnetbånd og CD-ROM med MARCposter.

RLG/RLIN

RLG conspectus on-line er den anden af de store databasesammenslutninger i USA. Det er ligeledes en non-profit organisation, som ejes og styres af medlemmerne (The Research Libraries Group, inc.). Medlemmerne er store universiteter og forskningsinstitutioner i USA. Desuden har RLG mange "programmatic Members", som er institutioner, der deltager i et eller flere af selskabets programmer. Det finansieres dels ved afgifter, som medlemmer betaler (meget høje afgifter), dels ved indtægter fra det automatiserede informationssystem RLIN (Research Libraries Information Network). RLIN er et meget udbredt databasesystem, og det tjener som onlinekatalog både for medlemmer af RLG og ikke-medlemmer.

RLG blev grundlagt i 1974 af Harvard, Yale og Columbia universiteterne. Desuden deltog New York Public Library, som er et akademisk bibliotek med lukkede hylder. Et af formålene med oprettelsen var at etablere et edbaseret bibliografisk system, som skulle skabe en forbedret informationssøgning, bl.a. ved mere fleksible søgemetoder, Boolske operatorer m.v. OCLC har først for nylig indført disse "dyre" søgefaciliteter. Her prioriterede man

andre ting højere, så som udlån og accession. - Der er temmelig strenge regler for medlemsskab af RLG, og kun 67 store forskningsinstitutioner har kvalificeret sig.

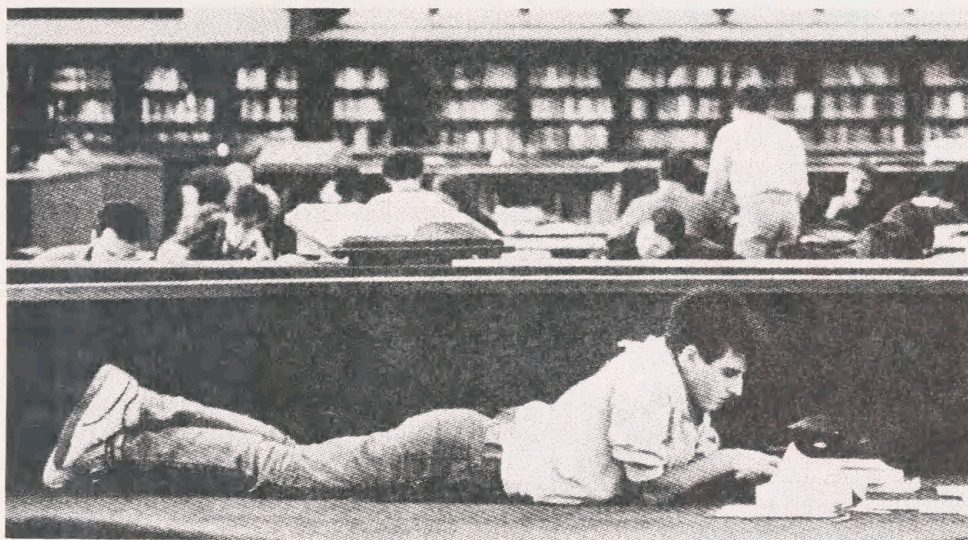
WLN

WLN (Western Library Network) hed indtil 1987 "Washington Library Network". Det blev dannet i de tidlige 70'ere. Senere har sammenslutningen fået medlemmer fra de omgivende stater i det nordvestlige USA, British Columbia og Canada. Det er den mindste af de 3 store datasystemer og er især kendt for at have bibliografiske poster af høj kvalitet, d.v.s. at de er genstand for omhyggelig kontrol m.h.t. korrekte navneformer og "subjekt headings". På grund af det høje niveau er WLN's software meget eftertragtet af biblioteker i andre lande.

Decentraliseringen i 80'erne

De store databasesystemer dominerede på bibliotekerne i USA i 10-15 år. Men i begyndelsen af 80'erne tog udviklingen en anden retning mod en større grad af decentralisering. Et stort antal nye bibliotekssystemer blev sendt på markedet. Bibliotekerne blev mere interesserede i at løse problemerne vedrørende edb-accession og udlån tættere på funktionerne. Gennem brugen af de store systemer var den nødvendige ekspertise blevet spredt rundt på bibliotekerne. Store firmaer som GEAC og VTLS bragte nøglefærdige systemer på markedet i 1980'erne. Disse systemer omfattede alt, både hardware og software til alle biblioteksfunktioner, Vi så flere steder disse "turn-key" systemer i brug, f.eks. GEAC på New York University Library og VTLS på National Agricultural Library i Washington og Newberry Library i Chicago. Personalet her var med til at udvikle systemerne og var glad for dem, lød det til. Det er især de store biblioteker, som benytter disse systemer. Et meget stort bibliotek som New York Public Library havde dog selv udviklet sit system.

Der er ikke længer en koncentration af viden vedrørende edb på universiteterne og i ledelsen af de store databaser. Bibliotekerne er fra at være passive medlemmer nu blevet aktive i arbejdet med at skabe egne systemer, eller de vælger frit mellem de hundrede af softwareprodukter, som findes på markedet i USA. Mikrodatamatrevolutionen, som jeg nævnte i indledningen, har også medvirket til denne demokratisering og decentralisering. De



Læsesalen på the University of Chicago Library.

store databasesammenslutninger OCLC, WLN og RLG producerede ikke "turnkey" systemer i begyndelsen af 80'erne og blev løbet over ende af edb-firmaerne. De udgør dog stadig hovedkræfterne inden for biblioteksautomatiseringen.

Amerikansk forskning af onlinekataloger

Onlinekatalogerne har nu været tilgængelige for lånerne i mere end 10 år i USA. Men de er stort set forblevet uforandrede i alle disse år. De var oprindelig i 70'erne designede til bibliotekspersonalet. Bibliotekarerne havde de første erfaringer med søgninger i de bibliografiske databaser fra Dialog. Med indførelsen af onlinekatalogerne i slutningen af 70'erne blev brugerskaren udvidet til også at omfatte lånerne. Dette havde betydning for forskningen i brugen af onlinekatalogerne. En del amerikansk forskning havde behandlet brugernes adfærd over for systemerne. Idet den beskæftigede sig med, hvor let det var at "samtale" med systemet. Kun få behandlede problemerne vedrørende søgning, nemlig om brugerne var tilfredse med resultatet af deres søgning, og hvad der evt. kunne gøres for at forbedre resultatet. Biblioteksforskningen synes i høj grad kun at supplere "common knowledge" hos bibliotekarerne. Den giver ingen indsigt i, hvorledes en onlinekatalog skal designes. I en artikel i *Journal of the American Society for Information Science* i maj 1988 ³⁾ stod bl.a. følgende: "... research has

not yet provided a satisfactory solution to the problem of interfacing between end-user and large scale databases. Despite a relatively large amount of literature about the subject, the research in information seeking and retrieving is in its infancy. It is still in an exploratory stage.... The key to the future of information systems and searching processes (and by extension, of information science and artificial intelligence from where the systems and processes are emerging) lies not in increasing sophistication of technology, but in increased understanding of human involvement with information".

Kognitionsforskningen har fået en stadig stigende betydning for andenerationen af informationssystemer, nemlig de eksperter-systemer og frontend-systemer, som er udviklet i 80'erne.

"Gateway" -løsninger og eksperter-systemer

Det er en udbredt opfattelse, at i fremtiden vil bibliotekerne kunne give adgang til alle former for bibliografiske systemer gennem et enkelt brugerfladesystem. (Spørgsmålet om brugerbetaling er lige så kontroversielt i USA som herhjemme.). Men "do it yourself" tanken er mere udbredt i Amerika, og informationsteknologien anvendes nu af ikke-eksperter, som ser på den som et redskab, og som ikke er så tolerante over for "uvenlige" og dårligt designede systemer. Man arbejder derfor for øjeblikket med forskellige systemer, som skulle gøre det muligt for ikke-professionelle brugere at betjene de bibliografiske databaser. De bibliografiske databaser er fremstillet af de store databaseværtfirmaer. Et par af de største i USA er Dialog Information Service, Inc., i Californien, BRS (Bibliographic Retrieval Service, Inc.) i staten New York og Wilsonline, som er et helt nyt firma i Bronx i New York. Søgninger i bibliografiske databaser har længe været en stadig stigende aktivitet inden for fag som biologi og fysik. Og i de sidste par år er der også kommet en række gode humanistiske databaser på markedet fra disse firmaer, således Arts and Humanities Citation Index hos BRS og Dialog, Religion Index hos Dialog og Art Index og Humanities Index hos Wilsonline. Der er derfor en del aktivitet i gang rundt omkring på universiteterne i USA for at lære forskere og studerende selv at benytte baserne til informationssøgninger. I den traditionelle forskning måtte forskere både have ideer og kunne beherske teknikken. Inden for den elektroniske informationssøgning har forskere været anset for at være passive. De skal have ideerne, men må få hjælp til det tekniske. Og det vil sikkert fremover stadig være nødvendigt med professionelle til at udføre komplekse og udtømmende



Læsesalen for "social sciences" og humaniora på Joseph Regenstein biblioteket i Chicago.

onlinesøgninger. Men de mere banale søgninger kræver ikke anden viden end den, som er nødvendig ved traditionelt forskningsarbejde.

En række "Gateway"-systemer er kommet i handelen i de sidste par år. De benyttes til at forbedre formidlingen mellem brugeren og det bibliografiske onlinesystem. Herved slipper benytteren for at lære de forskellige spidsfindigheder, som findes i de enkelte databaser. Sådanne systemer er Dialog's Knowledge Index og Sci-Mate Searcher. Sci-Mate er det bredeste, og det giver brugerassistance til en lang række amerikanske databasesystemer, såsom BRS, Dialog m.fl. og ligeledes til det franske Questel.

Disse systemer yder dog ikke nogen hjælp ved mere avancerede søgninger. De tilbyder adgang til de eksisterende baser i en forenklet form med menustyrede indgange og hjælp ved basevalget og formuleringen af søgestrategi. De eliminerer de sofistikerede søgefaciliteter og også nogle af de enkle. De er ikke i stand til at formulere de succesfulde strategier, som er den uøvedes svage punkt. Man har således endnu ikke ad kunstig vej kunnet erstatte den erfarne bibliotekar.

Ekspertsystemerne yder hjælp til brugerne på en anden måde. De optræder som eksperter på specielle områder, idet ekspertviden kan samles fra menneskelige eksperter og andre kilder og lægges ind i systemerne. Sådanne systemer er i udvikling for øjeblikket mange steder i verden. Også i Danmark arbejdes der med projekter inden for området.

I USA er der mange forsøg i gang på at designe ekspertsystemer til hjælp ved onlinesøgninger. CITE fra NLM (National Library of Medicine) tilbyder faciliteter med naturligt sprog og emneord. Et andet system, CANSEARC,

er udviklet til hjælp ved søgning af cancerlitteratur med teknik hentet fra kunstig intelligensforskning.

Der er mange spændende initiativer i gang inden for udviklingen af informationsteknologien. Og forskningen vedrørende kunstig intelligens hører til det mest interessante. Men litteraturen om emnet er vanskeligt tilgængelig. Den giver næsten ingen hjælp til løsning af de nære problemer, såsom hvorledes man baner vejen til vellykkede søgninger for brugerne, og hvorledes systemerne skal designes, så de kan udnyttes optimalt. Der er mange opgaver, som venter på at blive løst, inden for denne videnskab.

Noter

1. Imo report 89/2. Commission of the European Communities. Luxembourg. March 1989.
2. CD-ROM International nr. 10, september 1988, s. 3.
3. A Study of Information Seeking and Retrieving, JASIS, 39, 1988, s. 161 f.

Litteraturhenvisninger

- Charles R. Hilreth: Library Automation in North America, München. 1987.
The Impact of online Catalogs. ed. Joseph R. Matthews, New York. 1986.
Online 88, Conference Proceedings, Oktober 10, 11, 12. New York.
Ulla Retlev: Intelligente kommunikationbetragtninger over "Gateway"-løsninger m.v. DF-Revy nr. 7, 1988. s. 186-189.
Rethinking the Library in the Information Age-Issues in Library Research: Proposals for the 1990's. udg. af Office of Educational Research and Improvement, Washington. 1988.
Using Online Catalogs: a nationwide Survey. ed. Joseph R. Matthews m.fl. New York. 1983.