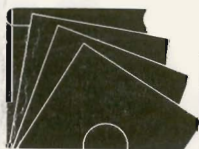


# h

# UMANISTISKE

# D A T A



NAVS EDB-SENTER  
FOR HUMANISTISK FORSKNING  
NORWEGIAN COMPUTING CENTRE  
FOR THE HUMANITIES

# 1-91

## Fra redaktørens side:

Redaksjonen spør seg med jevne mellomrom hva slags tidsskrift *Humanistiske Data* bør være. Det skjer mye interessant i kryssingsfeltene mellom de humanistiske fagene og en datateknologi i rask utvikling, og vi makter ikke å gi fullgod dekning av alt som skjer.

I fjor høst spurte vi også leserne hva de mente om HD. Vi sendte ut et spørreskjema til alle enkeltabonnenter og ba om reaksjoner på tidsskriftet. Spørreundersøkelsen får fyldigere omtale annetsteds i bladet, men hovedinntrykket er at leserne representerer mange fag og ønsker mer stoff om *sine* fagfelt. Vi kan ikke love å bringe stoff fra alle de fagområder som ble etterlyst, men blant de ønskene vi kan oppfylle, er å presentere Senterets tjenester. Dette nummer av HD bringer oversikter over den programvaren vi har kompetanse på og over Senterets databaser. Men det generelle problemet må vi antakelig leve med: Edb i humaniora er i ferd med å bli et omfattende fagfelt med et større behov for spesifikke applikasjoner enn for generelle løsninger.

**HDs fremtid:** HD kommer til å bli lagt om fra og med neste nummer. Bladet vil få A4-format som gjør det enklere å behandle billedmateriale, og vi tar sikte på 4 utgivelser pr år for å få stoffet litt raskere ut til leserne.

En annen grunn til at HD kommer til å endre karakter, er at tidsskriftets mangeårige redaktør, Kristin Natvig, har permisjon og søker lykken annetsteds. Etter vedtak i RHF vil stillingen som informasjonskonsulent ved Senteret være ubesatt i 1991, og informasjonsarbeidet er derfor fordelt på flere medarbeidere ved Senteret, som har bidratt med artikler, meldinger og referater. Dette er en ny arbeidsform, og er ikke bare en nødløsning: vi tror at vi på denne måten kan fange opp mer av utviklingen innen edb og humaniora. Det er når det gjelder å få inn større faglige artikler, at vi sterkest føler savnet av en som har øvelse i først å få folk til å love å skrive en artikkel og deretter ikke lar vedkommende i fred til artikkelen er levert... Men vi tar selvsagt imot artikler til trykking som før.

Som omtalt annetsteds i HD, har Senteret opprettet en elektronisk posttjeneste. Der vil vi bringe stoff om Senteret, og artikler som eventuelt ikke får plass i HD, vil vi reklamere for i e-mail-listene, slik at den som ønsker en artikkel kan hente den ut elektronisk.

Vi ønsker å skape et leseverdig tidsskrift – som kan bidra til å styrke edb-bruken i humanistiske fag. Det kan vi enklest oppnå hvis leserne gir oss reaksjoner på det vi foretar oss.

Herved overlates det siste HD i gammel form til lesernes nåde.

---

***HUMANISTISKE DATA***

---

***1-91***

---

---

***NAVF's EDB-SENTER FOR  
HUMANISTISK FORSKNING***

***The Norwegian Computing Centre for the  
Humanities***

---

---

## **HUMANISTISKE DATA**

---

utgis av NAVFs edb-senter for humanistisk forskning, Harald Hårfagresgt. 31, Boks 53 – Universitetet, N-5027 Bergen. Tlf: + 47 (0)5 212954/55/56. Fax: + 47 (0)5 322656. E-post: FAFKN@NOBERGEN.EARN

**Redaksjonsråd:** Jostein H. Hauge, Knut Hofland, Anne Lindebjerg, Kjell Morland (red.), Espen S. Ore, Øystein Reigem.

**Abonnement:** Abonnementspriser for 1991 er kr 70,- for enkeltpersoner, kr 140,- for institusjoner/firmaer (3 nr. pr. år). Gratis for abonnenter utenfor Norden.

## **OM DETTE NUMMER**

---

**Medarbeidere fra Senteret:** Ivar J. Gubberud, Arne Hallheim, Signe Marie Sanne, Eli Schilbred, Unni Sæther og Per Vestbøstad.

**Redaksjonen avsluttet 12. april.**

**Forsidebilde:** Bildet er tatt i Kristiania (Professor Dahls gate) av en ukjent fotograf rundt århundreskiftet. En negativ glassplate er scannet inn med en oppløsning på 300 punkter pr tomme. Deretter ble det invertert og gjort litt skarpere med programmet PhotoShop. Tilslutt ble det skrevet ut på Apple laserskriver med 53 rasterlinjer pr tomme. Se forøvrig artikkelen *Litt om bilder og datamaskiner og lagringsplass* i dette nummer av HD.

**Sats:** Xerox Ventura Publisher.

**Trykk:** Bergen Trykk as.

## **NAVFs EDB-SENTER FOR HUMANISTISK FORSKNING**

---

ble opprettet av Norges allmennvitenskapelige forskningsråd i 1972. Senteret skal arbeide på nasjonal basis for bruk av edb i de humanistiske fagene.

Sentrale oppgaver er edb-tjenester (veiledning og betalte oppdrag) og program- og metodeutvikling. Senteret holder kurs, seminar og konferanser om bruk av edb i humanistiske fag. Informasjonstjenestene omfatter også en rapportserie, årsmelding og elektronisk informasjonsformidling.

Senteret er sekretariat og operativt edb-organ for International Computer Archive of Modern English (ICAME) og utgir tidsskriftet *ICAME Journal*. Senteret driver også en informasjonstjeneste for humanistiske forskningsprosjekter og en offentlig tilgjengelig forskningsdatabase.

## **INFORMATION ON HUMANISTISKE DATA**

---

Humanistiske Data is published 3 times a year by the Norwegian Computing Centre for the Humanities. Editor: Kjell Morland. The journal is free of charge to subscribers outside Norway, Denmark, and Sweden. We print short contributions of all kinds, and welcome lengthy articles for consideration.



# INNHold

## ARTIKLER

|   |    |
|---|----|
| Litt om bilder og datamaskiner og lagringsplass. <i>Espen S. Ore</i> .....                | 5  |
| FotoMac videreutvikles. <i>Espen S. Ore</i> .....   | 10 |
| Dokumentasjonsprosjektet ved HF-fakultetet,<br>UiO. <i>Christian-Emil Smith Ore</i> ..... | 13 |
| STEMMA, a Robust Noun Stem Generator for Finnish.<br><i>Kimmo Kettunen</i> .....          | 26 |
| Multimedielaboratoriet ved AVH, UNIT. <i>Julie Feilberg</i> .....                         | 32 |
| Distribusjon av informasjon fra Senteret via elektroniske nett. ....                      | 36 |

## ANMELDELSER

|   |     |
|---|-----|
| PAST MASTERS. <i>Ralph Jewell</i> .....   | 40  |
| Mozart og Beethoven i HyperCard. <i>Espen S. Ore</i> .....  | 43  |
| Elastiske media i Bergen. <i>Øystein Reigem</i> .....   | 52  |
| Seminar om informasjonsteknologi i de humanistiske fag –<br>tendenser og utviklingsbehov. ....  | 56  |
| Seminar om lenking av nominative data. <i>Øystein Reigem</i> .....  | 63  |
| ONLINE 90. <i>Ivar J. Gubberud og Unni Sæther</i> .....   | 77  |
| Internasjonalt seminar om forskningsdokumentasjon.<br><i>Ivar J. Gubberud</i> .....   | 81  |
| Sentral for informasjon om forskningsprosjekter. <i>Ivar J. Gubberud</i> ..   | 85  |
| "Elektroniske bøker – multimedia oppslagsverk"<br>Konferanserapport. <i>Kjell Morland</i> .....   | 88  |
| A Taste of the Electronic Library – "Electronic Books – Multimedia<br>Reference Works": Conference in Bergen 21–23 Nov. 1990.<br><i>Espen Aarseth</i> ..... | 93  |
| Komitéinnstilling om ny teknologi på HF-fakultetet ved UiB. ....  | 99  |
| Diskusjonsseminar – interaktiv video i språkopplæring.<br><i>Signe Marie Sanne</i> .....  | 101 |
| Leserundersøkelse om Humanistiske Data. ....  | 103 |

## MELDINGER

---

|  |     |
|--|-----|
| Konferanser. ....  | 106 |
| Konferansekalender. ....   | 112 |
| ACH/ALLC '91. <i>Per Vestbøstad/Jostein H. Hauge</i> .....   | 119 |
| Nytt nummer av OLUFF. ....   | 133 |
| CHArt – forening og tidsskrift for kunsthistorie. <i>Øystein Reigem</i> ...                            | 136 |
| Datastøtta forskning i germansk språk og skandinavisk<br>middelalderspråk. <i>Per Vestbøstad</i> ..... | 138 |
| Artikler/rapporter i Humanistiske Data 1990. ....  | 141 |
| Nytt i biblioteket. ....   | 145 |
| Årsrapporter fra SIF og HUM-DOK. ....  | 146 |
| Program i bruk ved Senteret. <i>Arne Hallheim</i> .....  | 147 |
| TACT. Et program for indeksering og søking i<br>tekst. <i>Knut Hofland</i> .....                       | 152 |
| Konferanserapport/Proceedings: Electronic Books – Multimedia<br>Reference Works. ....                  | 159 |
| English Computer Corpora: Material available through ICAME. ....                                       | 162 |

## SUMMARIES

---

165

# LITT OM BILDER OG DATAMASKINER OG LAGRINGSPLASS

---

---

*Espen S. Ore*

Når vi legger inn bilder på en datamaskin, er det flere hensyn som må tas både ved innleggingen og ved etterbehandlingen. Først og fremst er det viktig å spørre seg hva bildet skal brukes til. Deretter er spørsmålet i hvilken grad man i tillegg skal ta hensyn til at andre kanskje ønsker å bruke bildet til noe helt annet. I billedsamlinger av den typen FotoMac er beregnet på vil det være temmelig korttenkt å lagre bildene slik at de tilfredsstiller øyeblikkets behov, men må legges inn på nytt for å dekke fremtidens. I det følgende vil jeg ta utgangspunkt i scanning av gråtonebilder (svart/hvitt fotografier) som enten skal vises på skjerm eller trykkes, men det meste av det som her blir sagt, gjelder også for f.eks. digitalisering av videobilder i farge.<sup>1</sup>

## **BILDER TAR STOR Plass**

---

Hvor stor lagerplass krever et bilde? Hvilke faktorer er bestemmende for det? Når et bilde digitaliseres, gjøres det om til en samling punkter, la oss si som et gigantisk sjakkbrett. Hvert punkt skal inneholde informasjon om gråtoneverdi (eller farge) for punktet. Denne informasjonen lagres som et tall. Dersom vi i ett punkt kan velge mellom verdier for 16 tall (0 - 15) trenger vi 4 bit til å lagre gråtoneinformasjonen i punktet. På den annen side vil vi bare kunne skjelne mellom 16 forskjellige gråtoner, og det gjør at overgangen mellom de forskjellige tonene blir skarpt synlige, vi får altså ikke den myke overgangen mellom gråtoner som vi er vant til i et fotografi. Hvis vi derimot setter av 8

1 Se omtale av *Tryllefløyten* i dette nummer av HD for et eksempel på et bilde som er digitalisert fra video.

bit pr punkt, kan vi velge mellom 256 forskjellige gråtoner. I praksis vil vi ikke ha noen synlig glede et høyere antall gråtoner enn dette. (Det er tvilsomt om man normalt vil se forskjell på et bilde med 64 gråtoner og ett med 256.)

Men det er ikke bare lavt antall gråtoner som kan gi skarpe og synlige overganger. Antallet punkter pr lengdeenhet (langs horisontal og vertikal akse) er også vesentlig. En normalt god dataskjerm har idag en oppløsning på ca 70 punkter pr tomme (eller snaut 30 punkter pr centimeter – som et resultat av kombinasjonen av at mesteparten av faglitteraturen idag er skrevet på engelsk og undertegnedes iboende latskap, vil jeg i det følgende bruke punkter pr tomme som måleenhet). Det vil si at et bilde på 8 x 11 tommer (nesten en A4-side) med 256 gråtoner krever 8 tommer x 70 punkter x 11 tommer x 70 punkter x 8 bit pr punkt. Dette blir 3.449.600 bit eller 431.200 byte. Dette høres mye ut, men hvis vi fordobler oppløsningen til 140 punkter pr tomme, må vi firedoble antallet byte til drøyt 1,6 megabyte. Man får *ikke* plass til mange slike bilder på en 40 MB harddisk!

### ***MEN HVA MED TRYKKKVALITETEN?***

---

Mens skjermen altså har ca 70 punkter pr tomme, kan en fotosetter på sin side ha over 2000. Dette virker umiddelbart litt håpløst, men det er en vesentlig forskjell mellom de to: mens hvert av skjermpunktene kan ha en av 256 gråtoner eller for den saks skyld en av 17 millioner farger (dette forutsetter selvfølgelig at man bruker dertil egnet datautstyr), kan et fotosetterpunkt bare være enten svart eller hvitt. For å gjengi gråtoner bruker fotosetteren rasterpunkter som er bygget opp av grupper av fotosetterpunkter. Hvis rasterpunktet er bygget opp av 16 x 16 fotosetterpunkter, kan hvert rasterpunkt gjengi en av 256 (egentlig 257) gråtoner. Hvis vi nå tar utgangspunkt i skjermens 70 punkter pr tomme og ganger det med 16, får vi 1120. En skjermoppløsning på 70 tilsvarer altså 1120 punkter pr tomme på en fotosetter.

Ut fra det ovenstående kunne man kanskje tro at finere oppløsning på fotosetteren og finere oppløsning ved innscanning alltid ville føre til finere bilder på trykk. Dette er ikke tilfelle. Trykkmetode og papirkvalitet spiller også inn. Avispapir egner seg f.eks. ikke til høyere rastertetthet enn drøyt 80 linjer pr tomme. Høyeste kvalitet glanset papir kan tåle opp til 175 linjer pr tomme.

### ***OPPLØSNING VED INNSCANNING***

---

Ut fra det som er skrevet ovenfor, skulle det være klart at man ikke uten videre har noen glede av å scanne et bilde med en oppløsning





*Fig. 1 a. Scannet med  
75 punkter pr tomme og  
skrevet ut med 75 linjer  
pr tomme*

*Fig. 1 b. Scannet med  
150 punkter pr tomme  
og skrevet ut med 75  
linjer pr tomme*



*Fig. 1 c. Scannet med  
300 punkter pr tomme  
og skrevet ut med 75  
linjer pr tomme*

på 10.000 punkter pr tomme. Men hvilken oppløsning bør brukes? Hvis vi har tenkt å forstørre bildet etter innsanning, må vi ta hensyn til det. Hvis vi vil beholde størrelsen som den er, må vi først avgjøre om bildet bare skal vises på skjerm eller om det også skal trykkes. Er vi bare interessert i å bruke det på skjermen, er 70 – 75 punkter pr tomme utmerket. (De forskjellige scannere har forskjellig nedre grense for oppløsning.) Ønsker vi å trykke bildet, blir det mer komplisert. Hvis vi tar utgangspunkt i en ønsket trykkkvalitet på 75 linjer pr tomme, får vi ikke nødvendigvis det beste resultatet ved å scanne med 75 punkter pr tomme. I analog/digital konvertering (som scanning er et særtilfelle av) finnes det en gylden regel som litt forenklet sier at vi må konvertere i to ganger den oppløsning vi har tenkt å bruke. (For spesialister: det er slik Nyquists samplingsteorem virker ved billedscanning/rasterisering.) Det bildet vi ønsker å trykke med 75 linjer pr tomme, bør altså scannes med 150 punkter pr tomme. Et eksempel på dette er vist i fig. 1a-b-c (se forrige side).

## MANIPULERING AV BILDER

Bildet i fig. 1 kommer opprinnelig fra en negativ glassplate og er tatt i Kristiania i begynnelsen av dette århundret. Fotografen er ukjent. Glassplaten er lagt inn i en vanlig scanner med høyeste oppløsning på 300 punkter pr tomme og med evne til å skjelne 256 gråtoner. Etter at bildet var scannet, ble det tatt inn i et elektronisk mørkeromsprogram.

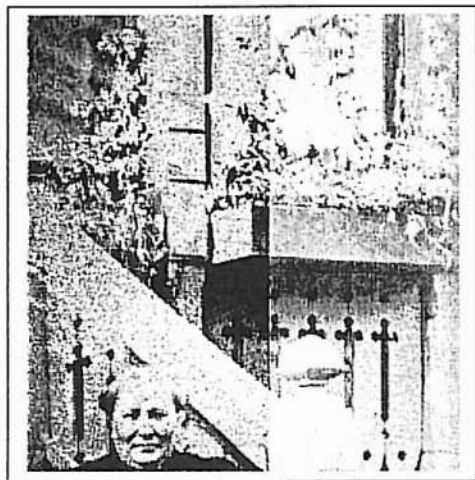


Fig. 2. På venstre side av bildet er flere gråtoner hentet frem med elektroniske filtre.

Der ble det først invertert (gjort om fra negativ til positiv), og deretter ble kontrasten øket ved at gråtonene i bildet (som klumpet seg ganske tett sammen) ble trukket ut over hele spekteret av 256 gråtoner. Det ble ikke gjort noe forsøk på å oppnå verdens beste resultat, bare en forbedring ved bruk av enkle midler. Programmer av denne typen som er brukt her, har imidlertid langt flere muligheter. I fig. 2 er et utsnitt av bildet fra fig. 1 forstørret, og gråtonene i halve bildet (venstre side) er på nytt trukket ut over hele skalaen.

Men disse programmene gir oss ikke bare muligheten til å forbedre de innskannede bildene. Vi kan også skape helt nye! Å skjære vekk deler av bilder og sette inn nye har vært utført i lang tid. Et noe ubehjelpelig eksempel på dette, som tok undertegnede temmelig nøyaktig 30 sekunder å lage, er vist i fig. 3. Dersom man tar seg tid til å gjøre arbeidet grundig, er det ikke vanskelig å forfalske bilder av hjertens lyst. Dette siste gir grunn til ettertanke når det gjelder den rolle fotografier kan ha som dokumentasjon eller bevismateriale i fremtiden.



Fig. 3. Enkel montasje.

## LITTERATUR

---

- Babčický, K. et. al. 1989, *Sort-hvitt bildebehandling med standard datateknologi*, Prosjektrapport 2-89, Grafisk Institutt, Oslo.
- Germundson, L og Olsson, B. 1990, *Bilden i tryck*, Halmstad.

# FOTOMAC VIDEREUTVIKLES

---

---

*Espen S. Ore*

I en tidligere artikkel beskrev undertegnede kort en prototyp av et fotolagringsystem for Macintosh. Prototypen var laget i Hypercard v. 1.2.x og brukte eksternt tilgjengelige funksjoner til å vise gråtonebilder på skjermen. Fra denne prototypen ble det sommeren 1990 utviklet en versjon (populært kalt versjon 0.3b) i samarbeid med *Roger Erlandsen* (Sekretariatet for fotoregistrering) og *Espen Aarseth* (Universitetet i Bergen). Denne versjonen hadde mer avanserte søkemuligheter enn den første, og den innførte et skille mellom søkeskjemaene (det brukeren ser) og dataene slik de faktisk er lagret. Dette skillet har to formål: det skal være mulig å definere forskjellige skjema eller vinduer mot dataene tilpasset de forskjellige brukeres behov, og brukeren skal ikke komme i direkte kontakt med data og lagringsstruktur – med de farer for uheldige endringer i databasen dette kunne medføre.

## FOTOMAC VERSJON 1.0

---

I løpet av høsten 1990 ble den siste versjonen av FotoMac demonstrert flere steder, og ved Universitetsbiblioteket i Oslo bestemte man seg for å prøve systemet til å lagre en større samling bilder av og med Fridtjof Nansen (se neste side). Med bakgrunn i dette, og siden flere andre institusjoner også viste interesse for et slikt system, ble det ved Senteret bestemt at det i løpet av 1. kvartal 1991 skulle utvikles en versjon 1.0 av FotoMac. (Vi bruker en versjonsnummerkonvensjon der 0.x betyr prototyper/utviklingsversjoner mens 1.x og høyere antyder at versjonene skal være funksjonelle og driftsikre – i den grad dette kan oppnås.)

FotoMac 1.0 blir utviklet i HyperCard 2.0. Dette gir klare fordeler i forhold til de tidligere versjonene:

- Kort kan være vilkårlig store (opp til 1280 x 1280 punkter)
- Mekanismen for å vise gråtonebilder (eller fargebilder) følger med HyperCard og er mer robust enn den som ble brukt i tidligere versjon av FotoMac



- Forhåndsdefinerte utskriftskjemaer brukes
- Kontroll over menyer

Blant de planlagte endringene i FotoMac v. 1 i forhold til tidligere versjoner er:

- Absolutt adskillelse av data og brukervindu

Dette gjelder ikke bare et skille mellom søkeskjema og data som tidligere, men også at det brukes et eget skjema for datainnskrivning og endring. På denne måten får vi enda bedre datasikkerhet, og det blir mulig å legge verifiseringsrutiner mm. mellom innskrivingen og den endelige oppdateringen av databasen. Dette betyr også at det strengt tatt ikke er nødvendig at data er lagret i HyperCard: vi kan tenke oss en større database som fysisk er lagret på f.eks. en UNIX-server, og der brukeren bare ser de samme Hypercard-baserte skjemaene som brukes nå.

- Muligheter for bruk av logiske operatører og kobling av søk

I den siste HyperCard 1.2.x-baserte versjonen av FotoMac var det laget en mulighet for å søke gjennom QBE (Query By Example). Det var også mulig å kombinere funnlister fra slike søk. Denne kombinasjonen ga samme effekt som et logisk *eller* mellom søkene. I versjon 1.0 skal det være mulig å bruke standardoperatørene *og*, *ikke* og *eller*. Det skal også lages en mulighet for større enn/mindre enn-søk, noe som f.eks. er svært nyttig når dateringen av et bilde er et vesentlig søkekriterium.

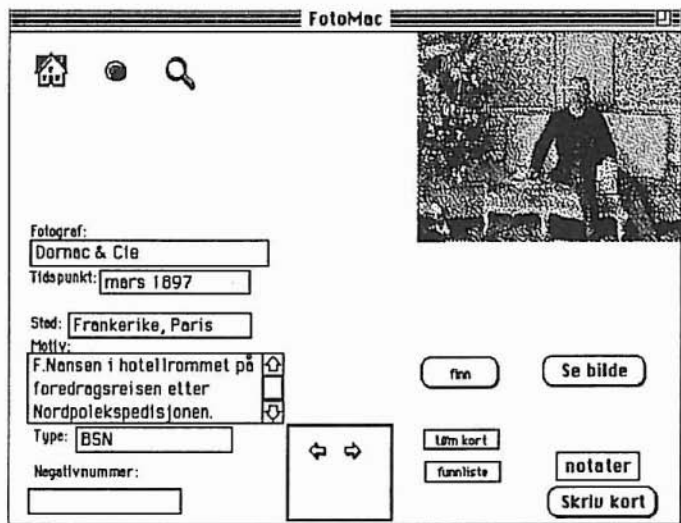


Fig. 1. Det "enkle" søkekortet. FotoMac inneholder også et mer avansert søkekort. (Se Fig. 2.)

## NANSENSAMLINGEN OG FOTOMAC

Universitetsbiblioteket i Oslo har en samling på ca. 8000 bilder av og med Fridtjof Nansen. De 4000 beste av disse bildene og referansedata til dem ønsker man å lagre digitalt. Siden høsten 1990 har det derfor vært et samarbeid mellom Senteret og UBO. Ved UBO har man allerede begynt å lagre data og bilder i den nåværende versjonen av FotoMac. Det er først, som et forsøk, lagret data om ca. 400 bilder, og ut fra erfaringene med dette forsøket er hovedregistreringen nå i gang.

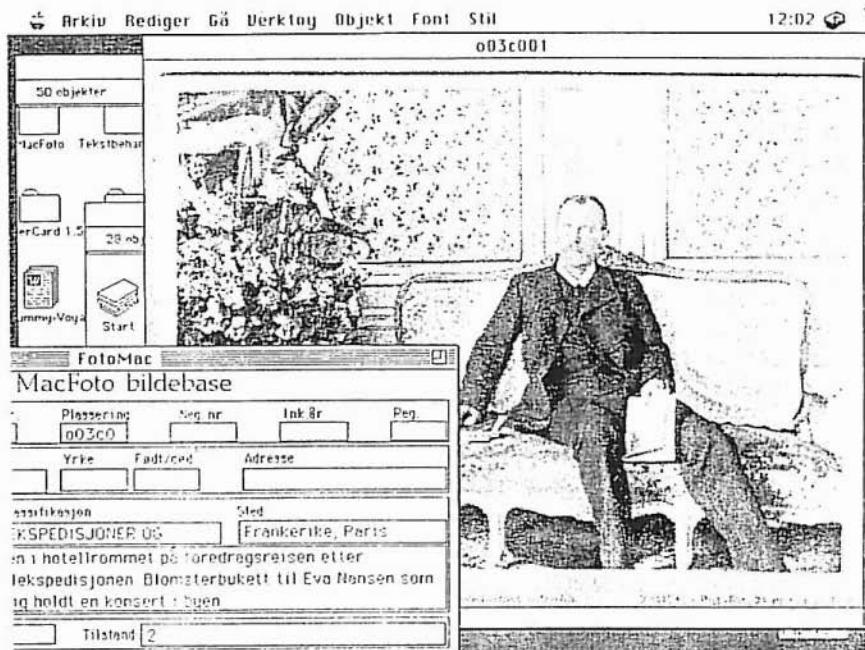


Fig. 2. Gråtonebildet vises. I nederste venstre hjørne ser vi litt av det "avanserte" søkekortet. (Kvaliteten på gråtonebildet er preget av at det er en skjermdump. Det ser bedre ut på skjermen.)

# *DOKUMENTASJONSPROSJEKTET VED HF-FAKULTETET, UiO*

---

---

*Christian-Emil Smith Ore*

## *INNLEDNING*

---

Det historisk-filosofiske fakultetet ved Universitetet i Oslo har i disse dager gitt klarsignal til å gjennomføre det første året av et seksårig dokumentasjonsprosjekt. Prosjektet har det ambisøse mål å bygge opp databaser over de mange store fagarkivene og samlingene som i dag finnes ved fakultetet. Det må her understrekes at det ikke er tale om bare å lage elektroniske kataloger over materialet, men derimot å overføre selve arkivmaterialet til moderne, datamaskinstøttede medier. Det endelige siktemål er et samlet elektronisk dokumentasjonssystem der alle interesserte lett kan få tilgang til den delen av vår kultur- og nasjonal arv som disse arkivene representerer.

Dokumentasjonsprosjektet fikk sin spede begynnelse for noe over et år siden, da det ble vedtatt i strategisk plan for HF-fakultetet at man skulle søke å bedre vilkårene for arkiver og samlinger ved blant annet å opprette offentlig tilgjengelige databaser. Dette nedfelte seg som konkret tiltak i virksomhetsplanen for 1990, og på initiativ fra daværende prodekanus Bjarne Hodne ble det i april satt i gang et tre måneder langt forprosjekt ledet av artikkelforfatteren for å få en samlet oversikt over arkiver og registre ved HF-fakultetets tradisjonelle samlingsavdelinger med vekt på de ulike arkivene og registrenes karakter, fysiske utforming og på hvilke datamengder de representerer.<sup>1</sup> Det forberedende arbeidet med prosjektet ble tatt opp igjen i oktober 1990. I begynnelsen av februar 1991 vedtok så fakultetsrådet at man i 1991 skulle gjennomføre

1 Rapporten *Dokumentasjonsprosjektet ved Det historisk-filosofiske fakultet, Universitetet i Oslo* gir en utførlig beskrivelse av prosjektet og fås ved henvendelse til Chr.-E. Ore, Inst. for lingv. og filosofi, Boks 1102 Blindern, 0317 Oslo.

det første året av den seksårige planen for prosjektet.

Prosjektet er stort i den forstand at bare dataregistreringen vil kreve mange hundre årsverk. Prosjektet er også faglig svært omfattende idet vi blant de aktuelle samlingsavdelingene finner Oldsaksamlingen, Norsk folkemusikksamling og de leksikografiske avdelingene. Prosjektet burde således være av interesse for mange ulike grupper innen humaniora generelt og humanistisk databehandling spesielt.

## **2 ARKIVENE**

---

De tradisjonelle samlingsavdelingene ved HF-fakultetet er Oldsaksamlingen, Folkemusikksamlingen, avdelingene for etnologi og folkloristikk samt leksikografi-, målføre- og navnegransking-avdelingene. Disse avdelingene har arkiver og registre av forskjellige størrelser, fra 17 000 dokumenter/kort ved Oldsaksamlingens register over faste fornminner til 3,5 millioner dokumenter/kort i bokmålsarkivet ved Leksikografisk avdeling, Institutt for nordistikk og litteraturvitenskap. Samlet er det tale om 14 millioner kort og dokumenter. Arkivene er ikke ensartede. Noen er rene kataloger, andre er registre over ordforekomster i løpende tekst (konkordanser), og til slutt har vi rene dokumentsamlinger.

### ***2.1 Leksikografi- og målføresamlingene ved Institutt for nordistikk og litteratur***

---

Avdeling for leksikografi (tidl. Leksikografisk institutt) består av tre underavdelinger: gammelnorsk, nynorsk og bokmål. Alle de tre avdelingene ble i sin tid grunnlagt med det for øye å lage ordbøker over de respektive sider av det norske språk. Hver avdeling har i dag store og svært omfattende vitenskapelige ordsamlinger. Disse vitenskapelige ordsamlingene ved Universitetet i Oslo er de eneste av sitt slag i Norge og er derfor av nasjonal betydning. Den gammelnorske samlingen representerer en betydelig del av vår nasjonal arv fra førreformatorisk tid. Samlingene ved bokmåls- og nynorskavdelingene dokumenterer betydning og bruk av ord i moderne norsk språk, men omfatter også store mengder informasjon om det norske språk i hele tiden etter reformasjonen. Nynorskavdelingens samlinger omfatter ordtilfanget i norske dialekter og gir således uvurderlig informasjon om sammenhengen mellom talespråk og skriftspråk i Norge. Samlingene utbygges og kompletteres fortløpende.

Samlingene består av såkalte ordbokssedler. En slik seddel inneholder oppslagsord i normalform, grammatiske opplysninger, opplysninger om kilde og kontekst med ordforekomst markert. Gammelnorskavdelingens



sedler er (naturlig nok) basert på skriftlige kilder (Fig. 1.). Det samme gjelder for bokmålsavdelingen. For å unngå unødig inntasting og lagring av ordenes kontekster, er det derfor meningen å opprette elektroniske tekstsamlinger i kombinasjon med relasjonsdatabaser. Det vil dermed være mulig å dra nytte av både fritekstsøkingsens fordeler og de strukturerte databasers fordeler.

sigla

Segle

segla, s, eit

- Hvorus går for utan ovanpåstein og segle.

Utydning: Seglet er festa på spinnulen, og seest lek

lak i ovanpåsteinen og driv han.

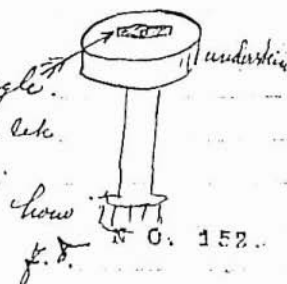


Fig. 1.

Nynorsksamlingen er ikke basert på et avgrenset tekstkorpus. Ordforekomstene er hentet fra både skriftlige og muntlige kilder, og er dels registrert av fagfolk, dels av leginformanter som gir opplysninger om sitt eget målføre (Fig. 2.). Det er derfor ikke mulig utelukkende å

tapp m

tapp, ein (tapp'on, tapp'ar, tapp'ans), kisel, svirre: slå tapp', eit spel, ein lek. Ein slær kiselu med ei reim som er fest til et svart's træ, ei litt svarva stong med hol i brøste enden til reim, so kiselu svirrar rundt. Det gjeld ved stendige slag med reimi å halda kiselu svirrande. (Jampfer setelen sl-er l's træ!).

110 727.

Fig. 2.

benytte en database med løpende tekst og derved komprimere datamengden. De uensartede innsamlingsmetodene gjør også selve sedlene svært forskjellige. De er både håndskrevne og maskinskrevne, og kan også inneholde utklipp fra skriftlige kilder. Sedlene har bare en grov felles rubrering, men inneholder dog de samme basisopplysninger: normalisert oppslagsform av ordet, målføre- eller skriftform og kilde. De fleste sedlene har dessuten en rekke andre og varierende opplysninger, f.eks. om betydning, grammatiske former, fraseologi, kontekst, geografisk utbredelse og brukstid. Noen sedler er illustrert (Fig. 3.). Innskrivningen av dette arkivet, som består av 3,2 millioner sedler, vil derfor være en utfordrende oppgave.

N 35

fullgöðr, adj  
full gott msn

(8) Sifka segir at konungr (9) vill þat æigi lata, oc hann  
sealfr vill þat hana (10) ef han skal fara. oc visar honum  
til þess skips (11) er verst var oc telr þo vera full gott  
til æigi (12) lengri ferðar. oc nu vill reginballd æigi fara  
næma (13) hann havi gott skip.                      þiðr. II 163;

Fig. 3.

Norsk Målførearkiv har et seddelarkiv som tildels inneholder de samme sedlene som nynorskarkivet. Men i tillegg kommer de håndskrevne Storm-listene. Disse utgjør resultat-tabellen av en stor undersøkelse av uttalen av 1150 utvalgte ord på ca. 900 steder i landet. Dataregistreringen av dette materialet vil kreve en del oppfinnsomhet. I tillegg til det skriftlige materialet har målførearkivet en omfattende lydbandsamling som danner basis for mye av det skriftlige materialet. I dag er det en selvsagt ting å kunne lagre lyd ved hjelp av en datamaskin. Det er altså mulig å lagre dialektprøver både som lyd og som tekst. Ved å legge opp de aktuelle delene av arkivet som en tekst/lyd-database vil en dermed få mulighet til direkte å kunne sammenlikne transkripsjon og taleprøve.

## ***2.2 Institutt for nordistikk og litteraturvitenskap, Avdeling for navnegransking***

---

Avdeling for navnegransking har store samlinger over norske stedsnavn. I samlingene inngår opplysninger om lokalisering, bruk og betydning. Avdelingens samlinger består i hovedsak av seddelarkiver.

Herredsregistret er det største arkivet. Dette består av ca. 450 000 sedler og ble grunnlagt i 30-årene. Arkivet er i hovedsak sortert alfabetisk under de gamle herredene, men består også av en rekke småsamlinger. Sedlene er til dels håndskrevne og ikke ført på en ensartet måte.

Bostedsnavnregistret består av ca. 90 000 sedler og inneholder opplysninger om bostedsnavn fra før 1910. Registret er basert på matrikkelen 1903–08 og er ordnet topografisk etter denne. Det er (vanligvis) ett kort for hvert gårdnummer. Dette kortet inneholder en liste over bruk og husmannsplasser for gårdnummeret. Deretter følger (i prinsippet) ett kort for hvert bruk og husmannsplass.

I tillegg finnes registre basert på de foregående, et register over seternavn, samt en rekke mindre registre. Avdelingen har også et lydbåndarkiv (jvf. det som er sagt ovenfor om målføreakivet).

## ***2.3 Institutt for arkeologi, kunsthistorie og numismatikk (Oldsaksamlingen)***

---

Avdelingen for faste fornminner registrerer faste fornminner i terrenget og har nå ca. 17 000 registreringer. Avdelingen har drevet dataregistrering av sitt arkiv siden 1985 og tok denne sommeren i bruk fritekstsystemet SIFT.

Avdelingen for løse fornminner har ansvaret for Historisk museums samlinger. Her er det flere registre. Det største er museets hovedkatalog. Hele katalogen er anslagsvis på 6000 sider maskinskrevet/trykt tekst. Katalogen kan enten leses optisk eller tastes inn. Den første løsningen er mest tiltalende og tanken er da å lagre artiklene som dokumenter i fritekstsystemet SIFT. Det er i forprosjektet gjort prøver med ulike programmer for optisk tegngjenkjenning. Resultatet er tildels oppløftende, men en grundig korrektur er nødvendig. Vi arbeider for tiden med enkle filterprogrammer for en automatisk grovkorreksjon.

Fotoarkivet ved Historisk museum er på ca. 30 000 negativer, med tilhørende register. Dette arkivet vil kunne avfotograferes og lagres som en bildedatabase i tilknytning til hovedkatalogen. Dette vil gi en unik mulighet til å koble tekstlig beskrivelse med bilde. Dette er nyttig i det faglige arbeid, og åpner store muligheter i undervisning ved universiteter og særlig i skolen.

Myntkabinettet har i dag ca. 200 000 mynter med et mer eller mindre komplett register. I tillegg til å få skrevet inn hele dette registret bør også myntene fotograferes slik at det blir mulig å lage en kombinert bilde- og register-database. En slik kobling av et register og bilder av myntene vil faktisk gi en komplett myntkatalog over kabinettets samling. Trykte verker kan enkelt produseres fra en slik base ved hjelp av setteprogrammer ("desktop publishing") som i dag er standard hyllevare.

Runearkivet ble for omkring hundre år siden skilt ut som et eget arkiv. Arkivet er nasjonalt og inneholder opplysninger om samtlige kjente (ca. 1600) runeinnskrifter i Norge.

Oldsakssamlingen har en rekke større og mindre arkiver i tillegg. Det vil selvfølgelig være interessant også å legge disse opp som databaser. Men dette vil komme etter de andre arkivene.

#### ***2.4 Institutt for kultur- og samfunnsfag, Avdelingene for etnologisk og folkloristikk***

---

Arkivene til avdelingene for folkloristikk og etnologi er i en slik stand at det kreves stor faglig ekspertise for å gjøre dem egnet for innskrivning.

Når det gjelder de eldre originalmanuskriptene, er disse ikke velegnet til innskrivning, idet de er vanskelige å tyde (Fig. 3.). Dog skulle ikke dette by på større problemer enn avskriften av kirkebøker, som er gjort ved Registreringssentralen for historiske data i Troms.

#### ***2.5 Norsk folkemusikksamling***

---

Folkemusikksamlingen har kommet svært langt i arbeidet med å data-registrere sine registre. Folkemusikksamlingen deltar i et prosjekt for katalogstandard for databehandling av folkemusikksamlingene i Norge. Selve katalogdataene ved samlingen i Oslo er på ca. 40 000 kort, hvorav mesteparten er registrert.

### ***3 HVA ER HENSIKTEN MED PROSJEKTET?***

---

Materialet som skal dataføres, representerer både direkte og indirekte informasjon. Informasjonen i de leksikografiske dataene er direkte i den forstand at den ikke henviser til gjenstander, men er interessant i kraft av sin tekstlige representasjon. Informasjonen i Oldsakssamlingens arkiver og kataloger er indirekte idet denne gir brukeren informasjon om fysiske gjenstandssamlinger.

Disse gjenstandssamlingene har et meget stort forskningspotensiale. Men som for alle store gjenstandssamlinger er det vanskelig å ha en fullstendig oversikt. Ved å legge opp registrene som databaser vil en få vesentlig bedre tilgang til og oversikt over dette materialet. Databasene



vil via de internasjonale datanettene også kunne brukes av forskere ellers i Norden så vel som i resten av verden.

Tilsvarende vil katalogopplysningene til Norsk folkemusikksamling ha stor nytte for folkemusikforskere, (hovedfags-) studenter og lokalhistorikere. Videre vil utøvere av folkemusikk samt utøvere, komponister innen andre genrer ha nytte av dem. Lokale folkemusikkarkiv knyttet til spelemannslag, muséer og bibliotek, vil ha store fordeler av en slik database.

Noe av det samme gjelder for etnologi og folkloristikk. Dersom avdelingenes arkiver og registre blir tilgjengelige som elektroniske databaser, vil dette åpne arkivene på en helt ny måte. En slik database vil være enestående i sitt slag og være interessant for folklorister, etnologer og alle andre som måtte være interessert norsk folkekultur.

Databasene vil også være aktuelle for kulturelt og historisk arbeid på det lokale plan, lokalarkiv og for bygdebokprosjekter med gårds- og brukerhistorie. Slike databaser vil også kunne hjelpe kulturadministrasjonen i kommuner, fylker og stat.

Men også annen offentlig virksomhet vil direkte nyte godt av databasene. Statens navnekonsepttjeneste har til nå vært knyttet til Avdeling for navnegransking ved Universitet i Oslo, men blir fra 1. juli 1991 desentralisert. I tillegg til i Oslo blir det konsulenttjenester i Bergen, Trondheim og Tromsø. Denne omorganiseringen skaper et stort behov for en felles database som er søkbar for alle de regionale tjenestene. I tillegg vil en slik database kunne kobles mot de såkalte GAB-registrene (grunn- adresse- og bygningsregistre), Universitetets database over faste fornminner (se nedenfor) og den vedtatte databasen ved Statens kartverk over navn på kart. En slik kobling vil kunne effektivisere det statlige og kommunale planarbeidet, jvf. Lov om kulturminnevern.

Oldsaksamlingens database for fredede faste fornminner danner grunnlaget for all planbehandling i kommuner og fylkeskommuner og for annen virksomhet som medfører inngripen i terrenget, e.g. kraftverk og flyplasser. I denne forbindelse er det også meget viktig å ha tilgang til oversikter over løse fornminner inkludert oldsaker og mynter funnet i Norge. Funn av slike gjenstander indikerer ofte at det er eller har vært faste fornminner (boplasser o.l.) i funnområdet. Riksantikvaren, fylkeskommunene og Oldsaksamlingen har alle et forvaltningsansvar etter Kulturminneloven og er dermed avhengige av tilgang til disse opplysningene.

Aktualiteten av opplysningene i Oldsaksamlingens registre illustreres ved at det nå er reist krav fra Miljøverndepartementet om at registret over faste fornminner skal kunne knyttes opp mot det såkalte GAB-registret (grunn-, adresse- og bygningsregistret). Dette registret er i dag lagt opp som databaser ved kommunedatasentralene. Utformingen av en

slik forbindelse er en naturlig del av dokumentasjonsprosjektet og bør løses i samarbeid med Riksantikvaren.

Hva er så hensikten med å gjøre de store leksikografiske samlingene elektronisk tilgjengelige, dvs. legge dem opp som databaser? I de senere tiår har det vært en eksplosiv utvikling av kommunikasjonsmidler, både for mennesker og informasjon. Vi lever derfor i en verden hvor landene knyttes sammen med stadig tettere bånd. Dette øker behovet for språklig referanselitteratur generelt og ordbøker spesielt. Den økte reise- og informasjonsutvekslingen over landegrensene gir seg utslag i et økt forbruk av slik litteratur, men også i et sterkere behov for stadig ajourførte verker.

De nasjonale ordsamlingene ved Universitetet i Oslo utgjør et viktig bakgrunnsmateriale for enhver ordbok som omhandler eller innbefatter det norske språket. En elektronisk database over disse samlingene vil bidra til å effektivisere dette arbeidet og muliggjøre utgivelse av flere typer ordbøker og dermed styrke det norske språket mot den stadig mer aggressive påvirkningen utenfra.

Dette bringer oss over til den kanskje viktigste effekten av å gjøre ordsamlingene elektronisk tilgjengelige. I løpet av de siste fem årene er en stor del av skrivemaskinene i både offentlig virksomhet og i næringslivet erstattet av personlige datamaskiner og elektroniske dokumentbehandlingssystemer av ymse slag. Denne utviklingen har knapt bidratt til papirløse kontor, men har i alle fall økt flommen av dokumenter i Norge så vel som i utlandet. Det viktige er imidlertid her at de fleste dokumenter og brev nå produseres ved hjelp av elektroniske tekstbehandlingssystemer. De fleste slike systemer er laget for å understøtte innskriveren språklig; de kan inneholde funksjoner for ordretting, oversettelsesstøtte og skrivestøtte generelt. Ulempen er at de i sin opprinnelse er engelskspråklige og gjerne har mangelfullt utviklede funksjoner for norske brukere. Utvikling av slike skrivestøtteprogrammer og -funksjoner kan sammenliknes med utvikling og redigering av ordbøker og annen referanselitteratur, men krever i tillegg omfattende elektroniske ord- og betydningsdatabaser.

Dersom Norge skal kunne delta i internasjonale språkprosjekter, er det absolutt nødvendig at det finnes uttømmende og elektronisk tilgjengelige databaser over det norske ordtilfanget. Hvis ikke, er det et smertelig faktum at vi i det internasjonale miljø rett og slett vil bli oversett, med de katastrofale følger det vil ha for Norge som språkkulturelt samfunn. Denne problematikken blir særlig aktualisert av den sterkere tilknytningen av Norge til EF som synes å være forestående.

Som det fremgår av beskrivelsen av ordsamlingene, inneholder mange sedler utfyllende opplysninger om oppslagsordets bruk. Disse opplysningene gir samtidig et innblikk i den lokale kultur ordene brukes eller

er blitt brukt i. På denne måten gir ordsamlingene mange og viktige opplysninger om norsk kultur og væremåte. Allerede i dag brukes sedlene som kilde for opplysninger om ulike emner innen folkeminne, etnografi, lokalhistorie osv. Men i dag er et slikt arbeid møysommelig, idet oppslagsordene er den eneste inngangen til materialet. Ved å legge ordsamlingene opp som en database vil man nærmest som en bivirkning få åpnet denne skattekisten full av informasjon om norsk folkekultur.

#### **4 INNSKRIVNING OG ANNEN DATAREGISTRERING**

---

Som det fremgår av arkivoversikten er det mange og forskjelligartede arkiver og registre ved HF-fakultetet. Samlet representerer disse enorme datamengder, ca. 11–12 millioner kort og dokumenter. Innskrivning og annen dataregistrering vil derfor måtte være et stort løft. Antall nødvendige årsverk vil avhenge av ambisjonsnivået, både når det gjelder grad av nøyaktighet og grad av kompletthet. Her må det understrekes at det er uaktuelt å gå på akkord med nøyaktigheten, idet dette vil gjøre alle språkarkivene verdiløse. Den regulerende faktor er altså hvor mye av materialet som skal skrives inn eller registreres på annen måte (avfotograferes, leses optisk).

Registreringen av data kan deles i to, innskrivning og innskanning av tekstlig materiale, og avfotografering av dokumenter og gjenstander. Innskrivningsdelen av prosjektet anslått til omlag 600 årsverk, heri medregnet korrektur og kvalitetskontroll. Det store innskrivningsarbeidet er tenkt utført av personer lønnet via sysselsettingsmidler. Faglig tilrettelegging, kvalitetskontroll og korrektur er tenkt gjort av hovedfagsstudenter og ferdige kandidater engasjert som vitenskapelige assistenter.

Universitetets erfaring med bruk av arbeidskraft finansiert via sysselsettingsmidler, har gitt tommelfingerregelen "ett årsverk investert gir maks. fem årsverk i produksjon". Vi går derfor ut fra forholdet én vitenskapelig assistent for hver fjerde sysselsettingsansatt i de følgende beregningene.

I kostnadstabellen under er hver vitenskapelig assistent-stilling kostnadsberegnet til brutto kroner 240 000. Hvert av de ufaglærte årsverkene er satt til 1500 timer. Et normalt årsverk er drøye 1700. Timeantallet er redusert, idet Universitetet bare skal betale for netto produksjon. Dette gir følgende pris for en vitenskapelig assistent-årsenhet:

|                           |             |
|---------------------------|-------------|
| 4 ufaglærte à kr. 45 000  | kr. 180 000 |
| 1 vitenskapelig assistent | kr. 240 000 |
| <hr/>                     |             |
| Pris pr. enhet            | kr. 420 000 |

Den totale arbeidsmengden er beregnet til 612 årsverk, eller 122 enheter pluss to løse ufaglærte årsverk. Dette gir en total kostnad på 51 780 millioner kroner (i 1990 kroner) for innskrivningen, hvorav ca. 60% er lønn til vitenskapelige assistenter og 40% er utgifter til de sysselsatte. Men man skal være klar over at disse 40% dekker 80% av arbeidet.

I begynnelsen av prosjektet vil innskrivningen foregå lokalt, i eller i nærheten av fagmiljøene, slik at man kan høste erfaring og se hvor stor innskrivningskapasitet det er mulig å bygge opp. Det er for 1991 bevilget midler til fire slike vitenskapelig assistent-enheter. Disse skal brukes i de leksikografiske avdelingene og i Oldsaksamlingen. På sikt kan innskrivningen flyttes til lokalmiljøer andre steder i landet.

En registrering av slike datamengder det her er snakk om, er en formiddabel oppgave. Det er viktig å respektere det faktum at det krever erfaring og kunnskap å få en slik innskrivning/registrering til å fungere.

## **5 SYSTEMUTVIKLING**

---

Prosjektet har ingen ambisjoner om å utvikle all programvaren selv. Dette ville være et høyst urealistisk mål. Men selv om mye kan kjøpes i butikken, må det foretas vurdering ut fra og tilpassing til prosjektets behov. Prosjektet består derfor dels i tilrettelegging av eksisterende metoder, men innebærer også nyutvikling av metoder og har således en klar forskningsmessig karakter.

### **5.1 Datamessige løsninger**

---

Ved de ulike avdelingene er det i bruk fire forskjellige databaseverktøy for PC (dBaseIII, ISIS, DataEase, 4th Dimension, FELTED). Videre er to ulike ikke-PC verktøy i bruk (SIFT, TRIP). En overordnet målsetning med dokumentasjonsprosjektet er å gjøre de ulike arkivene allment tilgjengelige i en nettverkløsning. Det sier seg selv at de enkelte avdelingenes løsninger må samordnes i en helt annen grad enn tilfelle er i dag. Det er imidlertid viktig å velge en løsning som gir størst mulig fleksibilitet.

Den tradisjonelle løsningen er en sentral(istisk) database. Alle arkivene administreres av én maskin, og databaseprogrammet som administrerer brukerne, deres søk og oppdateringer, kjøres på denne maskinen. Fordelen ved en slik løsning er at den er enkel, velprøvd og at det finnes ferdiglagede databaseverktøy basert på den. Fulltekstsystemet SIFT er foreløpig basert på denne løsningen.

Ulempene er imidlertid store. I en slik løsning vil databasemaskinens kapasitet være en eventuell flaskehals. De miljøer som har egne maskiner,

kan da bare bruke disse som terminaler mot databasemaskinen. Den lokale datakraft brukes altså bare til å få brukerens maskin til å oppføre seg som en terminal. Rapportgenerering o.l. blir derimot belastet den sentrale maskinen. Dette vil også føre til økt nettrafikk, idet ikke bare data, men også skjermbilder må overføres via nettet. Løsningen strider mot tendensen mot stadig mer distribuert databehandling.

Den sentrale databaseløsningen beskrevet over, er tradisjonell og noe gammeldags. I de senere år er løsninger med distribuert prosessering, dvs. at ulike maskiner gjør ulike deler av en oppgave, blitt stadig mer utbredt og er avgjort de løsninger som vil bli brukt i årene som kommer. Slike løsninger baserer seg ofte på den såkalte klient/tjener-modellen.

Dette er en relativt ny metode og det er derfor rimelig å bruke noe plass for å forklare den med et enkelt databaseksempel.

**Klient/tjener-modellen.** I tradisjonelle oppsett sitter brukeren ved en terminal og skriver sine data ved hjelp av tastaturet. Tegnene blir sendt til maskinen og behandlet. Maskinen sender så tegn tilbake til skjermen. For brukere som fra sin PC benytter de sentrale maskinene ved Universitetets sentrale edb-senter, (USE), betyr dette at PC-en sender hvert tegn de skriver over nettet og mottar over nettet alle tegn som skal skrives ut på brukerens skjerm. Ved bruk av tekstbehandlingssystemer, databaseprogrammer og andre programmer som ofte forandrer hele skjermbilder, er det store datamengder som må overføres over nettet. Klient/tjener-modellen er en løsning for å få redusert denne unyttige trafikken. For et databasesystem er idéen den at det i databasemaskinen og i den lokale maskinen (f.eks. en PC) er installert programmer som kommuniserer med hverandre. I den sentrale maskinen virker den delen av databaseprogrammet som foretar søk i og administrasjon av databasen. Dette er tjenerprogrammet. I brukermaskinen kjøres den delen som tar seg av skjermbilder og annen presentasjon av data. Dette er klientprogrammet. Grovt sagt kommuniserer de to delene ved at klientprogrammet sender beskjeder til tjenerprogrammet og får data og beskjeder tilbake. Trafikken over nettet og belastningen på den sentrale maskinen reduseres dermed dramatisk.

Skissen ovenfor representerer en enkel situasjon med en tjener og en klient. Modellen kan godt utvides til flere tjenere og flere klienter eller mer likeverdige partnere. Det viktigste her er å gi leserene en idé av hva såkalt distribuert databehandling innebærer.

Dette setter visse krav til databaseverktøyet, men vil samtidig øke fleksibiliteten i systemet. For å utnytte lokalitetsprinsippet bør fagmiljøenes data lagres på filtjenermaskiner i de enkelte miljøenes lokale nettverk. Sikkerhetskopiering ("backup") foretas over nettet av USE. Dette er altså en distribuert plassering av dataene. Men det betyr ikke

nødvendigvis at en får en såkalt distribuert database. Uttrykket "distribuert database" forekommer ofte i faglitteratur og i ymse datadiskusjoner der det ofte brukes uklart eller direkte feilaktig. Det er derfor på sin plass å utdype dette begrepet noe.

Med en database mener man gjerne data organisert etter en gitt logisk modell. Dataene kan godt være lagret i separate filer på separate plattelager ("disker"). Søkning og oppdatering vil foregå i henhold til den definerte logiske modellen.

En distribuert database er en database der ikke bare dataene kan være lagret ulike steder, men hvor også søkeprogram og andre databaseadministrerende program er spredt rundt på ulike maskiner. Database-systemet er altså bygget opp av separate enheter som kommuniserer seg i mellom. Men hele datasamlingen er organisert i henhold til en felles logisk modell. For brukere av systemet virker det som en enkelt enhet. Det vil altså være mulig å stille spørsmål som *Skriv ut navnet på alle gårder i kommuner der det er funnet greske mynter og er innsamlet eventyr med orientalske temaer*. Dette konkrete spørsmålet er noe søkt, men det illustrerer bredden i en slik base. En slik mulighet er selvfølgelig også til stede om alle arkivene legges opp som én database på en enkelt maskin.

Men styrken ved en distribuert databaseløsning ligger også i at selv om en del stopper, kan resten fungere. Siden de fleste henvendelser vil være av lokal art, vil en slik lokal stans kun ha en lokal virkning. Videre vil dette lokalitetsprinsippet redusere nettrafikken.

Det er likevel lite realistisk å tenke seg å legge de ulike HF-arkivene opp som en samlet distribuert database i overskuelig fremtid. Hovedgrunnen er at det finnes svært få kommersielt tilgjengelige program-systemer for konstruksjon av distribuerte databaser. De som finnes, er beregnet på standard arkivsystemer og dekker ikke HF-arkivenes behov for søking i fritekst.

## 5.2 Datateknisk status for prosjektet

---

I dokumentasjonsprosjektet har vi foreløpig valgt en kombinasjon av den sentrale løsningen og en distribuert løsning basert på klient/tjener-modellen. Det er som en start anskaffet en filtjener, en Decstation 5000/200 med 2 Gigabyte plattelager, tilknyttet Universitetets stamnett. På denne maskinen er det lagt opp fritekstsystemet SIFT og relasjonsdatabasesystemet INGRES. SIFT vil i første omgang bli brukt til Oldsaksamlingens arkiver og INGRES til ordarkivene og Folkemusikk-samlingens databaser. SIFT-databasene vil fungere etter den tradisjonelle sentrale modellen, mens INGRES-databasene vil bli lagt opp etter klient/tjener modellen.



Den innkjøpte maskinen er meget kraftig og de innskrevne dataene vil bli lagt på denne så langt kapasiteten rekker.

Hvor lenge dette varer, vil avhenge av innskrivningstakten, men også av hvorvidt det senere kjøpes inn arbeidsplassmaskiner som trenger en slik maskin som filtjener.

For tiden arbeides det med å med å modellere en databaseløsning for nynorsk- og for gammelnorskarkivet i INGRES. For gammelnorsk-avdelingen vil vi prøve å kombinere relasjonsdatabasen med et kanadisk fulltekstsøkeprogram (PAT) utviklet i forbindelse med den elektroniske utgaven av *Oxford English Dictionary*, Oxford University Press. For begge disse arkivene vil det bli laget en klient/tjener-løsning slik at Macintosh-brukerne ved de to avdelingene vil få et grensesnitt i Hypercard.

Som demonstrasjonsprogram vil vi også legge opp en interaktiv versjon av Landbruksordboka og bokmålsavdelingens nyordregister.

## **6 VIDERE AKTIVITET**

---

Parallelt med at innskrivningen starter, vil vi fortsette utviklingen av ryggraden til det samlede arkivsystemet. Dette blir bygd på UNIX-baserte filtjenere knyttet til Universitetets nett etter tjener/ klient-modellen som er beskrevet tidligere. Presentasjonssiden skal baseres på X-windows. I tillegg skal det lages klientprogrammer som kan stå som mellomledd mellom systemet og sluttbrukersystemer som ikke bruker X-windows, f.eks. Hypercard e.l.

Målet med dokumentasjonsprosjektet er å øke tilgjengeligheten til HF-fakultetets arkiver ved å overføre disse til moderne lagringsmedier. Innskrivningen og annen registrering er beregnet til å ta seks år. Utviklingen på både maskinvare- og programvaresiden er uhyre rask, og prisene går stadig ned. I løpet av innskrivningstiden vil tilbudet av dataløsninger ha utviklet seg enormt. Det er derfor viktig ikke å ha for store ambisjoner om utvikling av egen programvare, men i størst mulig grad benytte seg av tilgjengelig "hyllevare". Men det er også viktig å unngå å kjøpe omfattende og dyre ferdiglagde løsninger som etter kort tid kan virke som en tvangstrøye.

Konklusjonen er at innskrivningen og registreringen må foregå slik at all ønskelig informasjon i originalmaterialet blir bevart. Med en gang arkivmaterialet er gjort maskinleselig, er presentasjons- og organisasjonsmulighetene legio.

# *STEMMA, A ROBUST NOUN STEM GENERATOR FOR FINNISH*

---

---

*Kimmo Kettunen*

Finnish computational morphology has been studied rather intensively during the last ten years and there are nowadays many programs which either analyze or synthesize Finnish word forms or do both. The most prominent programs have been those based on the two level model developed by Kimmo Koskenniemi in the early 1980's. The champion's challenger has been a program named Morfo (Jäppinen & Ylilampi 1986), which has been developed in the Kielikone project. Also other programs exist, and some of them have been made outside the major projects as individual attempts. These include e.g. Eugene Holman's Finnmorf (1988), which is a CALL program for foreign students of Finnish.

I have recently finished a prototype of a Finnish noun stem generator, which I have named **Stemma** (a Finnish form of the English word *stem*). The program is able to produce all the differing stems and substems for an input noun which has been given to it. The program may not bring anything considerably new to computational modeling of Finnish morphology, but it is a reliable and robust tool program which can be adopted to many uses. The program has been implemented with the programming language **Icon**, which has not been used very much in computational linguistics although its characteristics are very well suited at least for prototypes and non-commercial research programs.

## *THE STRUCTURE OF THE PROGRAM*

---

The working of the program is based on string and substring matching of the input word. The program analyses the word from its end and according to its characteristics sends the processing to proper subprocedures. The most important sub-procedures are grade alternation, plural

formation and 12 pattern procedures of which each can handle one or several types of nouns (mostly one type). The program also has six small vocabularies, which contain either all the tokens of exceptional small paradigms (such as *vesi*-type) or exceptional words which are not affected by the grade alternation rules. Altogether some 600 words are included in the vocabularies or in the rules themselves. Otherwise the program relies only on substring matching and pattern analysis and does not need comprehensive lexicons or paradigm markings. The flow of the program can be described as follows:

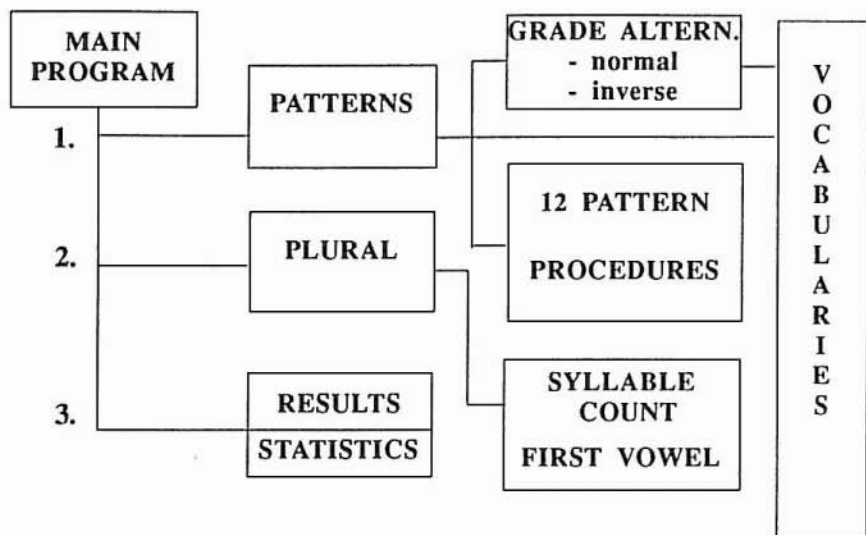


Fig. 1. Main Parts of the Program

The main program reads the input, splits compounds to parts, makes some basic stem and character variable definitions and calls for the other procedures. The actual processing of the word is begun by the patterns procedure, which first decides whether the word needs to be handled in grade alternation procedures. If grade alternation seems necessary, the word is sent to appropriate alternation procedure, either to weakening or strengthening. Otherwise it is matched against different pattern rules, which decide whether any one of the 12 special pattern procedures has to be invoked. After this the word is given to plural formation, which may call syllable count procedure and analysis of the

first vowel, if needed (only with words ending with *a* and *ä*). After this the processing is finished and the resulting stems can be given out. When the program stops, it also gives a short statistics, which states how many words it has processed, how many stems were produced and the arithmetic mean of the amount of stems per input word. Also the processing time used is given.

## **COVERAGE OF THE PROGRAM**

---

Stemma covers now almost completely the generation of Finnish noun stems. Also adjectives are covered pretty well, only comparison stems and alternations which are category specific to some types of multisyllabic adjectives are not handled yet. Pronouns and numerals are also outside of the program's present scope.

The only noun phenomena not covered yet are those few compound nouns which have double inflection, i.e. inflect both from the basic and modifying parts as for example *vanhapoika* -> *vanhanpoja*- ('a bachelor'). This kind of noun is very rare in Finnish. There are less than ten of them.

Normal compounds are handled right, if the last part of the compound is separated from the rest with /, e. g. *avioliitto* ('marriage'). If the last part has not been separated, the result may be right or wrong depending on the word. The program does not use comprehensive dictionaries and no other way of separating the parts of compounds is reliable.

Homographic nouns are not distinguished yet, and thus the program is not able to give all the stems for example for the noun *laki*, which is a common basic form to two differently behaving nouns (*laki*, *lain* 'law' / *laki*, *laen* 'crest'). Homographic basic forms are also not very common and they can be easily covered if that is necessary.

I have made several test runs on different amounts of basic noun forms. My largest test file includes 3736 nouns, which have been randomly selected from a corpus of some 35 000 non-compound words. When this test file is run, Stemma gives over 12 000 different forms as output (where of course the 3736 forms are the same as the input). The percentage of erroneous forms in this sample is **less than one**. Other smaller test corpora (with 1000 and 133 different basic word forms) have also given similar results, and it seems that the accuracy and coverage of the program is somewhere between 98-99.2 per cent. Stemma is thus a very robust and reliable prototype program for Finnish noun stem generation.

## ***POSSIBLE USES OF STEMMA***

---

A program like Stemma has not very much interest in itself, since noun stems per se are not very useful. But as a tool program for different applications Stemma has some interesting perspectives. It has at least four kinds of possible uses in computer aided language learning, morphological research, as a part of more sophisticated computational linguistic programs and as a part of an information retrieval system for Finnish.

A CALL program could easily be developed from Stemma by only adding some procedures that take care of concating the right suffixes to the stems given by Stemma. This could be done either straight or in such a fashion that the program shows first the stems and then asks the user to produce a certain inflected form. After getting the user's proposal, the program could check whether it is correctly formed.

As a research tool Stemma could serve in large scale quantitative analysis of Finnish noun stem formation, while it is able to produce the stems comparatively fast and reliably. It probably needs some changes for this use depending on the aims of the research, but these should be easily done.

While Stemma works only for generation, it could be used as a module in text generation e.g. in machine translation. As in the CALL application this would only need formulation of the rules, which take care of the production of right forms.

Information retrieval programs for Finnish texts are especially invalidated by the complex stem alternations occurring in Finnish words. For information retrieval purposes a program like Stemma could be used to produce all the altering forms of the search key. And as information retrieval is done almost totally with content words, i. e. nouns, Stemma's capabilities are readily suitable for the work. The only possible modification could be use of some simple mechanism to check that hits given by the stem forms are really possible inflected forms of the input key and not trash.

## ***TECHNICAL INFORMATION AND CHOICE OF IMPLEMENTATION LANGUAGE***

---

Stemma has been implemented with the programming language Icon, which has been developed at the University of Arizona in Tucson by Ralph Griswold and his colleagues (Griswold & Griswold 1990). Icon is a kind of a successor to Snobol, and has some very good features for linguistic programming. Of particular interest are Icon's good ready

made string manipulation functions and its rich repertoire of data structures.

Mats Eeg-Olofson (1987) has considered different criteria for choosing a suitable language to describe morphological phenomena in a computer oriented framework. According to him, a suitable programming language should be on a high level of abstraction, its data structures should help to model the phenomena being described etc. It seems that Icon fulfils most of these needs and is a good choice for anyone interested in morphological and other linguistic programming. Its string handling and pattern matching capabilities are excellent, it has a rich choice of data structures which include sets both for characters and as a general structure, and it has many ready made string functions which make programming in it easier than usual. Untyped variables of the language are also suitable for this kind of loosely structured work, where new variables are often needed in the midst of programming (c.f. Holman 1988).

The size of Stemma is now about 30 kilobytes of source code, which makes some 52 kilobytes of compiled i-code. The whole program has about 730 non-empty and non-commented lines which include also the six inherent vocabularies.

The implementation was done first with Icon's version 7.5 but it runs as well on version 8.0 with no modifications up- or downwards. On an ordinary 12 Mhz AT-clone the program handles some 3-5 nouns per second depending on the machine and the complexity of processing caused by the word.

Stemma runs on IBM compatible MS-DOS machines which have at least 512 kilobytes of RAM but it is also easily portable to any computer which runs Icon, including mainframes and workstations, which is due to Icon's good portability to different computing platforms.

## ***EVALUATION OF THE PROGRAM***

---

Palmer & Finin (1990) have used two different terms for two ways to look at computational linguistic programs. A program can be evaluated either as a **black box** model or as a **glass box** model. When a program is considered as a black box model the evaluator cannot see how the program is internally organized, and his sole evaluation criteria are based on the working of the program. When the internal organization of the program can also be studied the program is evaluated as a glass box model.

It is rather difficult to evaluate your own programs, but it seems to me that Stemma is working well at least from the black box point of view. It works reliably and has a high percentage of correctness. From



a strict linguistic glass box point of view it may have some inadequacies, since certain phenomena are not described in an orthodox linguistic way. But most of its rules are based on standard descriptions of Finnish morphology and thus it can be seen as a very concrete level implementation of Finnish noun stem synthesis.

## CONCLUSION

---

Computational linguistic programs are often considered difficult, time consuming and laborious to make. It seems that this is only partially true, and there exist certain areas which can be tackled reasonably well with basic knowledge of programming and normal linguistic analysis. It has been a learning process to work with Stemma and my experiences are very encouraging, since they show that any linguist who has a working knowledge of a good high level programming language such as Icon, can make a useful program of considerable coverage and complexity in just a few months. Virtuosity in programming is not always necessary in order to create something useful.

## REFERENCES

---

- Eeg-Olofson, Mats 1987. *Software Systems for Computational Morphology – An Overview*. Nordic Journal of Linguistics 10: 1–34.
- Griswold, Ralph E. & Griswold, Madge T. 1990. *The Icon Programming Language*. Second edition. Prentice Hall, New Jersey.
- Holman, Eugene 1988. *Finnmorph: a Computerized Reference Tool for Students of Finnish Morphology*. Computers and the Humanities 22: 165–72.
- Jäppinen, Harri & Ylilampi, Matti 1986. *Associative Model of Morphological Analysis: an Empirical Inquiry*. Computational Linguistics 12: 257–72.
- Koskenniemi, Kimmo 1983. *Two-Level Morphology: A General Computational Model for Word-Form Recognition and production*. Publications of the Department of General Linguistics, University of Helsinki.
- Palmer, Martha & Finin, Tim 1990. *Workshop on the Evaluation of Natural Language Processing Systems*. Computational Linguistics 16: 175– 81.

# **MULTIMEDIELABORATORIET VED AVH, UNIT**

---

---

***Julie Feilberg***

I januar 1990 ble et av de tre tradisjonelle språklaboratoriene ved AVH erstattet av et multimedielaboratorium. Den faglige ledelsen av laboratoriet ligger under *Institutt for anvendt språkvitenskap*. Utstyret ved laboratoriet er finansiert via Historisk-filosofisk fakultet, NAVF og AVHs midler.

Multimedielaboratoriet er utstyrt med 9 IBM PS/2, modell 55SX X61, 1 IBM PS/2, modell 70 121, og 1 IBM AT med kort for interaktiv video. Laboratoriet har også 5 Macintoshmaskiner: 1 Mac II, 1 Mac II cx, 1 Mac SE/30, 1 Mac IIfx 4/160 og 1 Mac IIfx/80. De to største Mac'ene er utstyrt med videokort og tilkoblet scanner, cd-rom og videoplatespillere. Fire av IBM maskinene er utstyrt med MIC-4000 kort for sanntids digital video, og er tilkoblet Sony Laservision LDP 3600D videodiskspillere. I tillegg er én IBM-maskin tilkoblet en Pioneer LD4100 videoplatespiller. Alle maskiner er utstyrt med ethernetkort og kobles via lokalnett mot Universitetets Vax-maskin. Laboratoriet har også cd-spiller for video og lyd, diverse lydutstyr, en VHS-spiller med digitalt bilde, og satelittmottaker og parabolantenne for mottak av internasjonale fjernsynssendinger.

Laboratoriet og det fagmiljøet som er knyttet til det, fungerer som et Apple multimediesenter.

## **MÅLSETTING FOR VIRKSOMHETEN VED LABORATORIET**

---

Multimedielaboratoriet brukes i undervisning ved flere forskjellige fagmiljøer ved AVH, men laboratoriets primære funksjon er å være forskningslaboratorium og forskningsverktøy for vitenskapelig ansatte og

doktorgradsstudenter ved AVH. Multimedielaboratoriet skal gi mulighet for bruk og utvikling av ny teknologi på de humanistiske fagenes premisser, og utviklingsarbeidet ved laboratoriet skal ta utgangspunkt i humanisters behov og kompetanse både når det gjelder forskning og undervisning.

Målsettingen for laboratoriet er at det skal

- være et forskningslaboratorium innen multimedia og edb-relatert virksomhet for alle miljøer ved AVH,
- være et undervisningslaboratorium for alle fagmiljøer,
- stimulere, initiere, koordinere og assistere multimedieundervisning og multimedieforskning.

## ***BRUK AV MULTIMEDIELABORATORIET***

---

Laboratoriet brukes i dag i forbindelse med prosjektvirksomhet og til kurs og seminarvirksomhet for ansatte, dr. art- og hovedfagsstudenter, som forskningsverktøy for ansatte og studenter med spesielt behov for utstyr som bare finnes på multimedielaboratoriet, ordinær undervisning i tilknytning til ulike fagmiljø, bl.a Institutt for drama film og teater og Institutt for anvendt språkvitenskap.

Laboratoriet skal også brukes i tilknytning til nye emner som er under utvikling: *EDB for humanistar* (20 vektall), *Bruk av multimedier i språkundervisning* (5 vektall), og innføringskurs i edb for alle hovedfagsstudenter (ca. 35 timer). Det vurderes også å innføre *Forprøver i grunnleggende datakunnskap* for samtlige nye studenter ved institusjonen.

## ***KURS OG SEMINARER VÅREN 1991***

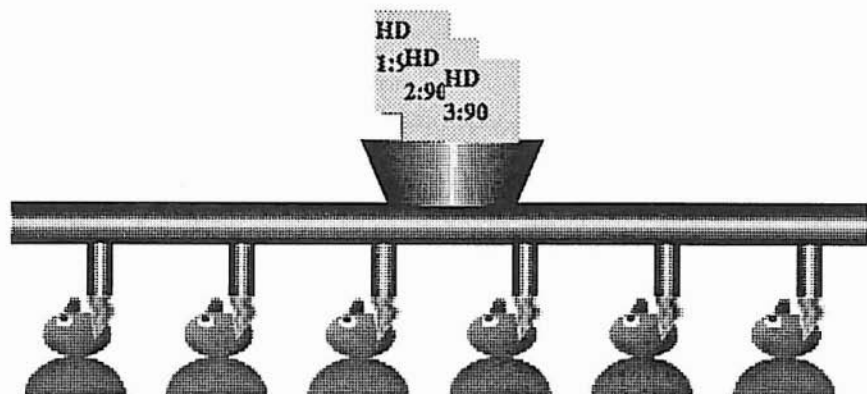
---

Et svært viktig skritt når det gjelder å skape et slagkraftig tverrfaglig miljø rundt Multimedielaboratoriet, er å heve den generelle kompetansen hos vitenskapelig ansatte, hovedfags- og doktorgradsstudenter med hensyn til innsikt i og bruk av den teknologien laboratoriet representerer. Følgende kurs og seminarer er planlagt ved Multimedielaboratoriet våren 1991:

| TEMA   | TIDSPUNKT                                     | KURSLEDER  |
|--|---|--|
| <b>PC/Pilot</b>  | Fredag 1. februar<br>12.15 – 14.00            | Olaf Husby   |
| <b>Kurs for hovedfagsstudenter og ansatte ved HF:</b><br>– grunnleggende datakunnskap<br>– introduksjon til Windows<br>– WordCruncher<br>– PowerPoint<br>– WP 5.1<br>– Statistikkprogram<br>– Bibsys | Mandag 11.<br>februar – Fredag<br>22. februar | Kirsti Rye<br>Ramberg<br><br><br><br><br><br>Arnt Dragsten |
| <b>Hypercard: ideen, pedagogikken og teknikken.</b>  | Fredag 8. mars,<br>12.15 – 14.00              | Terje Rydland  |
| <b>Hypercard: oppfølgingsseminar</b>   | Fredag 15. mars<br>12.15 – 14.00              | Terje Rydland  |
| <b>PC/Pilot: oppfølgingsseminar</b>  | Fredag 5.april<br>12.15 – 14.00               | Olaf Husby   |
| <b>Bruk av Vax og ressurser på nettet fra egen PC:</b>   | Fredag 12. april<br>12.15 – 14.00             | Kirsti Rye<br>Ramberg<br>Geir Braseth                      |
| <b>Idebehandling:</b>  | Fredag 19. april<br>12.15 – 14.00             | Lars Sigfred<br>Evensen                                    |
| <b>Idebehandling med " Framework" og "Word for Windows":</b>   | Fredag 14. mai<br>12.15 – 14.00               | Julie Feilberg   |
| <b>Elektronisk post (mail og ean) fra egen PC:</b>   | Fredag 30. mai<br>12.15 – 14.00               | Kirsti Rye<br>Ramberg                                      |
| <b>Ekspertsystemer:</b>  | Fredag 7. juni<br>12.15 – 14.00               | Jarle Rønhovd  |
| <b>Demonstrasjon av et program med interaktiv video – "Frameup":</b>   | Fredag 14. juni<br>12.15 - 14.00              | Lars Evensen   |
| <b>Evaluering av "Frameup" – et program med interaktiv video:</b>  | Fredag 21. juni<br>12.15 - 14.00              | Lars Evensen   |

Foredragsholderne representerer seks ulike fagmiljøer. Alle seminarer er gratis.

*Julie Feilberg er amanuensis ved Institutt for anvendt språkvitenskap, UNIT-AVH.*



## *DISTRIBUSJON AV INFORMASJON FRA SENTERET VIA ELEKTRONISKE NETT*

---

---

*NAVFs edb-senter for humanistisk forskning (heretter kalt Senteret), driver flere nettbaserte elektroniske tjenester.*

### *VERTSMASKINER*

---

Hovedmaskinen for Senterets elektroniske tjenester er en Sun 3/60 kalt NORA. NORAs fulle navn er NORA.NAVF-EDB-H.UIB.NO, og hun har IP-adresse 129.177.24.42. De fleste tjenestene drives fra denne maskinen.

Senteret har naturlig nok også en HELMER. Han er en Sun 386i med IP-adresse 129.177.24.41. Helmer:NAVf-edb-h står som en Apple-Share filtjener i sonen EtherMac ved Universitetet i Bergen. Denne tjenesten er kun tilgjengelig for Macintoshnett som er tilkoblet bredbåndsnettet ved UiB. Det er to volumer som er åpen for alle, GjesterLes og GjesterSkriv. Helmer forteller om seg selv gjennom filen *Les Meg* i GjesterLes.



Senteret benytter også NOBERGEN.BITNET som er IBM 4381-maskinen ved Universitetet i Bergen. Her har Senteret FAFSRV@NOBERGEN.BITNET, som er en filtjener startet for personer interessert i ICAME, *International Computer Archive of Modern English*. ICAME er engelsk språkmateriale som Senteret distribuerer. ICAME-informasjon ligger nå også på NORA, som etterhvert vil overta for FAFSRV.

## ***DISTRIBUSJONSLISTE HUMEDB***

---

HUMEDB er en distribusjonsliste for edb i humanistiske fag. En distribusjonsliste er en spesiell elektronisk postadresse som inneholder en liste over andre postadresser. Når en melding sendes til denne postadressen, blir meldingen automatisk videresendt til alle på listen. Pr. mars 1991 er det 80 navn på listen. Listen er åpen for alle, men en må tilføyes manuelt på listen etter å ha sendt et brev til administrator. Det fulle navnet på listen er HUMEDB@NAVF-EDB-H.UIB.NO og for administrator HUMEDB-ADM@NAVF-EDB-H.UIB.NO. Listen ble opprettet for et par år siden, men har ligget slumrende i påvente av at de humanistiske miljøer skulle bli mer modne for å ta i bruk elektronisk kommunikasjon.

Senteret sender ut egen informasjon på HUMEDB, i tillegg til opplysninger hentet fra andre elektroniske kilder. Deltakerne på listen oppfordres til å komme med innlegg, spørsmål, svar m.m.

I tillegg vil HUMEDB fortelle om artikler, data, programvare, m.m på Senterets filtjenere.

## ***HVA MÅ TIL FOR Å KUNNE BRUKE DENNE TJENESTEN?***

---

Du må ha en elektronisk postadresse på en vertsmaskin som har tilknytting til Uninett eller samtrafikk med Uninett. Ved mange institusjoner kan en i dag lese og skive post på sin PC eller Mac og bruke en "usynlig" Unix-maskin som postkasse.

Tilgang til listen er gratis. Dersom du ønsker å stå på listen, gi beskjed til den administrative adressen under, likeledes om du senere vil endre adresse eller ønsker å bli strøket fra listen igjen.

Adresse for innlegg HUMEDB:  
**HUMEDB@NAVF-EDB-H.UIB.NO**

Administrativ adresse HUMEDB:  
**HUMEDB-ADM@NAVF-EDB-H.UIB.NO**

## **FILTJENERE PÅ NORA**

---

På NORA ligger generell informasjon om og fra Senteret, artikler, informasjon fra andre kilder og programmer.

Det er to ulike måter å få tilgang til filene på NORA. Den første er med filoverføringsprogrammet FTP. Maskinen tillater såkalt anonym FTP.

**Bruk: FTP NORA.NAVF-EDB-H.UIB.NO**

**User: Anonymous**

**Password: Ditt navn**

**Kommandoer:**

|                               |                                 |
|-------------------------------|---------------------------------|
| <b>CD pub</b>                 | for å gå til hovedkatalogen     |
| <b>DIR</b>                    | for å se filliste               |
| <b>CD dirnavn eller CD ..</b> | for å forandre directory        |
| <b>GET filnavn</b>            | for å hente en fil              |
| <b>BINARY</b>                 | for å sette binær filoverføring |

Hva må til for å kunne bruke denne tjenesten? Du må ha adgang til en maskin som har såkalt TCP/IP forbindelse. Dette kan være en Macintosh eller PC med gratisprogram som NCSA Telnet/FTP eller HyperFTP. Det kan også være en vertsmaskin som kan brukes via serie eller modemforbindelse. Filene må da først overføres til vertsmaskin via FTP og så videre til den lokale PC eller Mac via Kermit eller lignende program.

Den andre måten er postbasert filoverføring. NAVFSERV@NORA.NAVF-EDB-H.UIB.NO er en postbasert filtjener på NORA. Noen av dere kjenner INFO@UNINETT.NO, og NAVFSERV har samme kommandoer som denne (er samme program). Med NAVFSERV vil det være tilgjengelig deler av det som er tilgjengelig via FTP. Ventelig vil den viktigste begrensningen for brukerne være at binærfiler ikke kan hentes direkte med NAVFSERV. Send et brev med *HELP* som *Subject:* for nærmere informasjon.

## **HVILKEN INFORMASJON LIGGER PÅ FILTJENERNE?**

---

Det er pr. mars 1991 følgende hovedkataloger:

|              |   |
|--------------|---|
| <b>ICAME</b> | Informasjon relatert til International Computer Archive of Modern English |
|--------------|---|

|                    |   |
|--------------------|---|
| <b>INFO</b>        | Informasjon om tekster, prosjekt, bibliografier |
| <b>KONFERANSER</b> | Oversikt og opplysninger om konferanser         |
| <b>MAC</b>         | Macintosh-programmer                            |
| <b>NAVFEEDB-H</b>  | Generell informasjon fra Senteret, referater    |
| <b>HUMEDB</b>      | Gamle innlegg i HUMEDB-listen                   |
| <b>OLUFF</b>       | Artikler fra bladet <i>OLUFF</i>                |
| <b>HD</b>          | Artikler fra <i>Humanistiske data</i>           |
| <b>NCCH</b>        | Engelskspråklig informasjon fra/om Senteret     |
| <b>NETTINFO</b>    | Informasjon om netttjenester                    |
| <b>PC</b>          | Programmer for PC                               |
| <b>UNIX</b>        | Unix programmer                                 |

Koordinator for de elektroniske tjenestene:

**Knut Hofland,**  
**KNUT@NAVFEEDB-H.UIB.NO** og  
**FAFKH@NOBERGEN.EARN**

**NAVFs edb-senter for humanistisk forskning**  
**Postboks 53 Universitetet**  
**N-5027 Bergen**  
**Tlf: + 47 (0)5 21 29 54/55/56**  
**Fax: + 47 (0)5 32 26 56**

# PAST MASTERS

---

---

*Ralph Jewell*

## INTELEX CORPORATION'S SERIES OF CLASSIC HUMANITIES TEXTS ON DISK

---

The *PAST MASTERS* series from the *InteLex Corporation* is a publishing venture that is speeding up the pace at which the meaning of the verb *to read* is undergoing far-reaching change, now that electronic media are tending to replace paper as the first choice for scholarly perusing and search in texts.

The pleasures and the rewards of reading a good tactile *real* book, printed on paper, and having obvious weighty mass, are not disputed. But it is obvious that for many purposes or tasks, or even moods, electronic texts work best. Anyone who has become familiar with using or preparing texts whose lengths are easily measured in terms of K's – just about everyone who uses a PC – will know the gains that may be won when texts are given electronically as *e-texts*, rather than by ink on paper.

Many a 1980's and 1990's scholar who has taken a PC into use has wished that the great classic texts of his subject – those true fountains of telling phrases, remarks, points to remember and authoritative distinctions – were *on disc* and as accessible for quick search as are his own files from the past few months' work. Where did *Plato* make his dramatic character, *Socrates*, say; "We learn Geometry through working out examples"? Or, where exactly did *Hobbes* say that "Words are wise men's counters – they do but reckon by them"? These are the kind of questions that crop up in the humanities, and they often crop up just when time is short.

What used to be *looking for* something or other in large texts, may now be *conducting searches* instead. The *looking for* that used to take ages, and perhaps was either never finished or finished only through good luck, may now, with electronic texts, take a very short time indeed, and end with definite results. Among those who are aware of

the gains that may be had from using electronic texts there is a demand for some new kinds of publishing. InteLex Corporation has seen this clearly, and has responded with something that is far more than a mere speculative commercial bid for a potentially lucrative market. It has published electronic editions of classic texts in the humanities. But it has done more. It has taken the most authoritative previous editions of these works, compared them, and in cases of discrepancies has judiciously resolved them or listed them as variants.

It has done still more. Taking the view that even electronic editions of very large amounts of text-material would still be somewhat cumbersome when accessed with normal word-processors, InteLex Corporation made a bold decision to publish them already fully processed for access by the very powerful text-retrieval programme *VIEWS* of the *Folio Corporation*. What one has with the publication of the InteLex editions of classics in the humanities is a combination of the best of thoroughly reliable editions of these works together with the facilities of what is probably the most impressive text-retrieval package available for use with personal computers.

The first releases include editions of major works of *John Locke*, *George Berkeley*, *David Hume*; *Thomas Hobbes*; *John Stuart Mill*, *Jeremy Bentham*, and *Henry Sidgwick*; leading modern translations of *Descartes*, *Leibniz*, *Spinoza*. Also available are composite databases for *Political Philosophy*, *Economics*, and one for *Kierkegaard's writings*. Due out in the summer 1991 are: complete works of *Spinoza* in Latin and Dutch; major works and selected correspondence of *Descartes* in Latin, French and (authorized by Descartes) French translation; *Plato* and *Aristotle* (in modern English translations); *Kant* and *Hegel* (in modern English translations and German).

To use PAST MASTERS is to have one's eyes opened to very interesting new opportunities and even new possibilities for scholarly work with large bodies of texts. An hour or so is all that is required for becoming familiar with the primary search and display functions of the built-in *VIEWS* text-retrieval programme, and these functions are impressive. Each data base, text base, or "infobase", is divided into small manageable units called "folios", which are usually paragraphs but may be footnotes or other easily and strategically distinguished small segments of text. Searching begins actually during the typing of a search request, and search requests may consist of words or phrases, logic operators, parentheses for nested searches, groups and wildcards. As the search request is being typed a scrolling window shows the words of the index that satisfy the request as it has been typed so far, and a results window shows how many folios satisfy the developing search request. The complete text of every folio satisfying the search

request is then displayed with the chosen word or words highlighted. Moving from folio to folio could not be simpler. There are also link-tokens that make it possible to move quickly from one part of the text to other parts that have an information-connection to the folio marked with the starting link-token. Items in a table of contents, for example, would be marked with such link-tokens, for immediate viewing of the relevant parts of the text.

VIEWS makes it extremely easy to criss-cross through these works according to however fancy and fleeting thoughts may strike, and in many kinds of work in the humanities, fancy and fleeting thoughts, of course, are the vital elements of the kind of pursuit that makes research deeply satisfying. The best and surest way to avoid unnecessary disagreement with this is to try it and see!

There is a surprising point to notice about how IntelLex e-text editions make very economical use of disk space. Although Folio VIEWS puts the whole text of the data-base and the index of every single word of the text into a single file, the resulting single file is just over half the size of the original text file. It manages this by using its own patented compression technique. This gives the result that for the Political Philosophy data-base, for example, the original data-base files totalled 11.5 MB approximately, while the Folio file amounted to only 7 MB.

Much might be said about this PAST MASTERS series. But it is not the intention, nor would it be appropriate, here to outline all or most of the facilities of the IntelLex PAST MASTERS series. The intention here is to pass on the message to those who have not yet heard, that such a series exists, that it is being steadily added to, that it does superbly well what it was meant to do and what many hoped it would be capable of doing. It will go a very long way towards giving a good name to the whole enterprise of arranging e-text editions of classic works as material for scholarly work. Those who decide to investigate PAST MASTERS further, have a great pleasure in store. It might even lead to philosophizing more vigorously about how words work, and what texts are meant to do – questions that the authors of many of the works in the PAST MASTERS series have themselves tried to treat!

*Ralph Jewell is Senior Lecturer at the Department of Philosophy, University of Bergen.*



# MOZART OG BEETHOVEN I HYPERCARD

---

---

*Espen S. Ore*

Det er ikke bare videoplater som kan styres fra en datamaskin. Det samme kan gjøres med CDer. Fra *Warner New Media* og *The Voyager Company* har det kommet flere produkter som kombinerer CD-lyd, HyperCard skjermbilder og Macintosh-lyd. Senteret har anskaffet de første produktene fra hver av produsentene: Mozarts *Tryllefløyten* og Beethovens *9. Symfoni*. Produktene har (selvfølgelig) mye til felles, men det er også klare forskjeller i måten stoffet blir presentert på.

## TRYLLEFLØYTEN

---

Tryllefløyten blir levert som tre CDer og en diskett. Disketten kan brukes uten CD-spiller, og den blir anbefalt brukt av "de som ennå ikke har anskaffet CD-spiller". På CDene ligger det installeringsprogram og HyperCardstakker. De inneholder også lydspor utover operaen. Disse brukes til musikalske eksempler i forbindelse med de løpende kommentarer man får mens man spiller *Tryllefløyten*. I tillegg finnes det MIDI-data for de som har CD-spiller med MIDI-utgang, og det finnes grafikk for CD+G-spillere. Denne grafikken skal kunne vises på fjernsynsskjermer når disse er koblet til CD-spilleren. (Undertegnede har aldri sett en CD+G-spiller og vet derfor ikke hvorledes dette virker.) HyperCardstakkene (både de som ligger på CDene og den frittstående fra diskett) har også innebygde styringskoder for videoplatespiller. Dette kan brukes mot videoplateversjonen av Bergmans filmatisering av *Tryllefløyten* (se fig. 1).

Installasjon og bruk av dette produktet er unødvendig tungvint. Ca 15 MB HyperCardstakker må kopieres inn på harddisken fra CDene. (Man trenger ikke kopiere inn alle 15 MB samlet med mindre man ønsker å spille hele operaen uten å måtte installere nye stakker for hver av de tre CDene.) Har man plass, er dette ikke så ille. Verre er det at man må bruke en egen Start-stakk som inneholder de nødvendige



Fig. 1. Fra Bergman's filmatisering av "Tryllefløyten". Ulrik Cold i rollen som Sarastro. (Bildet er digitalisert fra levende video med et program som fulgte med et DVA-4000 framegrabber/genlock kort, og deretter skrevet ut med programmet PhotoShop.

driverne for CD-spillere. I bruksanvisningen sies det at man må gi sin opprinnelige Start-stakk (engelsk: Home Stack) et nytt navn og så kalle den som fulgte med *Tryllefløyten* for Start. En bedre løsning er å kopiere HyperCard-programmet inn i samme mappe som resten av *Tryllefløyte*-stakkene og bruke det sammen med den egne Start-stakken der. Men har man flere kopier av HyperCard på disken, må man være litt forsiktig med å åpne stakker ved å dobbeltklikke på dem. Man har nemlig ingen garanti for hvilken versjon av HyperCard som starter. Dette er likevel et problem man kan leve med. Det samme skjer hvis man av kompatibilitetshensyn må ha både versjon 2 og versjon 1.2.x av HyperCard på samme harddisk (se nedenfor). Forøvrig virker stakkene utmerket med HC-versjon 2 selv om de er skrevet i 1.2.2. Hadde versjon 2 eksistert da de ble laget, kunne man unngått hele problemet med spesiell Start-stakk, men det er en annen historie ...

Når man har startet (dvs. satt inn en CD og åpnet den spesielle HyperCard Start-stakken), kan man velge mellom fire forskjellige sett med løpende kommentarer (se fig. 2a-d):



Fig. 2 a. Tysk libretto.



Fig. 2 b. Engelsk libretto.

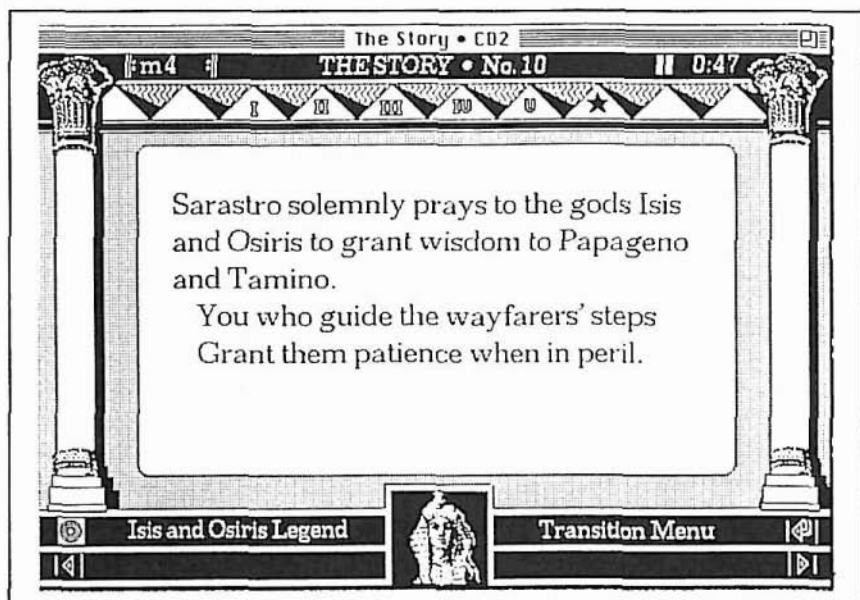


Fig. 2 c. Oppsummering av handlingen.

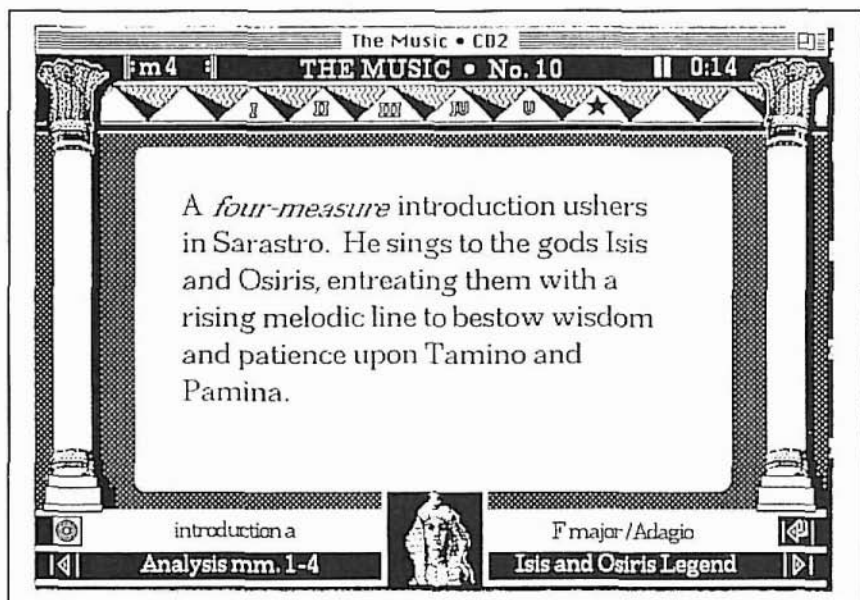


Fig. 2 d. Kommentar til musikken.

- engelsk libretto
- tysk libretto
- beskrivelse av handlingen
- musikalsk kommentar

Man kan fritt veksle mellom disse kommentarsettene mens operaen spilles. Det er også mulig å skrive egne notater som automatisk inneholder referanser til hvor i verket man befinner seg. Mens CDen spilles, dukker det opp knapper med hypertekstlige forgreninger til mer bakgrunnsinformasjon (se fig. 3). Siden CDene er spesiallaget, er det også lagt inn annen musikk enn *Tryllefløyten* (og også tale), og dette brukes bl.a. til å illustrere musikalske fenomener. En noe selsom opplevelse er et opptak fra *Carnegie Hall* fra førtiårene der *Florence Foster Jenkins* synger *Nattens Dronnings* arie.

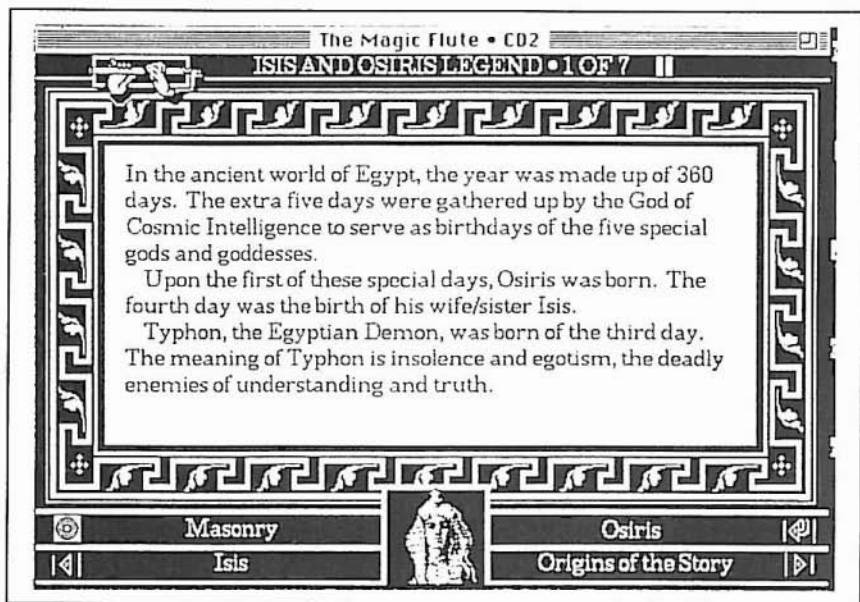


Fig. 3. En forgrening til mytene bak teksten.

## BEETHOVENS 9. SYMFONI

Dette produktet består av en vanlig lyd-CD og to HyperCard-stakker. Her kreves det ikke noen spesiell Start-stakk, men det var visse problemer med fontene som er brukt og med avspilling av digitalisert lyd, da jeg

forsøkte å konvertere stakkene til HyperCard v.2, slik at jeg fant det best å beholde en HyperCard versjon 1.2.5 på min harddisk.

I *Tryllefløyten* er det lagt an til at man skal høre musikken sekvensielt og foreta sidesprang når det dukker opp interessante forgreningsmuligheter. Dette er bare én av angrepsvinklene i Beethovens 9. (Valget "A close reading" i fig. 4) Alternativt kan man lese et lengre essay om Beethovens liv og musikk (se fig. 5) eller ett med en musikkvitenskapelig bakgrunn og analyse av 9. symfoni (se fig. 6)

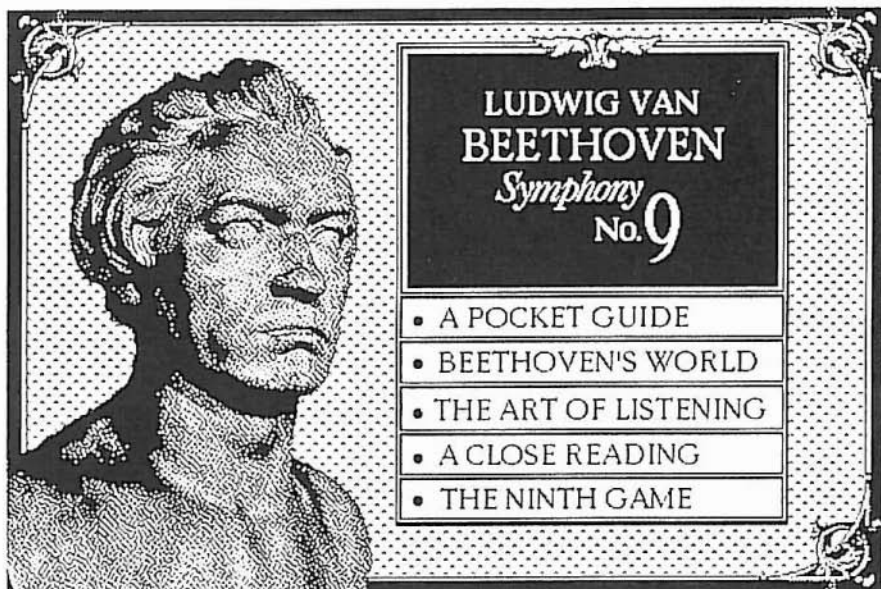


Fig. 4. Åpningskortet til Beethovens 9. Symfoni.

Velger man "A close reading" er grensesnittet i funksjon svært likt det som finnes i *Tryllefløyten*. Dette er ikke så rart, det er ikke så mange forskjellige måter å lage løpende hypertekstlige kommentarer til et musikkverk på. (Se fig. 7 a-c)

### **ER DET ENE BEDRE ENN DET ANDRE?**

Det er vanskelig å svare på dette. Programmet rundt Beethovens 9. symfoni er i det vesentlige laget av én person: musikologen *Robert Winter* fra UCLA. Dette kan være grunnen til at undertegnede finner at det "slår" hardere og mere direkte enn *Tryllefløyten*.

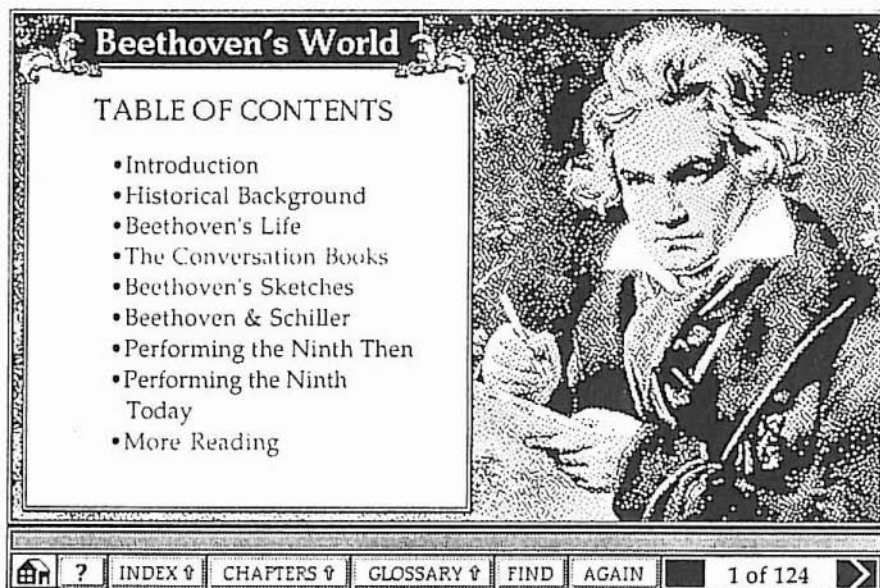


Fig. 5. Innholdsfortegnelsen til essayet "Beethoven's world".

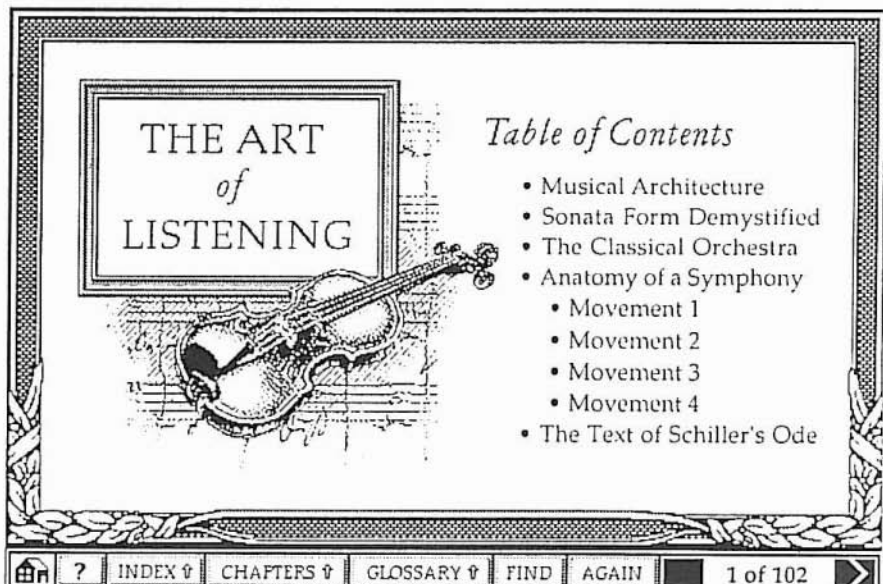


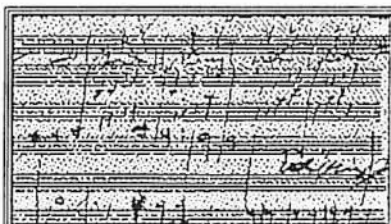
Fig. 6. Innholdet i essayet "The Art of Listening".



A CLOSE READING

MOVEMENT 4  
Presto (Extremely fast)

[1] A sharply •dissonant chord launches an apocalyptic curtain dubbed the "Horror Fanfare" in the nineteenth century. It halts as abruptly as it began.



DISSONANT  
—the description of an interval or chord that sounds harsh and/or unstable. The dissonance [the noun form] in this chord is a "minor second" of A and Bb, two notes directly adjacent to one another.

|         |            |             |      |   |       |         |       |
|---------|------------|-------------|------|---|-------|---------|-------|
| INDEX ▾ | MOVEMENT ▾ | NOTES       | 0:05 | <input checked="" type="radio"/> PLAY THROUGH | PAUSE | FIND    | AGAIN |
| ?       | GLOSSARY ▾ | 0:00 - 0:07 |      | <input type="radio"/> PLAY THIS CARD          |       | ◀ 217 ▶ |       |

Fig. 7a. Fotnoter er markert med en prikk.

A CLOSE READING

MOVEMENT 4  
Allegro assai (Very fast)

Statement 1: : a a' b a' b a'

[96] The cadence is followed immediately by the first statement, in the cellos and double basses, of the celebrated "Joy Theme." It contains only two phrases (a, b) and a variant (a'), arranged in the sequence a-a'-b-a'-b-a'. SHOW JOY THEME



SONATA-CONCERTO  
Opening Ritornello  
Horror/Recitative  
Recollection: Mvt. 1 2 3  
**Joy Theme: Strmtt. II 2 3 4**  
Exposition  
Development  
Recapitulation  
Coda

|         |            |             |  |   |  |         |       |
|---------|------------|-------------|--|---|--|---------|-------|
| INDEX ▾ | MOVEMENT ▾ | NOTES       |  | <input checked="" type="radio"/> PLAY THROUGH |  | FIND    | AGAIN |
| ?       | GLOSSARY ▾ | 3:15 - 3:22 |  | <input type="radio"/> PLAY THIS CARD          |  | ◀ 229 ▶ |       |

Fig. 7b. Her får vi muligheten til å se temaet i noteform.

The image shows a musical score for 'Gledestemaet som noter' in bass clef, 2/4 time, with a key signature of one sharp (F#). The score consists of six staves. The first two staves are marked 'a' and 'a'', the next two 'b' and 'b'', and the final two 'a'' and 'a''. Below the staves is a software interface with the following elements:

- Buttons: HIDE JOY THEME, Play from Mac, Play from CD
- Navigation: INDEX, MOVEMENT, NOTES, FIND, AGAIN
- Home icon, question mark icon, GLOSSARY, 3:15 - 3:22
- Radio buttons:  PLAY THROUGH,  PLAY THIS CARD
- Page number: 229

Fig. 7c. Gledestemaet som noter. Det kan spilles som "ren" datalyd eller hentes fra CDen.

## HVEM ER DETTE BEREGNET FOR?

Dersom undertegnede skulle diskutere i detalj i hvilken grad disse produktene kan brukes i musikkundervisning på ett eller annet nivå, ville det representere en erkennelse av Peter-prinsippet. Men som musikkglad amatør har jeg hatt stor glede av begge produktene. Og mine musikk lærere på realskolen og gymnaset som prøvde å lære meg noe om symfoniers arkitektur, kunne her hatt glimrende hjelpemidler.

# *ELASTISKE MEDIA I BERGEN*

## *SEMINAR OM INTERAKTIVE MULTIMEDIA*

---

---

*Øystein Reigem*

*The Elastic Charles* er et prosjekt ved det prestisjetunge *Media Lab* ved Massachusetts Institute of Technology. Det fulle navnet på dette prosjektet er *The Elastic Charles: A Hypermedia Journal*. *The Elastic Charles* er laget som et eksempel på hvordan informasjon kan bli formidlet i framtiden – et avansert elektronisk tidsskrift om en vil. *The Elastic Charles* handler om elven Charles, som renner gjennom Boston.

3–5. september ble det holdt et seminar i Høyteknologisenteret i Bergen hvor *The Elastic Charles* ble presentert. Arrangører var HF-fakultetet ved Universitetet i Bergen, NAVFs edb-senter, m.fl. Fra *Media Lab* stilte *Hans Peter Brøndmo* og *Glorianna Davenport*.

Brøndmo er nordmann, og er den som har utviklet *mikon*-konseptet og resten av verktøyet som brukes til å bygge opp hypermedia-anvendelser som *The Elastic Charles*.

Davenport har bakgrunn som filmskaper og fra arbeid med redigering av dokumentarvideo. Hun var dermed vant med en situasjon der mye relevant og interessant informasjon (video o.a) tas opp og samles inn, mens bare en brøkdel velges ut og presenteres. I tillegg er et medium med så lav oppløsning som TV ikke egnet til å presentere viktige typer informasjon som tekst, kart, m.m. Med den nye teknologien så hun en mulighet til å frigjøre seg fra begrensningene som lå i den gamle, og å skape en "ny litteratur", som hun sa. Et nøkkelpunkt for henne ble hvordan seeren skal lokkes til å interagere med det nye mediet, og der sto Brøndmos "elastiske" verktøy sentralt.

### *THE ELASTIC CHARLES*

---

*The Elastic Charles* er et eksempel på *hypermedia*, som er en ny måte å organisere og presentere informasjon på. *Hypermedia* vil for det første si at det er snakk om ulike typer informasjon – *tekst, grafikk, stillbilder, levende video og lyd*. For det andre er informasjonen stykket opp i

passer biter, som igjen er knyttet sammen med *lenker*. En følger disse lenkene og vandrer fritt rundt i informasjonen. Dette i motsetning til den *lineære* informasjonspresentasjonen vi vanligvis møter. Vi leser gjerne en tidsskriftartikkel pent og pyntelig fra start til slutt – i det minste er innholdet beregnet på det. Vi ser passivt et TV-program fra begynnelse til ende.

*Elastiske media (elastic media)* var forøvrig Davenports yndlingsbetegnelse for hypermedia-systemer. I det la hun at brukeren i slike systemer kan strekke, komprimere, avbryte og annotere (gjøre tilføyelser, notater til) sin opplevelse. (Husk på denne termen. Den kan være grei å ha den dagen du er lei av moteordene interaktiv video, multimedia, hypermedia.)

*The Elastic Charles* er utviklet på en Macintosh med ColorSpaceII og senere også ColorSpaceFx. Dette er kort for bl.a digitalisering av videobilder og for å legge tekst og grafikk over video. Programvaren er skrevet i HyperCard, men med en del eksterne rutiner skrevet i C (?).

Den som har sett HyperCard, vet at en forflytter seg i informasjonen ved å museklikke på "knapper", felter, ikoner, ord og annet i skjermbildet. Typisk er det at et ikon – en enkel grafisk figur – symboliserer et stykke informasjon. Ved et klikk på ikonet kommer denne informasjonen opp på skjermen. Når informasjonen er stillbilder eller levende video, er det i hypermedia-anvendelser vanlig å bruke *pikoner (picons – picture icons)*, som er små stillbilder. Nytt i *The Elastic Charles* er *mikoner (micons – moving eller motion icons)*. Et mikon er en kort sekvens av levende video som løper i løkke. Mikoner blir brukt til å symbolisere levende video.

Mikoner er spennende på den måten at de umiddelbart fanger oppmerksomheten, samtidig som de kan fortelle mer enn et pikon eller vanlig ikon. I *The Elastic Charles* kan mikonene i tillegg dukke opp *dynamisk*, f.eks i en sekvens levende video. Et eksempel er en sekvens som handler om forurensningen av Charles. Mens videoen løper, og en ser det forurensede vannet flyte forbi, kommer det plutselig til syne et mikon med bilde av en brovokter som har fulgt utviklingen i mange år. Klikker en på ham, fyller han skjermen og gir oss sin kommentar til saken. Etterpå vender en gjerne tilbake til elvesekvensen igjen, og et nytt mikon dukker opp – denne gangen av en kjent miljøaktivist. Klikker en på henne, får en høre hennes mening, osv.

## **THE ELASTIC HAUKELAND**

---

Seminaret var et arbeidsseminar, og etter å ha blitt presentert for *The Elastic Charles* og det elastiske verktøyet, fikk de rundt 30 deltakerne

i oppgave å lage en prototyp av et tilsvarende opplegg for informasjon om *Haukeland Sykehus* til pasienter eller mulige framtidige pasienter. På forhånd var det lagt til rette noen eksemplarer av en videoplate med materiale om sykehuset. Deltakerne ble fordelt på seks maskiner, hvorav to var fullt utstyrt med videoplatespiller, monitorer og Color-Space-kort. En MacRecorder var også tilgjengelig for innspilling av ytterligere tale og lydeffekter.

Arbeidet ble noe hemmet av mangelen på fullt utstyrte arbeidsstasjoner, men ellers var deltakelsen en meget interessant opplevelse. Fem grupper fikk tildelt hver sin oppgave:

- interaktive kart og andre hjelpemidler for orientering
- oversikt over sykehusets tjenester
- inn- og utsjekking
- informasjon/ordliste, spørsmål i forbindelse med språk og kultur
- introduksjon til sykehuset og til resten av informasjonen

I en fellesesjon til sist på seminaret ble de enkelte delene sydd sammen til et komplett system av Brøndmo.

Nå er det tvilsomt om den resulterende prototypen ville egnet seg i praktisk bruk. Bl.a hadde flere av prosjektdeltakerne vært i det lekne hjørnet – eksempelvis:

- Innsjekking via "Hva feiler det Dem?"-meny med valgmuligheter som "Hjertestans", "Hjerneblødning" og "Vondt i baken".
- Meny med "Cut", "Paste", "Undo", etc i oversikten over kirurgiske tjenester.
- Kommentar "Deres avskjed med sykehuset kan foregå på flere måter ..." med etterfølgende panorering til Møllendal kirkegård.

(Tragi)komisk var det også at deltakeren fra *Haukeland Sykehus*, som også spilte pasient i videoen, ikke kunne være til stede på seminaret, da hun lå på sykehuset med brukne ribbein.

## ***ELASTIC I BERGEN***

---

Ved seminarets slutt ble all programvare både for *The Elastic Charles* og *Haukeland* lastet ned på en server hos EDB-senteret ved Universitetet i Bergen. I tillegg ble kopier av begge de aktuelle videoplatene værende i Bergen. Fra serveren har vi ved NAVFs edb-senter kunne hente inn systemene til vår Mac II med ColorSpaceII og Fx.

Elastic-verktøyet har innebygde drivere for Pioneer 4200 NTSC videospiller. Pioneer Electronics Norge A/S har lånt Senteret en spiller slik at vi har kunnet finpusse resultatene fra seminaret og vise dem i våre lokaler.

## ***DET SOM GJØR HYPERMEDIA VANSKELIG***

---

En rekke interessante problemer oppstår når en skal organisere informasjon som hypertext eller hypermedia (og fra nå av svikter oversettelsen): Et av dem er "*granularity*", dvs at informasjonen i et hypermediasystem må være brutt ned i passende deler. Delene må være store nok til å være selvstendige, men heller ikke for store. Et annet er "*kontinuitet*", dvs hvordan en sikrer at informasjonen som når brukeren, fortsatt har en kontinuitet, selv om den er satt sammen av deler som kan komme i ulikt utvalg og rekkefølge. Systemet må også på et hvert tidspunkt kunne *avbrytes* ("*interruptability*").

Kontinuitet har med *kontekst* å gjøre. For å sette informasjonen i en sammenheng kan en også bruke vanlige teknikker med å presentere *oversikter* før detaljinformasjon. Ellers er *kontekst-modeller* et interessant konsept. F.eks kan en tenke seg at systemet holder rede på hvilke tema brukeren vet noe om allerede.

Hypermedia-systemer bør ha en "*limited look-ahead*", dvs at systemet skal ligge i forkant og finne fram og organisere den informasjon som brukeren vil kunne komme til å be om. Dette er et vanskelig problem, som en ikke har kommet langt i løsningen av. En annen ønskelig egenskap er "*graceful degradation*", dvs at systemet, som nødvendigvis vil inneholde informasjon av begrenset omfang, på en elegant måte kan signalisere at brukeren er i ytterkanten av systemets kunnskap – at det ikke er mer å hente for brukeren i den retning hun beveger seg. Systemet bør evt lede brukeren bort fra kanten og inn igjen i kunnskapen.

# **SEMINAR OM INFORMASJONS- TEKNOLOGI I DE HUMANISTISKE FAG – TENDENSER OG UTVIKLINGSBEHOV**

**BERGEN 5. OKTOBER 1990**

---

---

Dette seminaret var en del av det evalueringsarbeidet som pågår ved NAVFs edb-senter for humanistisk forskning. Direktør, *Jostein H. Hauge* åpnet seminaret med et ønske om at deltakerne ville beskrive sine fagområder og utviklingen videre framover på en slik måte at det kunne bidra til å avklare behovet for tjenester fra Senteret.

Seminardeltakerne kom fra HF-fakultetenes edb-seksjoner og fra en rekke humanistiske institutter ved universitetene – med hovedvekt på språkfagene. Deltakerne hadde derfor ulike opplevelser av forholdet mellom humaniora og edb-bruk, og presenterte en variert ønskeliste for Senterets videre arbeid.

Edb-bruken i humanistiske fag er kanskje bare i sin barndom. Enkelte fagmiljø er fremdeles lite interessert i edb, men ettersom utstyret har blitt kraftigere og billigere, lagrer mer og tar mindre plass på skrivebordet, har det blitt flere og flere edb-brukere også blant humanistene. Kostnadene er et av de store problemene for humanistiske forskere som tradisjonelt har hatt lite utstyrsbehov sammenliknet med andre grupper. Men en deltaker mente at humaniora i årene som kommer vil få et utstyrsbehov som kan sammenliknes med det naturviterne fikk i forrige århundre. Fagene vil kreve laboratorier for forskning og undervisning.

*Kirsti Rye Ramberg* fra Edb-tjenesten for humanistiske fag, UNIT-AVH, tok opp rammebetingelsene for edb i de humanistiske fag. Hun mente faktorer som budsjett, personell og tilgjengelighet av maskinressurser var vanskelige, men ikke håpløse å påvirke. Det var langt vanskeligere å påvirke utviklingen av maskinvare og kommersiell programvare.



Tidlig på 80-tallet sa en gjerne: *Tenk hva vi kunne fått til hvis maskinene var slik.* I dag utvikles stadig kraftigere og rimeligere maskiner, som ikke setter grenser på samme måte som tidligere. Utfordringen er både å se de nye mulighetene og å formidle dem til andre. Ellers går det slik som en nå opplever mye av innen multimedia. Multimedia ser ut til å ha "tatt av" (i hvert fall hvis en kan konkludere ut fra den spalteplass den får), men utviklingen er kommersielt drevet. Humanistene plukker smuler som så flikkes på for å imøtekomme deres behov. Ramberg ville ønske at humanistene i langt sterkere grad enn i dag kunne bidra med sin kompetanse når standardene ble definert. Dessverre opplever vi at humanistene til tider lar seg blende av teknisk briljanse. Også humanistiske fagmiljø har arrangert seminar over multimedia, der teknikken stod i sentrum uten noen pedagogisk og faglig referanseramme. Det ideelle ville være om pedagogikk, fag og forskningskompetanse kunne spille sammen. Hvordan kan vi komme på offensiven med vår kompetanse? Hvordan kan Senteret og universitetenes edb-tjenester hjelpe?

## **REKRUTTERING OG ALFABETISERING**

---

Generelt ble det oppfattet som viktig at Senteret synliggjør seg mer i brukermiljøene, bl.a. ved at medarbeiderne besøker disse miljøene og demonstrerer maskin- og programvare for å skape interesse for edb.

*Julie Feilberg*, Inst. for anvendt språkvitenskap, UNIT-AVH, tok utgangspunkt i to grupper brukere (eller potensielle brukere) – de vitenskapelig ansatte og de edb-kyndige. Hun ga følgende beskrivelse av en gjennomsnittlig vitenskapelig ansatt: Han kan WordPerfect, har kanskje et WordPerfect-kurs, har kanskje kjøpt manualen, men kjenner ikke andre tekstbehandlingsprogrammer. Han kjenner f.eks ikke idébehandlingsprogrammer som Framework, PC-outline og MORE. Han vet ingenting om presentasjonsprogramvare, og lager sine transparenter med tusj. Han kjenner ingen programmer for datastøttet læring innen sitt fag og vet ikke hvordan slike programmer fungerer. Han vet ikke hva multimedia betyr og hvordan denne teknologien brukes. Han vet ikke hvordan databaser brukes. Nye generasjoner av studenter har et visst forhold til edb, og vil f.eks gjerne bruke edb i oppgaveskriving. Mange ansatte viker imidlertid unna.

Feilberg så et behov for å informere denne gruppen, og for å vise den hvilke hjelpemidler som finnes, f.eks video til bruk i undervisningen. Det er imidlertid vanskelig å få gruppen med. Feilberg ønsket derfor hjelp til å inspirere, og hun ønsket en oversikt over hvem som gjør hva innen de enkelte fag. Hun så gjerne at en historiker, en musikkviter, etc kunne vise hvordan disse bruker edb i det daglige. De ansatte må

informeres om utstyr og verktøy det ikke tar lang tid å sette seg inn i og få nyttige resultater ut fra. De ansatte må se den umiddelbare nytten i å bruke den nye teknologien.

For å makte en slik oppgave må Senteret imidlertid rydde mye av dagens aktiviteter unna. Annen informasjonsvirksomhet vil måtte lide. Senteret skal likevel ikke ha alt ansvar for at relevant informasjon kommer ut. *Jan Oldervoll*, Inst. for samfunnsvitenskap, UiTø, mente at edb-tjenestene må stå for mye av informasjonsformidlingen.

*Bjørn Sørenssen*, Inst. for drama, film og teater, UNIT-AVH, påpekte en annen bekymringsfull situasjon, nemlig et gap mellom studentene og de lærere som ønsker å bruke teknologien i undervisningen. Studentene er ikke slike datafreakere som mange tror. Ofte begrenser deres kunnskap seg til tekstbehandling. Han etterlyste en edb-alfabetiseringskampanje også for studentene.

## **KUNNSKAPSOMRÅDER**

---

Flere deltakere fremhevet bestemte fagfelt eller kunnskapsområder. Mange mente at Senteret burde fungere som et nasjonalt kompetansesenter innenfor multimedia/hypertekst/hypermedia/interaktiv video, og innenfor bildeprosessering. Eksempelvis var det en deltaker som foreslo at Senteret laget en multimedia-plate for å sette denne teknologien på dagsorden. En av Senterets edb-konsulenter foreslo da *Snorres Kongesagaer* med lyd, bilder og evt. video som emne for en CD-ROM. Materiale kunne hentes fra andre historiske kilder, skuespill og hørespill. En slik plate kunne selge bra også utenfor spesialistenes rekke.

Lyd og tale ble nevnt som eksempel på et område der Senteret ikke har satset så mye som ønskelig, mens Senteret allerede har gjort en god del arbeid innenfor OCR – Optical Character Recognition – innscanning og automatisk gjenkjenning av tekst.

Det ble en viss diskusjon om Senteret skulle arbeide med pedagogikk. Det ble hevdet at forskning og undervisning – innhenting og formidling av informasjon – var nært forbundet. Senteret har ellers arbeidet lite med de pedagogiske sidene av edb-bruken. Unntaket er at Senteret huser en forsker-NAVF som er i ferd med å fullføre et undervisningsopplegg basert på interaktiv video.

Fjernundervisning er også et arbeidsfelt der det er et åpent spørsmål om Senteret bør sette inn ressurser.

## **INFORMASJONSFORMIDLING**

---

Blant forslagene til konkrete oppgaver som Senteret burde ta seg av, var det mange som gjaldt informasjonsformidling.

Deltakerne merket seg med interesse at Senteret nå har lagt til rette elektroniske postsystem og oppslagstavler der det kan knyttes kontakt mellom brukere i inn- og utland. Senteret har operativ en filtjener for engelsk språkmateriale, programmer for språkbehandling og informasjon om Senteret som benyttes av forskere fra hele verden. Senteret ser klart behovet for å få ut informasjon om slike opplegg for elektronisk post. Etterhvert har mange fått nødvendig utstyr og programvare for kommunikasjon, og tiden er kanskje moden for mer utstrakt bruk av slike system.

*Richard Pierce*, Klassisk institutt, UiB, mente at noe av det viktigste med edb var muligheten for å komme seg ut av "humanioras lukkede verden" og få impulser fra forskere som arbeider med liknende problemstillinger innenfor andre fag. Han tenkte på alle de utgravninger som ikke er publisert, og så fram til en ny form for publikasjon, der denne informasjonen gjøres tilgjengelig over nettet.

Han ønsket større mengder informasjon lagret sentralt og tilgjengelig for forskeren. Han ønsket seg tilgang til elektronisk lagrete innholds-oversikter fra tidsskrifter på mange fagfelt, også ikke-humanistiske. Hele artikler burde også være tilgjengelig elektronisk. Til alt dette trengs det en organisasjon. Han var oppmerksom på de økende muligheter det er for å ta store informasjonsmengder inn lokalt på sin egen maskin.

Et annet forslag som gjaldt informasjonsspredning, gikk ut på å gi fagpersoner mulighet til å være ved Senteret en viss tid for å tilegne seg kompetanse som de kunne bringe tilbake til fagmiljøene. I den grad dette vil gjelde frikjøp av forskere, er det kanskje ikke Senterets oppgave å formidle slike ønsker overfor NAVF. En annen mulighet er å engasjere stipendiater. Nå har Senteret en ordning for forskere/NAVF som knyttes til Senteret i 1-3 år, men kortere stipendopphold ved Senteret ville også være nyttig.

Pierce foreslo at Senteret skulle hente inn som stipendiater egnede personer fra miljøer med behov, for så å sende dem dit den nødvendige kompetansen ligger, det være seg Nansen-senteret, Multimediaboratoriet ved UNIT-AVH eller andre steder. Etterpå må personene tilbake til Senteret og formidle erfaringene og kunnskapen via filserveren. Det skal altså være en rundreise fra hjemsted via "lærested" til formidlingssted.

## **PRODUKT OG TJENESTER**

---

Det ble også foreslått at Senteret skulle gjøre mer for å utvikle faste produkt og tjenester: digitaliserte kart, databaser, bibliografiske databaser og dataleverandørvirksomhet à la Norsk samfunnsvitenskapelig datatjeneste. Senteret burde også se oppbyggingen av et norsk tekstkorpus som et satsingsområde.

## UTSTYRSPARKEN

---

En del av innleggene omhandlet utstyrsparken. Humanistiske forskere får nå et økende behov for utstyr. Studentene må også være med i denne utviklingen og må få adgang til maskinvare og opplæring.

Jan Oldervoll ønsket seg en historisk arbeidsstasjon – hvor forskeren kan arbeide med varierte data og mange typer programvare. Mye kommersiell programvare kan brukes i historiefaget, men ikke alltid uten problemer. Oldervoll er med i en europeisk gruppe som skal vurdere hva som kan brukes og hva som må utvikles. Hypertekst og hypermedia kan brukes, men analysedelen av slike systemer er altfor primitiv. Blob (binary large object – en type databasefelt som kan lagre vilkårlige, store dataobjekter, f.eks digitaliserte bilder og kart) kan brukes til å lagre andre objekter i vanlige databaser, men skikkelig programvare er egentlig ikke på markedet ennå. Jan Oldervoll mente kommersiell programvare ikke kan dekke alle historikernes behov. Dette er ikke bare fordi historikerne er spesielle (for det innrømmet han at de er), men også fordi historiske *data* er spesielle.

Oldervoll påpekte også behovet for feiltolerante systemer, dvs systemer som tåler data med feil og unøyaktigheter. Dagens systemer er ikke feiltolerante nok. Han så behovet for en annen logikk enn den som stort sett gjennomsyrrer datasystemer. Han har selv blitt glad i fuzzy-logikk, men mener den er vanskelig å legge inn i systemer. (Fuzzy-logikk: F.eks at et utsagn ikke nødvendigvis må være sant eller usant, men kan ha en verdi et sted i mellom.)

En stor grad av integrasjon er nødvendig pga at historikeren ønsker mange kilder på en gang. Nettverk er også viktig. Historikere blir etter hvert avhengige av nasjonale og internasjonale nettverk. Det er nødvendig å ha en institusjon med spisskompetanse. Bildebehandling og multimedia er områder av felles interesse for hele spektret av humaniora.

Interaktiv video (IV) er en investeringstung aktivitet, der utstyret dessuten ikke har blitt særlig mye billigere med årene. Kostnadene fører til at det bare har blitt gitt "klattbevilgninger" til IV-tiltak. Senteret fikk ros for sin innsats for å introdusere IV. Det ville kanskje være mulig å oppnå innsparing nasjonalt hvis Senteret ble et laboratorium for utprøving av ulike teknologier, slik at ikke alle miljøene må investere i dyrt utstyr. IV gir interessante muligheter, men det er usikkert hvilken teknologi som vil bli dominerende.

Lise Opdahl, Engelsk institutt, UiB, mente at undervisningen i fremmedspråk vil forandre seg svært mye. De første fem årene vil undervisningsprogrammene ha en støttefunksjon, og endringen vil ikke være så radikal, men i løpet av en tiårsperiode vil arbeidsmåte og evaluingsformer bli så mye endret at lærerkrefter vil bli frigjort fra rutine-

arbeid. Opdahl så til og med for seg at studenter kunne levere inn disketter ved eksamen, og at et program kunne gi besvarelsene et automatisk evalueringsforslag.

Opdahl kritiserte dagens språklab. Lærere med store grupper får for kort tid til å veilede hver student, og studentene kan risikere å bruke tiden på å øve inn feil. Hun ønsket seg digitale lab'er med lærerens eller en innfødts stemme. Det finnes allerede programvare som kan kontrollere og korrigere studentens uttale. For døve og tunghørte finnes systemer som gir tilbakemelding på om vokaler er riktig uttalt.

Det kom også et forslag om at Senteret kunne låne ut maskiner og programvare, men en slik oppgave må i tilfelle vurderes av Senterets styre/RHF.

### **NASJONALT SENTER MED SPISSKOMPETANSE**

---

Flere av deltakerne mente det var ønskelig med et Senter som kunne fungere som "lokomotiv" og utvikle spisskompetanse på bestemte områder over kort tid. Det ble hevdet at Senteret tidligere hadde spredt ressursene over for mange felt. Det måtte – med en spissformulering – "være riktigere å sette 3 personer på en sak i 2 år". En slik satsing kunne gi resultat.

*Jan Oldervoll* ville ikke se på Senteret som en institusjon som skal produsere ting. Da ville Senteret bli halvgodt. Et godt Senter er mer et slags grensesnitt. En fornuftig bruk av edb i historie forutsetter en *historiker*, men denne har bruk for bl.a databaseteknikk, multimedia og interaktiv video, og gjerne tunge teknikker som bildeanalyse, semantisk analyse, språkoversettelse, osv. I en historisk arbeidsstasjon trengs alle disse komponentene, men problemet er å vite hvor de finnes. Hvem har tid til å finne ut, og å hjelpe?

Problemet med Senteret i dag er at de er for tunge på edb-siden og for lette på den faglige. Oldervoll ønsket at Senteret skulle være et forum hvor en kan møte personer som ikke bare har edb-bakgrunn, men også fagfolk, som historikere.

Hvis Oldervoll hadde disponert Senterets budsjett, ville han brukt det slik at det for det første kunne være til stede en del fagfolk på halv til hel tid. For det andre ville han ha en god del edb-folk som støttespillere til fagfolkene. Edb-folkene måtte makte både å holde seg orientert og å kunne viderefremde kunnskapene.

Han ville imidlertid kutte ut "brød-og-smør"-oppgaver, som han sa. Andre kan ta seg av problemer på lavere nivå. Senteret skulle virkelig bli et spiss-senter. En forutsetning er imidlertid en bredde på fagfolkene en får inn. Oldervoll ville med en slik modell vente at 1/3 av prosjektene ble vellykket, og mente at det var bra! Mye mer ville komme ut av

de midler som ble investert på denne måten enn tilsvarende investeringer ellers i forskningsverdenen.

Noen av deltakerne mente det var et behov for å avklare ansvarsfordelingen: når kompetansen var kjent, burde Senteret kunne overlate ansvaret for den videre formidlingen til HF-fakultetenes edb-seksjoner og de enkelte institutt. Senteret burde alltid kunne være i forkant med å utprøve ny teknologi og fortelle om den, men deretter legge oppgaver bak seg og overlate til fagmiljøene å anvende kunnskapene.

Avslutningsvis sa direktør Hauge at det antakelig ville bli nødvendig å redusere Senterets arbeidsområder og konsentrere innsatsen på færre felt. Ved dette kan gode sider ved Senterets virksomhet bli skadelidende for at de beste skal styrkes.

# SEMINAR OM LENKING AV NOMINATIVE DATA

---

---

*Øystein Reigem*

29.–31.10. 1990 ble det holdt et *Seminar om lenking av nominative data* ved Riksarkivet, Oslo. Arrangør var *Gunnar Thorvaldsen*. Seminaret var lukket med inviterte deltakere. Totalt deltok 16 personer. En bevilgning fra RHF's informasjonsteknologiprogram dekket utgiftene til seminaret, inkludert reise- og oppholdskostnadene for de tilreisende.

## LENKING OG NOMINATIVE DATA

---

Nominative data vil si kilde-data hvor enkeltindivider lar seg identifisere ved navn. Typiske historisk-nominative data er folketellinger, kirkebøker og emigrantprotokoller.

Lenking vil si å kombinere informasjon om enkeltpersoner eller grupper av personer fra ulike kilder eller fra ulike innførsler i samme kilde, for dermed å kunne rekonstruere de enkelte individers eller grupperes livsløp. Eksempelvis vil en historiker kunne ønske å rekonstruere livsløpet for de enkelte personer i et sogn ut fra opplysninger i kirkebøker, folketellinger, emigrantprotokoller, matrikler, osv. Et beslektet problem er *kobling*, der en ønsker å finne relasjoner mellom ulike personer, så som slektskapsforhold.

## SEMINARET

---

Seminaret tok utgangspunkt i det arbeidet som gjøres på flere felter i Norge. Hovedsiktemålet var imidlertid å hente inn kunnskaper og impulser utenfra for å oppdatere historikere og informatikere mht status på området internasjonalt. Nøkkelpersonen i så måte var *Gérard Bouchard* fra SOREP, Centre Interuniversitaire de Recherches sur les Populations, Université du Québec à Chicoutimi, Canada. Ved SOREP samarbeider tre universiteter i Quebec-provinsen. Her anvender en avanserte teknikker for automatisk lenking, en legger opp data i relasjonsdatabaser, og



forskningen er tverrfaglig med et sterkt innslag av sosialhistorie. Bouchard er direktør for programmet, og sammen med professor *Hans Christian Johansen* fra Historisk Institut, Odense Universitet, fungerte han som hovedforeleser på seminaret.

Første dag var ment som en introduksjon til temaet lenking. Fokus var på igangværende lenkingsprosjekter i Norge og Norden. Dagen ga en bakgrunn for seminarets mer internasjonalt orienterte del. Andre dag ble innledet av Johansen, og så var Bouchard gitt resten av dagen. Bouchard fortsatte tredje dag, og seminaret ble avsluttet med en paneldebatt.

Jeg konsentrerer meg om Bouchard og Johansens redegjørelser for lenkingsproblematikken og deres syn på hvordan problemet best angripes. Et tema som bare indirekte berøres, er hvordan SOREP driver sin virksomhet, noe som kunne fortjent en egen artikkel. Men aller først er det på sin plass med en rask innføring i temaet lenking:

## **LENKING**

---

Som et enkelt eksempel på lenking kan en tenke seg en kombinerings av individopplysninger fra en folketelling og en emigrasjonsprotokoll. Da må en ta for seg hver innførsel i den ene kilden og prøve å finne en innførsel for samme person i den andre.

For den uinnvidde kan dette høres ut som en likefram sak, men ofte er det vanskelig eller umulig å avgjøre om to innførsler gjelder samme individ. Historiske kilder inneholder mange feil, unøyaktigheter og utelatelser. Eksempelvis kan det i navneopplysningene om en og samme person være store variasjoner fra kilde til kilde og fra innførsel til innførsel. Et annet eksempel på en problematisk opplysningstype er alder, som ofte er svært upålitelig. Det at kildene gjerne er svært spredt i tid, kompliserer også lenkingen.

I sitt innlegg kom *Eli Fure* fra Avdeling for historie, UiO, med en del konkrete eksempler på feil og problemer i et kildemateriale hun arbeidet med. Hun hadde støtt på navnesubstitusjoner som "Monsdatter" i stedet for "Hansdatter", "Johanne" i stedet for "Johannes" og "Johan" for en person som ble døpt "Karl Anton". Overgangen fra patronymikon ("Hansen", "Hansdatter") til fast etternavn skapte mange vansker i arbeidet hennes. Navn som "Hansen" har vært begge deler. Personer med dobbeltnavn var ofte oppført med bare ett av navnene. Uekte barn var ofte problematiske, både når det gjaldt navn og andre opplysninger. En uekte datter av far kunne være ført opp som pleiedatter. Feil person kunne være ført opp som begravet. Alder var ofte feil. Dupliseringer forekom, til og med av hele familier i folketellinger og dåpsinnførsler i kirkebøker.

Lenking kan gjøres med rent *manuelle* metoder, og det har historikere, lokalhistorikere, bygdeforskere og slektsgranskere lenge gjort. Kanskje mer interessant er lenking med *maskinelle* metoder. Her finnes det to ulike hovedstrategier. Den ene er *interaktiv* lenking, som bringer kildene opp på dataskjermen. Lenkingsprogrammet kan gjøre analyser og komme med forslag til lenker, men avgjørelsen om hvilke lenker som virkelig skal knyttes, overlates til operatøren. Interaktiv lenking har derfor mange likheter med manuell lenking, og mange av dennes sterke og svake sider. I *automatisk* lenking er det programmet selv som knytter lenkene ut fra ulike kriterier.

Både ved automatisk og manuell lenking vil en måtte regne med lenkingsfeil. Selv habile (manuelle) lenkere gjør feil. Generelt er det to typer feilsituasjoner som kan oppstå – *overlenking* og *underlenking*. Overlenking vil si at det etableres flere lenker enn det er belegg for ut fra kildematerialet. Ved overlenking blir opplysninger om flere personer satt sammen til én person, og en får en falske livsløp. Underlenking vil si at mange lenker som burde vært etablert, ikke blir det. Enkelt personer figurerer som flere personer, og med sparsomme opplysninger. I et underlenket materiale blir representativiteten i det lenkede materialet i forhold til populasjonen dårlig. Begge typer feil har negative konsekvenser for forskning på materialet.

Det er mest nærliggende å ha *enkeltpersoner* som *lenkingsenhet*, og det er da også mest brukt. Andre alternativer er *grupper av personer* som ektepar eller far-barn-par. Fordelen med grupper er at en får et rikere sett av lenkingskriterier. En ulempe er at de forhold en har valgt ikke alltid er stabile. I en del sammenhenger er det gunstig å bruke *bosted* som enhet.

Alternativ til lenking er å analysere kildemateriale med statistiske metoder på høyt abstraksjonsnivå. Slike metoder har lenge vært i bruk i historieforskningen og er fortsatt aktuelle. Vår tids historikere har imidlertid stor interesse for enkeltindividene. Forskerne er i dag opptatt av å finne ut hva som ligger bak de enkeltes valg. Som et verktøy for sosiodemografisk livsløps- og familierekonstitusjon er dermed lenking svært aktuelt for historikerne i dag.

Lenking er også aktuelt i andre fag. Innen epidemiologi og arvelighetsforskning er det gjennomført store lenkingsprosjekter.

## **AUTOMATISK LENKING**

---

Ved automatisk lenking av et materiale må en bestemme hvilke kriterier en vil bruke for lenkingen. (For så vidt er det prinsipielt og metodologisk viktig å ha klare kriterier og regler ved manuell lenking også.) En velger på forhånd ut visse felt, slik som navn, fødselsår/alder, kjønn,

osv, og så gjør en sammenlikninger og beregninger på disse feltene for å bestemme hvor like eller ulike kandidatene for lenking er. Resultatet av beregningene er en poengangivelse eller skåre. En høy skåre indikerer at innførslene omhandler samme individ.

For å unngå å måtte sammenlikne alle mulige par av kandidater, må materialet først ordnes, slik at sannsynlige kandidater samles i undergrupper. Det kan f.eks gjøres ved å sortere på en "standardisert" versjon av navn, slik at en tar hensyn til naturlig forekommende variasjoner.

For selve sammenlikningen kan det her være illustrerende å vise kriterier og regler for "matchskåring" brukt i en britisk undersøkelse på et italiensk 1800-tallsmateriale (Kitts et. al. – se litteraturlisten). For hver variabel, så som navn, andre navn (slektninger, ektefelle), fødselsår, osv, er det ulike regler for poengberegning. Disse enkeltskårene multipliseres med hverandre for å gi den endelige skåren for de to individinnførslene:

|                                 |  |         |       |
|---------------------------------|--|---------|-------|
| <i>navn</i>                     | alle navn identiske:   | 3       | poeng |
|                                 | noen navn samme:   | 2       | poeng |
|                                 | ingen felles navn:   | 0       | poeng |
| <i>andre navn</i>               | alle identiske:  | 3       | poeng |
|                                 | mangler i den ene innførselen:                               | 2       | poeng |
|                                 | ingen felles.  | 0       | poeng |
| <i>(utregnet)<br/>fødselsår</i> | 0-2 års forskjell:   | 3       | poeng |
|                                 | 3-5 år:  | 1       | poeng |
|                                 | 6-9 år:  | 0,5     | poeng |
|                                 | 10-15 år:  | 0,3     | poeng |
|                                 | 16-24 år:  | 0,2 (?) | poeng |
|                                 | > 24 år:   | 0       | poeng |
| <i>kjønn</i>                    | samme kjønn:   | 1       | poeng |
|                                 | forskjellig kjønn:   | 0       | poeng |
| <i>yrke</i>                     | samme yrke:  | 3       | poeng |
|                                 | yrke innen samme hovedgruppe eller samme sosiale gruppe:     | 2       | poeng |
|                                 | mangler yrkesopplysninger i den ene eller begge innførslene: | 1       | poeng |
|                                 | ulik hovedgruppe eller sosial gruppe:                        | 0,3     | poeng |
| <i>sivilstatus</i>              | identisk sivilstatus:  | 1,5     | poeng |
|                                 | ulik sivilstatus, men mulig kombinasjon:                     | 1       | poeng |
|                                 | umulig kombinasjon av sivilstatus:                           | 0,1     | poeng |
| <i>fars patronym</i>            | identiske patronym:  | 5       | poeng |
|                                 | patronym mangler i den ene innførselen:                      | 1       | poeng |
|                                 | forskjellige patronym:                                       | 0,1     | poeng |

De enkelte reglene/skårene må bestemmes ved å studere materialet på forhånd, og reglene vil ikke kunne overføres direkte til et annet materialet.

Før selve sammenlikningen gjøres det ofte en "standardisering" for å slå sammen naturlig forekommende stavemåter av navn, og for å formalisere andre opplysninger. Denne standardiseringen og den som gjøres for å organisere materialet i undergrupper, har ulike formål og kan være ganske forskjellige.

For å vise hva standardisering av navn for sammenlikning vil si i praksis, kan vi her se eksempler på resultatet av en (noe hardhendt) navnestandardiseringsalgoritme som benyttes i et norsk prosjekt. Lars Nygaard fra Institutt for informatikk, UiO, har utviklet et interaktivt lenkingssystem, og systemet brukes av Eli Fure:

Alle de følgende fornavnene blir omgjort til standardformen "EMBRET":

|                  |                  |                  |
|------------------|------------------|------------------|
| <i>Ambret</i>    | <i>Engebrigt</i> | <i>Ingebreth</i> |
| <i>Embret</i>    | <i>Ingebert</i>  | <i>Ingebrigt</i> |
| <i>Engbret</i>   | <i>Ingebredt</i> | <i>Ingebrit</i>  |
| <i>Engebret</i>  | <i>Ingebrekt</i> |                  |
| <i>Engebreth</i> | <i>Ingebret</i>  |                  |

Alle de følgende patronymikon blir omgjort til standardformen "PETER":

|                     |                 |                  |
|---------------------|-----------------|------------------|
| <i>Peders.</i>      | <i>Pehrs.</i>   | <i>Petersen</i>  |
| <i>Pedersd.</i>     | <i>Pehrsd.</i>  | <i>Peterson</i>  |
| <i>Pedersdatter</i> | <i>Pers.</i>    | <i>Petters.</i>  |
| <i>Pedersen</i>     | <i>Persd.</i>   | <i>Pettersd.</i> |
| <i>Pederson</i>     | <i>Person</i>   | <i>Petterson</i> |
| <i>Peers.</i>       | <i>Perss.</i>   | <i>Petterson</i> |
| <i>Peersd.</i>      | <i>Persson</i>  | <i>Pæærds.</i>   |
| <i>Peerson</i>      | <i>Peters.</i>  |                  |
| <i>Peersson</i>     | <i>Petersd.</i> |                  |

## **HANS CHRISTIAN JOHANSEN**

---

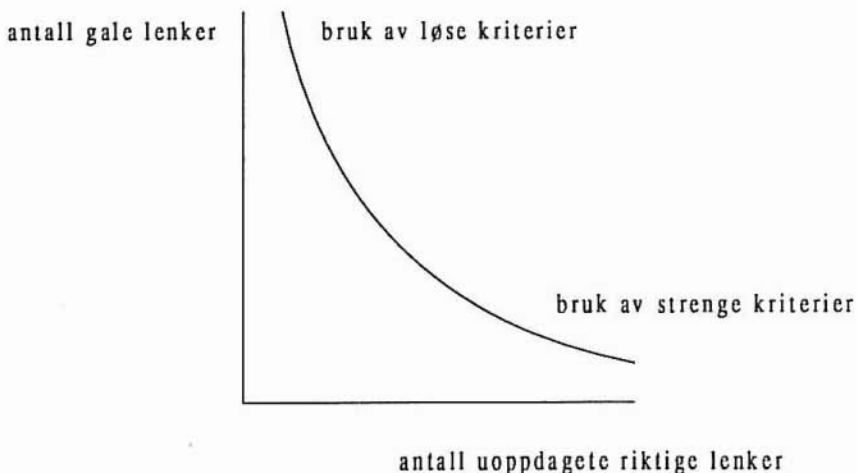
Hans Christian Johansen innledet seminarets andre dag med et foredrag om lenking på nordisk kildemateriale. Han startet med å reformulere tidligere deltakeres uttalelse "om vi trenger automatisk lenking" til "når vi trenger automatisk lenking".

Johansen så flere viktige ulemper med manuell (og interaktiv) lenking: Manuell lenking er arbeidskrevende. En kan ikke etterpå dokumentere

hva som er gjort. Prinsippene for arbeidet kan endre seg underveis, og skjevheter introduseres i materialet.

Dokumenterbarheten av lenkingsprosessen var ellers et tema som opptok mange av seminardeltakerne. Det var bred enighet om at dette var viktig.

På nyere kilder med personnummer som nøkkel er automatisk lenking en enkel prosess. På eldre kilder, derimot, må en gå varsomt fram. En kan lett få fullstendig ulike resultat med forskjellige lenkingskriterier. Johansen illustrerte situasjonen med følgende diagram:

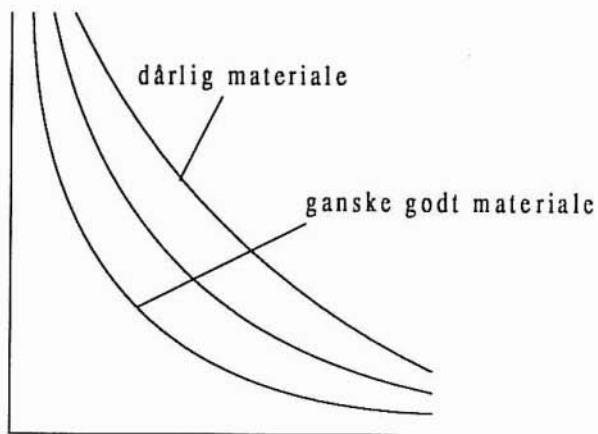


Ved bruk av strenge kriterier vil en få uopdagete riktige lenker. Ved bruk av løse kriterier vil en få feilaktige lenker. Det er viktig å finne den rette balansen.

Dette var også en problemstilling som gang på gang dukket opp i løpet av seminaret. Spesielt fra de utenlandske forelesernes side ble det understreket at avveiningen måtte gjøres ut fra formålet med forskningen. Det ulike utgangspunktet for demografiske og genetisk baserte studier illustrerer dette godt: Demografen er opptatt av materialets representativitet og vil foretrekke et overlenket materiale, mens epidemiologen og arvelighetsforskeren er mer avhengig av at de enkelte lenker er pålitelige.

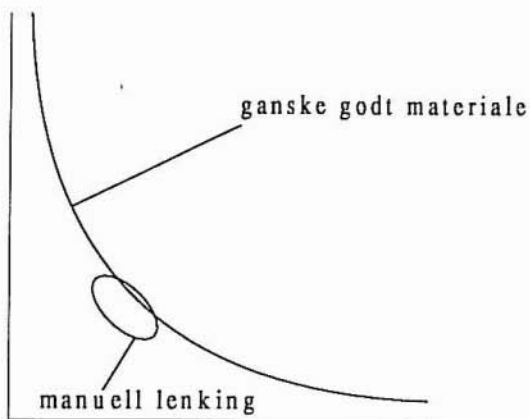
Antall feil vil naturlig nok avhenge av kvaliteten på materialet, noe Johansen illustrerte slik: (Se neste side)

antall gale lenker



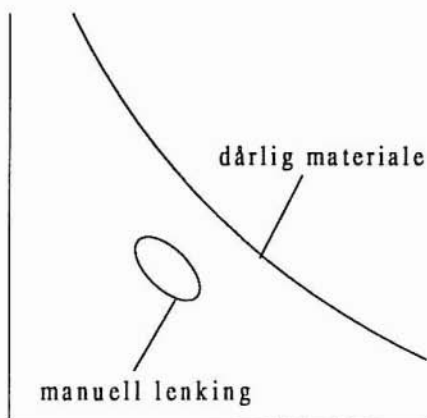
antall uoppdagete riktige lenker

Ved manuell lenking på godt materiale vil en gjerne havne rundt midten av kurven, og kanskje med noe bedre resultat, dvs færre feil av begge typer.



Siden automatisk lenking er raskere, og siden en har en viss mulighet til å velge hvor på kurven en ønsker å havne, ville Johansen velge

automatisk lenking på et slikt materiale. På et dårlig materiale ville kvaliteten kunne bedres forholdsvis mer med en manuell lenking:



Her ville Johansen valgt en kombinasjon av automatisk og manuell lenking.

Johansen tror at automatisk lenking ville gi like god kvalitet som manuell på (dansk) materiale etter 1830, mens tidligere materiale, spesielt før 1780, ville være mer problematisk. Selv hadde han brukt metoder tilsvarende dem som Fure/Nygaard benytter, på 1700-tallsmateriale.

### **GÉRARD BOUCHARD**

---

I et foredrag på seminarets andre dag gjorde Gérard Bouchard en grundig gjennomgang av viktige sider av lenkingsproblematikk og -arbeid:

#### ***– Strategi for datainnsamling***

---

Det er best å arbeide på hele populasjoner, som fylke (county) eller region. Ikke alle vil ha råd til kostnadene med dette, og en må da arbeide på utvalg, f.eks sogn. Flytting er problematisk for slike geografisk små utvalg. Det kan hjelpe å arbeide på flere sogn. Mange forskere er opptatt av representativitet, men et sogn er i det minste representativt for seg selv. Noen har brukt en utvalgsstrategi som har gjort at en



likevel kunne arbeide på f.eks nasjonalt nivå. F.eks tok en i et prosjekt i Rouen for seg alle familier med navn på "B", og Bouchard anså prosjektet for vellykket. Ved slike bokstavutvalg er det svært viktig å kompensere for skjevheter.

### ***– Dataevaluering***

---

En må vite hva slags analyser som vil være mulig på de aktuelle data, og også hvilke lenkingsmetoder en kan bruke. Disse to problemstillingene er nær beslektet. Han henviste til Johansens kurver og sa at en vanligvis må velge mellom de to typene feil.

### ***– Kontroll ved dataregistrering***

---

Bouchard så kontroll ved dataregistrering som et viktig punkt, om enn av mindre omfang enn andre problemer. Kostnadene ved hver feil er store. Ved SOREP benytter en dobbeltregistrering og dobbelt korrekturlesing. Arbeidslagene ruller mellom de ulike oppgavene. Både de ulike typer feil og feilraten varierer med ytre forhold, som f.eks været. Ved SOREP gir en registratorene av og til en ekstra fridag og høster gevinsten i form av en 3–4 gangers kvalitetsøkning etterpå.

### ***– Behandling av navnevariasjon***

---

Det finnes ulike typer navnevariasjoner. Én type er stave-/ortografiske varianter. Disse endrer ikke det fonetiske og kan standardiseres før lenking. En bør ikke slå fonetisk ulike former sammen før lenking. Standardiseringen må styres av et sett spesifikke regler slik at standardiseringen blir uniform. Ved SOREP har en selv laget verktøy med 75 regler.

En annen type navnevariasjoner er fonetiske varianter, men som sagt bør håndtering av disse utsettes til selve sammenlikningen.

Den siste typen er navnesubstitusjon, hvor helt ulike navn forekommer i stedet for hverandre. Det er meget vanlig med slike substitusjoner, og en person som ikke har arbeidet med historisk-nominativt materiale vil neppe tro hvilken variasjon som faktisk forekommer.

### ***– Nødvendigheten av å matche alle potensielle kandidater***

---

En kan ikke sjekke alle mulige par av kandidater mot hverandre. Følgende strategi må brukes: Først lage undergrupper ("pockets") som teoretisk inneholder alle kandidater for lenking. Så må en sjekke alle mulige par innen hver undergruppe. Det er også viktig å ha en prosedyre for å få tak i dem som har unnsloppet grupperingen.

### ***– Hvilke felt som skal brukes som identifikasjon***

---

Noen variabler blir kriterier for lenking, og det er viktig å velge riktig. Disse variablene kan etterpå ikke brukes for analyse, for det vil gi sirkulære resonnement. Eksempelvis vil det i en lenkingsprosess der bosted inngår i lenkingskriteriene, velges lenker som gir en stabil befolkning. På et slikt materiale kan en ikke studere mobilitet. Bouchard ville selv ikke bruke yrke eller bosted som kriterium.

### ***– Sammenligne navnedata***

---

En må prøve å unngå underlenking. Det er viktig med et egnet verktøy for sammenlikning. Bouchard nevnte *Soundex*, en mye brukt standardiseringskode, men denne mente han på den ene siden var for hardhendt for navnestandardisering før lenking, og på den andre siden ga for lite variasjon ved sammenlikningen. Ved SOREP brukes en algoritme som sammenligner navn tegn for tegn med hensyntagen til utelatelser og annet. Algoritmen gir en skåre fra 0–100 som sier hvor ulike eller like navnene er.

### ***– Garantere koherens og uniformitet av lenkene***

---

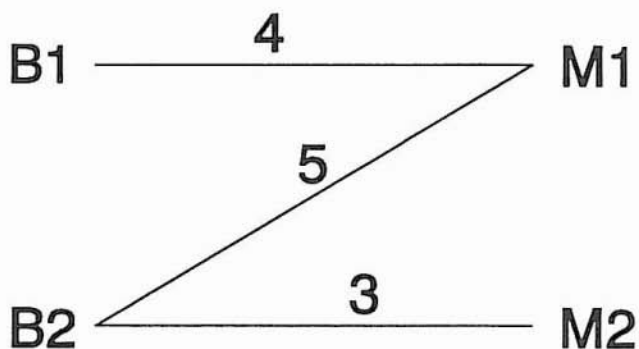
En må sikre at lenkingen følger samme mønster for hele materialet, og at en nytter samme valg av "cutoff-point", dvs skranke for hva som skal godtas som likhet. Hvis flere personer arbeider med lenkingen, kan det være vanskeligere å få et uniformt resultat. En må ha detaljerte retningslinjer. Automatisk lenking er en fordel her.

### ***– Lenkingsmetoden må kunne løse opp i klynger***

---

En må ha en metode for å takle tvetydigheter og ambivalens. I et materiale vil det ofte oppstå klynger ("clusters") av konkurrerende potensielle lenker. Et eksempel på en klynge er følgende, hvor opplysninger fra fire kirkebokinnførsler ikke entydig kan lenkes: (Se neste side)

Her er det snakk om minst to personer fra to ulike giftermål (M1 og M2) og to ulike dåpsbegivenheter (B1 og B2). Muligens er M1 og B1 samme person, likeledes M2 og B2, men det er også en sterk sannsynlighet for at M1 er B2, og at det dermed er tre personer totalt. Klynger kan bli store og kompliserte og en må ha en metode for å løse opp disse. I noen tilfeller er én lenke sterkest, i andre har flere lenker høyeste skåre. Det finnes to hovedfilosofier: Den ene går ut på å lage så mange lenker som mulig, og en velger den løsning som gir



best skåre for klyngen totalt. Da beholder en ikke nødvendigvis de sterkeste lenkene. Den andre filosofien går ut på å beholde den sterkeste lenken. Begge filosofiene er gode, og hvilken en velger er avhengig av formålet med den senere analysen. I eksemplet over ville de to filosofiene gi ulikt resultat.

*– Evt inkludere ikke-nominative felt for teknisk sjekk*

---

Det kan være ønskelig å holde noen ikke-nominative felt (andre felt enn navnefeltene) utenfor lenkingen for å spare dem til en teknisk kontroll etterpå. Eksempler på felt som kan brukes er alder ved død og ved giftermål. Hvis en tok disse feltene med under selve lenkingen, kunne noen gode lenker bli kuttet ut, og resultatet ville bli underlenking. Her er det igjen avhengig av hva en er mest redd for – over- eller underlenking.

*– Vurdere kvaliteten på lenkingsarbeidet*

---

For å kunne vurdere kvaliteten på det endelige resultatet, må en gå til andre kilder. Det kan hende at slike kilder ikke finnes. Hvis eksterne kilder finnes, og en korrigerer materialet ut fra disse, må en være oppmerksom på faren for å introdusere skjevheter. Hvis en f.eks har data som spenner over et lengre tidsrom, vil det ofte bli flere rettinger i den nyeste delen av data. Hvis en bruker yrkesopplysninger, kan en introdusere stabilitet. Vurdering er vanskelig, kostbart og tidkrevende, men nødvendig.

## ***– Om bruk av folketellingsdata***

---

Folketellingsdata er ikke longitudinelle, men viser tverrsnitt av befolkningen på bestemte tidspunkt, og det er dermed mye som ikke kan leses ut av folketellingsdata. Et nytt ektepar som dukker opp i en telling kan f.eks ha giftet seg siden forrige telling, men de kan også være innflyttere. Bruk av folketellingsdata bør derfor sjekkes mot kirkebøker o.l. Alle undersøkelser av geografisk mobilitet basert på folketellingsdata, har svakheter. En har ikke kontroll mot opplysninger om dødsfall, og dermed blir omfanget av emigrasjon lett overestimert, muligens i alvorlig grad. Kjernefamilier versus enslige er også et problem. En kan lettere lenke familier enn enslige, og skjelheter introduseres. Alder i folketellingsdata er ofte feil – i mye større grad enn en utenforstående ville tro.

## ***– Organisering av data***

---

Det er nødvendig å strukturere data, og i dag finnes mange kommersielle databasesystemer som kan brukes. Implementering på stormaskiner er ofte tungt, og her byr personlige datamaskiner på fordeler. En kan imidlertid ikke regne med å finne det perfekte system.

## ***– Personvern***

---

Personvernsspørsmålet kan dukke opp, gjerne avhengig av hvilket tidsrom data stammer fra. F.eks er data etter 1890 ikke tilgjengelige i England. (Michael Drake i kommentar etterpå: Situasjonen er kaotisk i England. Alle kilder unntatt folketellinger kan skaffes, men en må betale.) Som en kuriositet nevnte Bouchard et tilfelle fra Frankrike, hvor etterkommerne av en adelig familie gikk til rettssak da de følte at offentliggjøringen av opplysninger fra 1700-tallet gikk på familieæren løs.

## ***– Noen få empiriske regler***

---

Bouchard gikk så raskt gjennom en del viktige empiriske regler, dels som en rekapitulasjon av hva han hadde sagt tidligere:

- Unngå koding i størst mulig grad
- Standardiser ikke før lenking (dvs kun fonetisk like variasjoner)
- Ta vare på de regler som brukes ved (den automatiske) lenkingen, slik at lenkingsarbeidet kan dokumenteres
- Bevar muligheten for å endre lenker etterpå
- Ta alltid de letteste lenkene først

- Bygg en pålitelig base først ved bare å ta de sikreste lenkene. Bygg så videre på denne etterpå
- Begynn med det vanligste, og ta det uvanlige og unntakene etterpå
- Det er ønskelig at maskinen (programmet) alltid tar avgjørelsene
- Vurder datakvalitet og formålet med forskningen

Bouchard så store fordeler i formelle, stor-skala prosjekter. På den andre siden var han åpen for prosjekter i liten skala og ville være forsiktig med å ta motet fra små og individuelle prosjekter.

## **NORGE ETTER SOREP**

---

I et nytt foredrag samme dag presenterte Bouchard SOREP og gjennomgikk oppbygging og virkemåte for SOREP-systemet, men som sagt blir ikke dette direkte omtalt i dette referatet.

Flere deltakere lot seg stimulere av Bouchards redegjørelser. Etter en dyp tenkepause presenterte Nygaard en skisse av en norsk variant av SOREP-systemet, med de nødvendige hensyn tatt til særegenheter ved norsk kildemateriale. Han dømte systemet ad hoc for NOREP. Bouchard mente at Nygaards utkast hadde mye for seg.

Oldervoll trodde et NOREP ville være mulig, men ikke brukbart i alle situasjoner. Til et stort anlagt system trenger en programmerere og historikere som til enhver tid kan vedlikeholde og fininnstille systemet. Han så for seg 5–6 personer arbeide med systemet i 10 år. Et slikt system ville ikke kunne gis til forskeren i "bushen".

Thorvaldsen tenkte ut og beskrev hvordan han ville modifisere sitt eget lenkingssystem i lys av erfaringen på seminaret.

Bouchard sa avslutningsvis at det tar tid å bygge et godt automatisk system, men at det kan betale seg. Forutsetningen er at en satser stort. Han mente at siden det ikke fantes noe tilgjengelig lenkingssystem i Norge, kunne en nå gripe sjansen, bygge opp et system og skape lokal presedens. Han mente det kanskje var lettere å skaffe penger hvis en trakk inn epidemiologisk forskning. SOREP ønsker å utvide på denne måten. Også i USA er det lettere å skaffe penger for epidemiologisk forskning.

## **LITTERATUR**

---

For dem som har interesse av å sette seg dypere inn i lenkingsproblematikken, kan det være av interesse å vite hva seminardeltakerne fikk tilsendt som bakgrunnsmateriale:

Kitts, Arno et al: *The Reconstitution of Viana do Castelo*. AHC 1990. (Chapter 4.)

- Winchester, Ian: *Record Linkage in the Microcomputer Era: A Survey*. Demografiska databasen, 1985.
- Bouchard, Gérard: "The Processing of Ambiguous Links in Computerized Family Reconstitution". (Historical Methods, 1986-1.)
- Johansen, Hans Chr: "Record-linkage units: Persons, families, dwellings?". Seminarinnlegg, 1984.
- The Cambridge Group: "Automatic record linking for family reconstitution". (Local Population Studies No 40, 1988.)
- Buchanan, Anne V et al: "Reconstruction of Genealogies from Vital Records: The Laredo Epidemiology Project". (Computers and Biomedical Research 17, 1984.)

(I tillegg delte Lars Nygaard ut materiale om norsk 1800-talls kilde-materiale og det interaktive lenkingssystemet.)

### **DELTAKERNE**

---

Deltakerne på seminaret var:

- Gérard Bouchard* fra SOREP, Centre Interuniversitaire de Recherches sur les Populations, Université du Québec à Chicoutimi, Canada;
- Hans Christian Johansen* fra Historisk Institut, Odense Universitet, Danmark;
- Ivar Fonnes, Hege Brit Randsborg og Gunnar Thorvaldsen* fra Riksarkivet;
- Elisabeth Haavet* fra Senter for humanistisk kvinneforskning, UiB;
- Eli Fure og Sølvi Sogner* fra Avdeling for historie, UiO;
- Kåre Andersen* fra Avdeling for humanistisk informatikk, UiO;
- Lars Nygaard* fra Institutt for informatikk, UiO;
- Jan Oldervoll og Michael Drake* fra Institutt for samfunnsvitenskap, UiTø;
- Margunn Skjei Knudtsen* fra Historisk institutt UNIT-AVH;
- Ola Tovmo* fra Dovre dataarkiv;
- William H Hubbard*, Department of History, Concordia University, Montreal, Canada (for tiden også bosatt i Haugesund);
- Øystein Reigem* fra NAVFs edb-senter for humanistisk forskning.

# ONLINE 90

LONDON, 11.-13. DESEMBER 1990

---

---

*Ivar J. Gubberud og Unni Sæther*

*The 14th International Online Information Meeting* ble arrangert av *Learned Information* på utstillings- og konferansesenteret Olympia i Kensington. Det er et årlig arrangement med deltagere fra hele verden. Det var som vanlig to deler, konferanse og utstilling. Utstillingen i år var større enn tidligere, med hele 162 utstillere. Konferansen hadde nesten 2400 deltager, mens antallet besøkende til utstillingen var nærmere 8000.

Online i London er et kjempearrangement med parallelle sesjoner med foredrag og parallelle serier med produktpresentasjoner. Det var egne sesjoner for henholdsvis fransk-, tysk-, italiensk-, nederlandsk- og spansktalende delegater.

Det var som sagt mange utstillere og mye interessant å se. Det var en sterk vekst innenfor CD-ROM og mange databaser som er online, kommer også på plate. Produksjonsprisene for plater er gått ned og har gjort det mulig å lage produkter som krever bruk av mange plater – for eksempel tidsskrifter hvor sidene er lagret grafisk som bilder. Det var systemer med over 100 plater.

Det har også skjedd en del forbedringer når det gjelder online søking. Flere av de store databasevertene som ESA med Hyperline, DATA-STAR og ECHO har fått meny-drevne søkesystemer.

Konferansen ble åpnet av D. Raitt som bl.a. mente at online ikke lenger er hva det var. Online databaser er tilgjengelig på optisk medium og online-vertene ser ut til å gjøre stor bruk av CD-ROM. Det er nødvendig for informasjonsindustrien å ta i bruk ny teknologi, men han konkluderte med at det fremdeles er nødvendig med trykt litteratur og online databaser.

Emnene på konferansens sesjoner var følgende:

- multimedia-applikasjoner i dag og i fremtiden
- implementasjon av CD-ROM-systemer
- dokument-indeksering



- kjemisk søking
- utdanning og undervisning
- behovet for databaser
- informasjonsaktiva og korporativ intelligens
- online søking og software
- informasjonmarked – markedsføring
- nettverk og gateways
- databaser og informasjonspolitikk
- copyright i elektronikkens tidsalder
- integrering av informasjon – utvikling og forandringer
- globale utfordringer for bedriftsinformasjon

*John Gale, Information Workstation Group, USA*, åpnet sitt foredrag med følgende uttalelse: "Multimedia is a slippery term". Han presenterte status for multimedia i USA og hvordan markedet sannsynligvis ville utvikle seg i perioden 1990–94. Forbrukermarkedet fikk det største segmentet av kakediagrammet. Fra 1993 mente han at Japan antageligvis ville inneha markeds kontrollen.

1990-årene er spådd å bli online-verters og produsenters kvalitetsperiode sa *Reva Basch, Aubergine Information Services, USA*. Hovedfaktoren er en voksende bevegelse blant brukergrupper som ikke lenger betrakter online som en nyhet, men en tjeneste som ikke bør være immun for kritikk. I USA har denne gruppen tatt i bruk uformelle "trusler" som å skrive artikler og arrangere elektroniske konferanser for å øve innflytelse på online-verter og -produsenter. Gruppen ønsker en dialog med informasjons-industrien for utveksling av synspunkter, og de har klart å få onlineindustriens oppmerksomhet. Profesjonelle søkere har ervervet seg mye kunnskap som de deler med hverandre og med databaseprodusenter etc. Denne gruppen av søkere er ikke bare kunder mente hun, men også en markedsfører for onlineindustrien.

*Michael Temple, Ketron Inc., USA*, presenterte en markedsundersøkelsesmetode for å evaluere og skaffe seg oversikt over online-brukeres ønsker og behov. Onlineindustrien vokser både nasjonalt og internasjonalt og dette fører til skjerpet konkurranse mellom eksisterende og nyetablerte online-sentre. Nåværende markedsandel skal økes samt at nye tjenester også skal oppnå maksimum markedsandel. For å oppnå ønsket markedsandel må onlinetjenestene være i stand til å fange opp kundenes behov og ønsker, men samtidig må de holde øye med konkurrentenes tilbud. Den mest verdifulle bruken av undersøkelsesmetoden er i kombinasjon med en PC-basert markedssimulator. Ved bruk av dette verktøyet kan en skaffe seg oversikt/prognoser over estimerte markeder (gamle og nye) og over markedsandel av eksisterende og nye produkter.

*Graham Harris, SoftKlone (UK) Ltd. Storbritannia*, rettet oppmerksomheten mot brukergrensesnittet. Han mente at onlinedatabaser ofte er lite brukervennlige i utformingen. Det er 3 elementer som bidrar til dette: 1) Den ønskede informasjonen som skal være i basen – klassifiseringen og oppdelingen av informasjonen i dokumenter som igjen skal deles opp i felter. 2) Søkedelen av basen. 3) Terminalen. Disse 3 komponentene må utformes på en best mulig måte med brukerne for øye. Han trakk videre frem forskjellige typer terminaler og mente at få, men gjennomtenkte endringer kan gjøre det mye lettere for brukerne. Han konkluderte med at onlineverktøyet ikke har vært godt nok, men med noen endringer kan en oppnå større brukervennlighet.

Prisen for beste foredrag ble vunnet av *Péter Jacsó fra University of Hawaii, USA*. Han presenterte på en oversiktlig og grei måte den ideelle CD-ROM-arbeidsstasjon for 1990-årene. De enkelte begrensede faktorene i en PC ble gjennomgått og han kom med konkrete forslag til hvilke valg man bør ta med tanke på CD-ROM. I sannhet et oversiktlig og matnyttig foredrag som ikke hadde for mange tomme fraser.

Et interessant emne var problemstillingene rundt paneuropeisk multi-lingualisme i *CELEX* databasen. Dette er et datamaskinbasert dokumentasjonssystem for det felles lovverket i EF. Alle lover skal være tilgjengelig online på EFs ni offisielle arbeidsspråk. Dette skaper store problemer, som blir løst via implementasjon av internasjonale standarder for koding av språklige spesialtegn og for kommunikasjon. Nærmere opplysninger om dette kan interesserte få fra *Panagiotis E. Alevantis, CEC, IMCO, Brüssel*.

*SGML (Standard Generalised Markup Language)* er et systemuavhengig og applikasjonsuavhengig metaspråk som brukes for å merke skjematisk oppsatt tekst. Foredraget av *Brian Robinson, Hatfield Polytechnic, England*, tok særlig for seg applikasjoner i SGML som er nå under utvikling. Det er utviklet programmer som tillater brukerne å fylle ut komplekse skjema på egen PC. Skjemaet kan deretter sendes via datanettverket for automatisk behandling i stormaskinmiljø langt unna dataleverandøren. Spørrener har full kontroll over alle aspekter av datainnsamlingen; sikkerhet, synlige felter og brukerassistanse er tilgjengelig online.

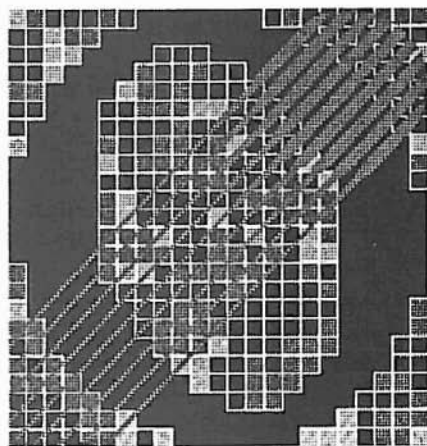
Database over databaser – én inngang til alle databaser, er en gammel drøm som en har arbeidet med i USA helt siden tidlig i femtiårene og framover til sekstiårene, da kritikere så mulighetene til å knytte alle databaser sammen som en realisering av Orwellianske skrekkevísjoner om en nasjonal databank. Senere har brukere av de forskjellige åpne databaser, bibliotekarer, historikere, forskere og journalister forstått at alle de forskjellige databasesystemene; organisasjon, søkespråk og generell infrastruktur reduserer nytteverdien og oppfyller ikke ønskene om

god tilgjengelighet. *Forest Woody Horton jr., Information Management Press, USA*, tok for seg denne problemstillingen på et teoretisk plan og forsøkte å sette opp forslag til konsept for løsninger på problemet database over databaser.

Ingen konferanse om databaser er komplett uten en sesjon om copyright-spørsmål. I London var det fire foredrag om opphavsrettigheter, som fremstilte problemet fra meget forskjellige vinkler. En gikk inn for et utvidet system av betaling og innskrenkninger i datastrømmen. En annen mente at lovverket måtte forandres for å tilpasses informasjonsflyten i samfunnet, mens en tredje mente at betaling for bruk ville presse seg fram som det mest kostnadseffektive i relasjon til systemer uten betaling som snart ville dø hen. Det ble ikke gitt entydige svar – og ingen hadde vel ventet det heller. Debatten var imidlertid opplysende og fikk satt problemet i relieff.

Foredragene er publisert i: *ONLINE INFORMATION 90, 14th International Information Meeting, Proceedings. Learned Information (Europe) Ltd., Oxford, England, 1990. ISBN 0 904933 75 X.*

# ONLINE<sup>90</sup> INFORMATION



14th International Online Information Meeting

Proceedings

# INTERNASJONALT SEMINAR OM FORSKNINGSDOKUMENTASJON

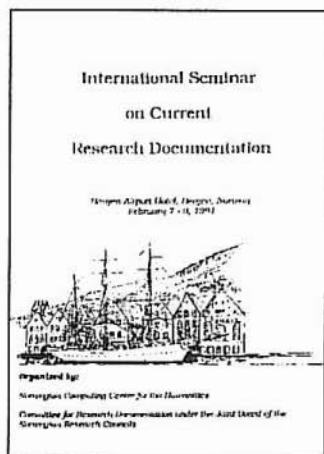
---

---

Ivar J. Gubberud

## INTERNASJONALT SAMARBEID

---



Et område som stadig blir viktigere, er dokumentasjon av pågående forskning. Dette kom tydelig fram på det internasjonale seminaret som ble avholdt 7-9. februar 1991 på Bergen Airport Hotel. Blant foredragsholderne var det representanter fra elleve europeiske land.

Det var *Utvalg for prosjektinformasjon* under *Forskningsrådenes samarbeidsutvalg* og *Sentral for informasjon om forskningsprosjekter* ved NAVFs edb-senter for humanistisk forskning, som stod som arrangører av det internasjonale seminaret om dokumentasjon av pågående forskning.

Seminaret var lagt opp som en bred mønstring av det arbeid som foregår i en

rekke nordiske og europeiske land for å skape og videreutvikle informasjonskilder om den forskning som utføres.

## STORT BEHOV

---

Behovet for oversikt over pågående forskning synes å stige i takt med økende omfang av forskning, økende behov for å legitimere virksomheten og behov for å utføre strategisk planlegging av forskningsarbeidet. I forbindelse med formidling av forskningsaktiviteter til industri og næringsliv og allmenheten kan dokumentasjonssystemer for igangværende forskning være sentrale virkemidler. I vårt land har forskningsrådene og flere av universitetene i løpet av det siste tiåret utviklet en systematisk registrering og dokumentasjon av forskningsarbeid.

Seminaret besto av presentasjoner fra inviterte foredragsholdere fra flere europeiske og nordiske land. I tillegg ble det gitt presentasjoner av norsk virksomhet på feltet. På programmet stod følgende foredrag:

- "The harmonization of databases on current research within the European Community", *I.R. Perry*, Directorate General for Science, Research and Development, Commission of the European Communities, Belgium
- "EXIRPTS, Exchange of Information on Research Projects", *Helena Fernholm*, Styrelsen för Teknisk Utveckling, Sweden, *F. Naldi – Gloria Bordogna*, CNR – SIAM, Italy
- "Establishment of a national research data base – the Danish approach", *Birgit Pedersen*, DANDOK, Denmark
- "Research documentation in government institutions in Finland", *Maria von Hertzen*, NORDINFO, Finland
- "The Norwegian Situation. The Needs for National Collaboration", *Ivar Gubberud, Jostein H. Hauge, Unni Sæther*, Norwegian Computing Centre for the Humanities, Norway
- "Collecting data for an ongoing research information system, the case of the Technical Research Centre of Finland", *Pirjo Sutela*, Statens Tekniska Forskningscentral, VTT, Finland
- "A policy for harmonized information system for R & D in France: Research units, ongoing research projects; The French Ministry of Research's demands and needs for science and technology", *M. Choukroun*, Ministère de la Recherche et de l'Enseignement Supérieur, France
- "Research Documentation in UK: Problems and Opportunities", *K.G. Jeffery*, Rutherford Appleton Laboratory, United Kingdom
- "Collection and retrieval of scientific information in Flanders", *Martin Ruebens*, Department of Research Policy and Technology, Ministry for the Flemish Community, Belgium
- "The Czechoslovak Information Structure and Online Services", *Vera Jiráková*, CSAV, Základní knihovna, Ustředí vedeckých informací, Czechoslovakia
- "The PROTEAS database", *Madeleine Campbell*, Longman Cartermill Ltd., United Kingdom
- "Experience with existing research databases in Germany and outlook to a future possible European network of research databases", *H. Behrens*, Fachinformationszentrum Karlsruhe, Germany
- "Availability of information and end-user search performance and the quality of information; strategies for the nineties", *Auke van Dijke, Johan van de Walle*, The Netherlands Bureau of Research Information (NBOI), the Netherlands

Det var denne gangen deltagere både fra departementsnivå, nasjonalt institusjonsnivå og fra faginstusjonsnivå. De forskjellige nasjonale løsningene varierte fra det absolutte bunn-nivå og opp til et meget avansert og sofistikert framtidsrettet nivå. Det norske nivået kunne absolutt hevde seg i europeisk sammenheng, selv om det er mange problemer å løse før vi kan si oss fornøyd.

## **FAGLIGE SPØRSMÅL OG PRAKTISK BRUK**

---

Det ble også avsatt tid for demonstrasjoner, presentasjoner av produkter, utveksling av informasjon og sosial kontakt. Seminaret ga derfor en svært god mulighet til å bli kjent med oppbygging og bruk av forskningsinformasjon i en rekke ulike land.

I tillegg til presentasjonene av ulike dokumentasjonstjenester, ble oppmerksomheten rettet mot følgende temaer:

- \* Målsettingsspørsmål. Hvorfor arbeider vi med forskningsdokumentasjon?
- \* Hvilken tjenester kan et dokumentasjonssystem for pågående forskning gi?
- \* Hvilke brukergrupper finnes for prosjektinformasjon?
- \* Dokumentasjonsstandard. Fellesformater.
- \* Internasjonal utveksling av prosjektinformasjon
- \* Prissettingsspørsmål
- \* Opphavsretts spørsmål
- \* Brukerstøtte ved utnyttelse av dokumentasjonssystemer for igangværende forskning

## **DOKUMENTASJON**

---

CERIF-manualen som er utgitt av EF-kommisjonen i Brussel vakte stor interesse. Dette er et forslag til et felles internasjonalt format for dokumentasjon av forskningsinformasjon i europeiske forskningsdatabaser, og et forslag til klassifikasjon av forskningsfelt. Denne manualen kan være nyttig i det videre arbeidet for å internasjonalisere norsk forskning og for å bringe internasjonal forskning raskere fram til store brukergrupper.

Responen på denne konferansen var meget god. Det meldte seg nesten 50% flere deltagere enn først estimert. De aller fleste deltagerne var meget fornøyd med konferansen og arrangementet som var av den samme høye standard som NAVFs edb-senter er blitt kjent for i de senere år. Dette ga seg utslag i et ønske om en oppfølgingskonferanse enten på nordisk basis, eller en europeisk konferanse sentralt i Europa.

I første rekke vil det i løpet av kort tid foreligge en grundig rapport fra konferansen. De som ønsker å sikre seg denne rapporten tidlig, kan allerede nå bestille den fra: *SIF, Postboks 53, Universitetet, N-5027 Bergen*. Utsendelse vil skje i løpet av annet kvartal 1991 til deltagere og dem som bestiller nå.



# **SENTRAL FOR INFORMASJON OM FORSKNINGSPROSJEKTER**

---

---

*Ivar J. Gubberud*

## **GRUNNLAGET FOR FELLESDATABASEN**

---

FSUs Utvalg for prosjektinformasjon vedtok 27. april 1990 etter mange års forarbeid å sette i gang et prøveprosjekt for overføring av prosjektinformasjon fra alle forskningsrådene i Norge. Datainnleggingen var fullført høsten -90.

Ansvar for avlevering av data ble lagt på de enkelte forskningsrådene, mens ansvaret for tilrettelegging ble lagt til den allerede eksisterende online-databasen i Bergen, som allerede inneholder alle NAVF prosjekter.

## **1. DATABASENE**

---

Databasene for forskningsdokumentasjon ligger på en IBM 4381-maskin som tilhører Universitetet i Bergen og er plassert på Høyteknologisenteret i Bergen. Basen bruker søkesystemet SIFT, som er et system for søking i fri tekst, utviklet av Statens Datasentral.

Databasene er tilgjengelig døgnet rundt via DATEL, DATAPAK og UNINETT. Det finnes en bemannet feilmeldingstjeneste mellom kl 8.00 og kl 22.00. Brukere kan få manuell veiledning og søkeassistanse mellom 8.00 og 16.00, og i tillegg er det utarbeidet et detaljert veiledningshefte. Tilgjengeligheten er således blant de beste som finnes. Alle tjenestene er uten belastning for bruker.

## **2. PROSJEKTDATA**

---

Data blir mottatt fra de fem norske forskningsrådene, enten direkte eller via enkelte dedikerte mellomledd for kvalitetssikring. NORAS og NAVF/RSF bruker NSD i Bergen til dette, mens NAVF/RHF bruker HUM-DOK ved NAVFs edb-senter for humanistisk forskning i Bergen. Der legges det inn opplysninger om forskningsprosjekter med eventuelle tilhørende

publikasjoner innenfor alle områder av norsk forskning. De øvrige rådene overfører en elektronisk lesbar utgave av sine prosjektkataloger. Det blir således en forskjell i kvalitet på data avhengig av hvilken grad av raffinering som har skjedd. Dog synes det som om alle rådernes data inneholder en kjerne av felles opplysninger som de selv anser som tilstrekkelige og som er overlappende.

Totalt antall prosjektdokumenter i fellesdatabasene pr. 31/12 1990 var 25.278. FSU prosjektdata akkumulert til og med 1990 fordeler seg slik:

|                       |       |                    |
|-----------------------|-------|--------------------|
| NFFR:                 | 315   | fra 1990           |
| NLVF:                 | 331   | fra 1990           |
| NTNF:                 | 593   | fra 1990           |
| NAVF/RSF-NORAS:       | 9.611 | fra 1984           |
| NAVF/RHF:             | 4.839 | fra 1984           |
| NAVF/RMF:             | 652   | fra 1990           |
| NAVF/RNF:             | 2.188 | fra 1986           |
| Spesialkartlegginger: | 6.749 | punktundersøkelser |

### 3. HVA FINNES I FELLESDATABASEN?

---

Fellesdatabasen inneholder informasjon om prosjekt som spenner fra jomfrutrualer i Stillehavet, via korsang til astrofysikk og miljøvern. Dette åpner muligheten for interessante tverrfaglige søk. Særlig for informasjonsmedarbeidere, journalister og samfunnsplanleggere vil dette skape muligheter til å trekke fram ny informasjon basert på data fra fellesdatabasen.

Det finnes nasjonale problem- og satsningsområder som det vil være av stor nytte å kunne sammenstille og evaluere. I denne forbindelse er det lett å trekke fram et knippe områder som *miljø, alger og stråling* som er representert innen mange forskningsfelt. I alt ble det funnet *1380 dokumenter* om miljøforskning i Norge ved et søk i NAVFs database som bare tok noen få minutter. Ville det i det hele tatt ha vært praktisk mulig å finne disse opplysningene på annen måte? En måtte i så tilfelle ha brukt 20–25 kataloger og lest hvert eneste ord i prosjektbeskrivelsen for å være sikker på å ha fått med alle relasjoner. Oppsiktsvekkende nok viser det seg at det også innen faget kunsthistorie pågår forskning som omfatter *alger*.

Tverrfaglig informasjon av denne kategori er interessant i seg selv, men også sett i et overordnet og prinsipielt perspektiv.

#### **4. ERFARINGER MED FELLESDATABASEN**

---

Erfaringene fra samarbeidet med en fellesdatabase for informasjon om pågående forskning kan summeres opp slik:

- Alle data foreligger allerede tilgjengelig i elektronisk form.
- Ordningen med en fellesdatabase er **meget rimelig** i forhold til den **økte informasjonsverdi** dataene får ved sammenstilling og at de gjøres elektronisk søkbare.
- Fellesdatabasen representerer også en form for **gjenbruk av data** fra søknadsbehandlingssystemer og katalogproduksjon.
- Blant norske brukere er **interessen stor** for å kunne søke tverrfaglig og over flere bevilgningsår.
- Fellesdatabasen kan inngå i et **internasjonalt nettverk** av forskningsinformasjon, som vil gjøre norsk forskning mer kjent i utlandet.
- Tverrfaglige søk, som spenner over flere bevilgningsår og ikke avgrenses til registrerte søkeord, er neppe mulig å gjøre i praksis uten **søkesystemer for fritekst**.
- Hvis vi skal lykkes i å spre norske forskningsdata i utlandet, må all dokumentasjon av prosjekter inneholde **nøkkelinformasjon** også på engelsk.
- Tilgjengeligheten til databasen er god, men forskjellige søkespråk i norske og utenlandske databaser er et problem som det for tiden **ofres stor oppmerksomhet** på og som det er naturlig å anta at kan bli løst innen rimelig tid.

I tråd med ovennevnte gode erfaringer og perspektiver vil vi videreføre samarbeidet i tiden framover. Interesserte brukere vil kunne kontakte SIF, Postboks 53, Universitetet, 5027 Bergen, eller på telefon + 47 (0)5 21 29 57 for nærmere opplysninger.

# **ELEKTRONISKE BØKER – MULTIMEDIA OPPSLAGSVERK**

## **KONFERANSERAPPORT**

---

---

**Kjell Morland**

Denne konferansen, som ble arrangert i Bergen 21.–23. nov. 1990, var en storstilt presentasjon av elektroniske bøker og multimedia-prosjekt med nærmere 20 foredrag og et stort antall demonstrasjoner.

Konferansen viste på den ene siden en fantastisk verden av elektronisk tilgjengelig informasjon og underholdning, og demonstrerte de mulighetene som ligger i å kunne presentere tekst, lyd og levende bilder slik at brukeren kan bevege seg fritt og raskt mellom de ulike elementene. På den andre siden ble de tekniske nyvinningene satt i perspektiv og sammenliknet med de konvensjonelle trykte mediene.

Konklusjonene må bli at ikke alt det gode er nytt, og at ikke alt det nye er helt godt, ennå. De elektroniske mediene kan opplagt bli et godt supplement til dagens teknologier og den konvensjonelle boken. De byr på store praktiske fordeler ved lagring og oppdatering av store informasjonsmengder, men det er et stykke fram til de elektroniske bøkene blir allemannseie.

Store oppslagsverk begynner nå å bli utgitt elektronisk. På konferansen ble det vist flere elektroniske leksika og ordbøker, som f.eks *Compton's Multimedia Encyclopedia* fra Britannica, den svenske *Nationalencyclopedia*, begge på CD-ROM. En CD-ROM (dataversjonen av CD-platen) er en ferdig innspilt plate med lagerkapasitet som tilsvarer et 20-binds leksikon pluss lyd og bilder. Det ble også vist en rekke ordbøker utgitt på diskett. En nyhet som ble presentert under konferansen, var databaser fra det norske *Lovdata* på CD-ROM.

*Bela Hatvany*, SilverPlatter Information Group, talte varmt for den elektroniske boken, fordi den er så lett å søke i. Han mente det var umulig å gå tilbake til å slå opp i papirutgaven av *Encyclopaedia Britannica* etter at man var blitt vant med den elektroniske utgaven. Boken var død, mente han.

*Graham Brown-Martin* fra Next Technology, Cambridge, beskrev

utviklingen av den digitale CD'en med sin store lagringskapasitet som grunnlaget for en ny generasjon av personlige datamaskiner. Foreløpig er det et problem at levende film tar så stor lagerplass når den digitaliseres, men snart vil det være mulig å komprimere de store tallmengdene, slik at en CD-ROM vil kunne inneholde 72 minutter video – nesten en hel spillefilm. Og når det nye lagringsmediet kombineres med en PC, kan brukeren hente frem tekst, lyd og bilde med bare noen tastetrykk. Hittil har dette utstyret vært for kostbart til at det har vært noe stort marked for CD-ROM-produksjoner, men snart vil en integrert multimedia-maskin koste så lite som 1.000 dollar i USA, og da kan dette utstyret bli like vanlig som videospilleren idag.

Graham Brown-Martin mente at «brukergrensesnittet» også må forbedres: Brukerne må få enkle anvisninger på hvordan man går inn i og ut av programmene. Personlig testet han gjerne programmene ved å se om hans egne barn og hans foreldre kunne bruke dem. Hvis de fant frem, kunne programmene gå ut til publikum. Han viste et eksempel fra et dyreprogram. Spørsmålet til brukeren var «What would you like to know?», med mulighet for å velge mellom ulike dyrebilder. Han syntes absolutt det var et bedre spørsmål enn et han hadde sett i et annet program: «Would you like to do a boolean search?»

Øyvin Rannem, Grafisk Institutt, mente at det var lett å se seg blind på de nye tekniske mulighetene. Her var det viktig også å være visjonær, og gjøre noe for å styre utviklingen. Det nye med teknologien er at den gir oss adgang til mye informasjon på ett sted: musikk, film, bilder, tekst. Denne informasjonen kan vi ha tilgjengelig på andre måter – i bokhyllen, som video eller på plate. Forskjellen ville i alle fall være noen skritt og noen meter! Og den nye spontaniteten må bety en forskjell, mente Rannem. Den vil fremheve sammenhenger på tvers av fagskillene. Derfor vil multimedie kreve ny pedagogisk tenking. Kanskje kan noen av de gamle fagskillene oppheves. Han antydte at boktrykkerkunsten kanskje var ansvarlig for fagskillene ved skoler og universitet, og dermed at det var Gutenberg og hans oppfinnelse som hadde delt opp verden.

Deretter tok Rannem for seg begrepsparet sekvensialitet/ikke sekvensialitet: Vi sier at de trykte mediene er sekvensielle, mens elektroniske medier er ikke-sekvensielle, og bedre enn de sekvensielle fordi vi oppfatter den ikke-sekvensielle presentasjonen som verdinøytral. Men nå når vi først har satt søkelys på begrepet, ser vi at det er mye ikke-sekvensialitet i trykte medier, som f.eks i aviser og leksika. Her kan leseren bevege seg frem og tilbake og danne egne forbindelser, og dermed skape sin egen kontekst for det leste – i prinsippet på samme måte som i hypertekst, der brukeren kan bevege seg mellom ulike tekstelementer eller bilder.

En annen foredragsholder påpekte tilsvarende at den profesjonelle leser vil lese en vitenskapelig artikkel ikke-sekvensielt. Han leser kanskje sammendraget først, og deretter leser han litt her og litt der.

*Petter Henriksen*, Den norske forleggerforenings utvalg for elektronisk forlagsvirksomhet, etterlyste en god informasjonsarkitektur for det nye mediet. Informasjonsstrømmen er nå så formidabel at den har blitt karakterisert som et «Informasjons-Ragnarokk», og det har blitt viktigere enn noensinne å strukturere informasjonen.

Informasjon har alltid hatt en eller annen struktur: Papyrus-rullen er strengt sekvensielt strukturert, mens Bibelen er ordnet hierarkisk, med 2 hoveddeler med flere bøker inndelt i kapitler og vers. For det nye mediet blir informasjonsarkitekturen spesielt viktig, mente han. I en dårlig organisert bok kan du alltid finne frem ved å bla lett frem og tilbake, men «550 MB med slapp informasjons-arkitektur er en katastrofe». Her så han en fremtid og en utfordring for forleggere og humanister. Virkeligheten er multimedial, siden vi mottar informasjon gjennom ulike kanaler. Utfordringen blir å kople disse mulighetene sammen på en teknisk god måte slik at brukeren ikke mister oversikten.

Begrepet «interaktivitet» ble også problematisert. Interaktivitet innebærer at brukeren har mulighet for å velge hvor han eller hun vil bevege seg i databasen. Dette er et honnørbegrep innenfor multimediene, men forfatteren *Tor Edvin Dahl* gikk begrepet nærmere etter i sømmene. For det første ga han eksempler på at den samme type valgfrihet finnes i en del bøker, der leseren inviteres til å delta og bygge opp fortellingen. For det andre viste han – med basis hovedsakelig i dataspill – hvordan interaktiviteten stort sett foregår på programskaperens premisser. Leseren får rett nok flere valg, men i slike spill er *interaktivitet* lik *problem-løsning*: «målet» for spillet er å finne skatten, drepe draken e.l, ved hjelp av en opplysning eller nøkkel som spilleren vet ligger et eller annet sted underveis. Det gjelder bare ikke å overse den.

Den interaktiviteten Dahl kunne ønske seg, var en type samhandling mellom bruker og datamaskin basert på brukerens «kreativitet og innlevelse». Og på den bakgrunn er faktisk boken bedre enn edb-programmet. For romanen gir oss mulighet for å leve med, leve oss inn i følelser og bilder og svelge alt rått. Her er ingen valg, men her er opplevelse og innlevelse. Det som kjennetegner de valgene som datamaskinen byr brukeren, er en type ironisk distanse. Distansen skapes ved at brukeren settes på prøve, og programskaperen har kontrollen. Dersom du som spiller/bruker virkelig er kreativ, «så skjærer det seg fullstendig». Kreativitet er bare akseptabelt på programskaperens premisser.

Dahls konklusjoner ble indirekte imøteggått av to av foredragsholderne som begge presenterte undervisningsprosjekt som ga begrepet interaktivitet mer innhold. En av disse var Professor *George Landow*, som

presenterte *The Dickens Web*, et undervisningsopplegg utviklet av Institution for Research in Information and Scholarship (IRIS) ved Brown University. Dette er et hypertext-opplegg, der ulike typer tekst og bilde kan trekkes frem på skjermen, og der det er lagt inn elektroniske lenker som peker frem og tilbake mellom tekstene. Brukeren kan følge de forbindelsene han eller hun ønsker. I databasen ligger Dickens' verker, tekster av andre forfattere som har påvirket eller blitt påvirket av Dickens, ulike typer opplysninger om Victoriatidens England og litteraturkritiske artikler. I tillegg kan studenten opprette sine egne dokument i programmet og kopiere inn tekstelementer og skrive en artikkel, som kan bli stående til neste bruker.

Landow sa at et av målene for opplegget var at studenten skulle se forbindelser mellom ulike tekster og sette disse sammen slik at han eller hun ved lesingen stilte sine egne spørsmål og dannet sine egne hypoteser. I motsetning til i de dataspillene som Tor Edvin Dahl analyserte, var det her snakk om selvstendige tekster som studenten kan lese og tolke på nye måter. Læreren/programskaperen har bare begrenset kontroll over hva studenten kan trekke ut av programmet.

Et annet undervisningsprogram viste også hvilken kvalitet multimedia kan bringe til undervisning og forskning. *Elli Mylonas'* presentasjon av Perseus-prosjektet, som pågår ved Harvard University, dannet en verdig avslutning på konferansen. Hensikten med dette prosjektet var å sette sammen en hypertextuell database om det klassiske Hellas, både litteratur og arkeologiske data, for å bringe sammen ulike fagfelt og gjøre materialet lett tilgjengelig. Tekstbasen, som er lagret på en CD-ROM, inneholder store deler av den klassiske greske litteraturen, et gresk-engelsk leksikon, en historisk oversikt og en klassisk encyclopaedia. Dessuten var en del levende film og 6.000 fargefotos av funnsteder, vaser, mynter og skulpturer, i tillegg til kart og tegninger i svart/hvitt lagt inn på videoplate.

For et prosjekt av denne størrelsen er det elektroniske mediet praktisk. Det vil gjøre informasjonen lett å lagre, enkel å oppdatere og den kan utgis i ulike versjoner. Studentpris på CD-ROM'en vil være ca US\$150, og vil bli utgitt av Yale UP.

Deretter gav Elli Mylonas en demonstrasjon av produktet. Hun gikk inn i noen av de greske primærtekstene: Aiskylos skriver et sted om Delphi. Da er det mulig ved noen tastetrykk å slå opp i leksikonet på «Delphi» og lese om templet, se hva leksikonet skriver om en kunst-historisk term («stylobate temple») eller å se videoopptak av tempelruinene.

Det er mulig å finne ut hvor mange ganger en forfatter bruker bestemte ord, og om andre forfattere bruker dem, eller noen av dem. Det er mulig å gå direkte til ordbok eller leksikon og få utfyllende



opplysninger. Alt i alt ble dette en meget overbevisende demonstrasjon av mulighetene med multimedia – en demonstrasjon av en svært effektiv læringsmaskin.

Konferansen ble arrangert i samarbeid mellom NAVFs edb-senter for humanistisk forskning, Grafisk Institutt og Den norske forleggerforenings utvalg for elektronisk forlagsvirksomhet.

# ***A TASTE OF THE ELECTRONIC LIBRARY***

## ***"ELECTRONIC BOOKS – MULTIMEDIA REFERENCE WORKS" CONFERENCE IN BERGEN 21–23 NOV. 1990***

---

---

*Espen Aarseth*

One is easily impressed by conferences arranged by the *Norwegian Computing Centre for the Humanities*. We were impressed by its national conference *Computers in the Humanities* in 1988, where an efficient crew kept three or four parallel sessions going for an audience of 140. At this conference, with an audience of 175, in which the Norwegian Publishers' Association and the Norwegian Institute for the Graphic industries, Oslo, were co-arrangers, things went mostly without a hitch – and anybody who has arranged a conference and relied on heavy technical equipment (including prototypes) knows that it is impossible to get everything to function on time. It was therefore somewhat comforting to the rest of us to see that there were some small glitches, not many, but enough of them to make those responsible appear human, after all.

This time the program was far more linear than at the 1988 conference, since there was only one main program line, which made rather a contrast to the topic – or at least to one of the topics: non-linear books and their consequences for authors, publishers and readers – in short: for the information society.

### ***A NONLINEAR BOOK – WHAT IN THE WORLD IS THAT?***

---

It is a text in which the data (sentences, encyclopaedia entries, pictures etc) are not organized neatly in rows, one page following another, but are organized with the aid of pointers and links, to put it simply. It may be in the shape of a hierarchy (tree structure), or in the shape of more or less well organized networks of the spaghetti type. Normally (but not necessarily) one reads such texts – usually termed *hypertexts* – by using a computer. When sound, pictures and film sequences are

added, it is called hypermedia or interactive multimedia.

If we compare the reading of a conventional book to a train ride, where the route is established and the view exactly the same from one ride to the next, a non-linear book may be compared to a journey into unknown land, where the readers choose their own paths and explore the landscape. In addition to the demands which this naturally makes on the readers, who have to seek information *actively* in order to get any, authors also meet new challenges.

When an author organizes a quantity of material into a conventional, linear book, it is done through phrasing (use of words), exclusion (determining what should not be included) and disposition (plot, narrative structure, the order of the elements). These elements make up what may be called a *rhetoric*, in one sense of the word.

A non-linear book permits a different type of rhetoric. Exclusion becomes less important, since the author may include pointers to material which would be left out of a linear version, but which here can invite the reader to take small excursions, without being in the way, so to speak.

The phrasing will also be affected: Since a word or a phrase in an ordinary sentence may be a pointer to a different part of the text, these words will gain in importance and attract attention – they may even become more important than the sentence in which they are placed. Awareness of this fact will change both the author's way of writing, and the reader's way of reading.

In a non-linear text (or hypertext), textual elements are not placed in any specific order, but are organised in a network, in which many different orders are possible. Thereby the old rules of disposition are waived. Aristotle's thesis about a beginning, a middle and an end does not apply to non-linear stories. In its place we need rules and conventions for how to organise a non-linear text – and for how the reader should navigate it. These questions were the focus of the conference, and they are a source of much criticism of hypertext. In a fragmented network of pointers and textual bits, it is easy to get "lost in hyperspace", as the saying goes.

The constructors of hypertext face a dual challenge: they must develop ways of organising (partitioning) the material and find pedagogically adept conventions for pointing, which can form a standard for hypertext as a form of communication. They also need to create navigational systems which give the readers full overview and good means of "transportation". In this process, well-founded criticism is indispensable, while banal and negative criticism based on fear of the unknown, will derail and damage the process.

It takes a long time for a new medium to find its form, cf. the

evolution of "film language", and it is therefore important to look beyond the limitations which will become visible in the first weak attempts. "Literature is debugged" says *Ted Nelson*, the hyperguru of hypertext, meaning that the system consisting of authors, publishers, typographers, books, distribution and readers is honed to a fine point over the centuries, and therefore functions as well as it does. At the same time he sees hypertext and electronic publishing as a natural extension of this system – and the substantial interest in the conference as well as the composition of the participants indicate that his view is shared by many.

### ***THE MEDIUM WAS THE MESSAGE – BUT MORE THAN DISC JOCKEY PALAVER***

---

Let us return to the conference. A varied selection of Norwegian and foreign representatives for the various interest groups – publishers, media specialists and product developers – were invited to present papers, and in general the quality of the speakers was very good and the variation useful and mutually inspiring.

When media technology is up for discussion, there is always the danger of what I term disc jockey palaver – focussing on technical standards and acronyms (CD-ROM, CDTV, CDI, DVI, CDV, CVD, etc), mixed with babble of a brilliant future and with what comes close to echoing proud mothers cooing over each other's offspring, "Oooh, how *small* it has become!" Standards and novelties are important, but not for themselves.

Fortunately, there was not very much of that at this conference, but there was, on the other hand, a lot of show – and at times it was quite impressive. First showman to the rostrum was *Graham Brown-Martin*, who participated in BBC's project *Domesday*. He gave an entertaining and informative introduction to the development of multimedia over the past ten years, from what he called "interrupted video" to applications which have become more fully integrated and "content rich". He also showed some examples of three-dimensional animations with music, made by children of varying ages. They were a kind of rock videos, funny and surprisingly well made and experimental. His final rather optimistic message was that we as Europeans, with our long tradition of art and culture, might play an important role as designers in relation to technological giants like the USA and Japan: "Europe could lead!"

*Knut Brustad*, representing the Norwegian company of *SYNTEC*, made a rather sombre and earthbound contrast to Brown-Martin. Brustad has been the editor of *LaserAtlas of Møre og Romsdal County*, a multimedial

presentation of the county aimed at the tourist industry. His experience from this work was the topic of his presentation, and his story was both useful and depressing. Obviously, multimedia production in Norway is not the place to go looking for easy money and a comfortable job. There were many problems, ranging from marketing to text processing. (According to Brustad, text was in fact the worst kind of data to handle!) Among his advice to future record producers, I took particular note of the following: "Learn, and learn quickly", and "Information is more expensive than technology". His conclusion was that the Norwegian market is not yet ready for such productions, and he advised would-be producers of multimedia records to wait – "in an intelligent manner".

A different approach to the record market was presented by *Tomas Bjöersdorff*, marketing manager of *Esselte Special Publishers, Stockholm*, who talked about the development and marketing of CD-ROM. He pointed out the curious fact that this "distribution medium" as he called it, was developed without anybody defining a target group. If one had tried to do so, the CD-ROM might never have gotten off the drawing board. There simply was no market! Bjöersdorff also presented interesting figures for the distribution of CD-players in Norway compared to Sweden: In Norway the total number of players is 4–500, while the Swedish government has purchased 2,000 players for the libraries alone.

### **CRITICISM AND REFLECTION – FROM DIFFERENT ANGLES**

---

Those of us who desired more theoretical and principled considerations on hypermedia and non-linearity, were not disappointed either. *Øyvind Rannem, The Norwegian Institute for the Graphic Industries*, started that discussion in a presentation that covered a lot of ground and provided a critical perspective on hypertext and multimedia, in a humorous and pointed form. He speculated on what the "interactive" element in these media really consisted of, and emphasised three key terms: spontaneity, control and focus on the user. Texts are and have always been sequential, he said, and described hypertext as a kind of interface. One may accept or reject his rather negative view of hypertext as "antistructured information", but he made some very apt observations, and his clearly formulated critical remarks would have been a fruitful starting point for a further discussion.

*George Landow, Brown University*, taking his starting point in a demonstration of the Intermedia system, discussed hypertext as an educational medium. He predicted enormous possibilities for the medium – in the long run, making reference to the Gutenberg "revolution" which took about 350 years.

Landow warned against placing too much emphasis on non-textual education. Text is more flexible than other forms of information, and hypertext is text. He also touched the problems of publishing, distribution and copyright, and foresaw that in the future today's hypertext would look quaint.

Author *Tor Edvin Dahl* pointed out that interactivity in art is no novelty, and illuminated the problem with a number of examples of interactive fiction. He saw severe limitations in today's "adventure games", a genre with which he was thoroughly familiar, and felt there was a need for more structure and dramaturgy in order to make it exiting and rewarding. This writer, however, does not see a need for more dramaturgy. However, not even after Dahl's presentation was there time for general discussion, so that polemic will have to rest.

*Gunnar Liestøl*, *Department of Media and Communication, University of Oslo*, regarded the computer in the context of media history, and showed how the computer has annexed other media (mail, TV, text, pictures, etc) to become a multimediu. His humorous and pointed observation in that connection was that the computer became a medium when digital pornography made headlines.

He also touched the problem of linearity, and felt that it was necessary to distinguish between time and space: temporal non-linearity is "impossible", while spatial non-linearity actually should be termed "multi-sequentiality" – a term also used by Landow. Without going too far into these chronotopical problems, I will still permit myself to point out a small disparity in the terminology: Ted Nelson uses the term "non-sequential", where others say "multi-sequential". Something might probably be gained by a degree of standardisation, though I would not say that Landow and Liestøl should be the ones to change their terminology.

The wittiest analogy of the conference was made by *Erling Maartmann-Moe* from *the Norwegian Computing Centre, Oslo*. He compared the free jumping possibilities of today's hypertext to the infamous "go to"-paradigm in earlier programming. Today, too, we are completely free to create chaos, he said, and called for hypermedia's parallel to object-oriented programming.

### ***PERSEUS – A GEM!***

---

The most precious item in the conference (as well as the best argument I have seen for the use of multimedia in the humanities) was *Elli Mylonas's* presentation of the Perseus project, developed at *Harvard University*. The Perseus project combines classical and antique Greek literature, art, history and archeological material in a hypercard-based

multimedia installation, available on diskette, compact disc and video disc.

Perseus offers approximately 40 – later a total of 60 – classical texts by authors like Pindar, Aeschylus and Sophocles in the original and in English, interactive maps of archeological sites, stills, video sequences and dictionaries. The user may make his own paths through the material, so the system is relatively open. The width of the material, and the elegance of the user interface, together with the fact that Perseus may be used for both education and research, makes this a completely unique academic product, which I presume to foresee will become a model for future humanist multimedia productions.

Once again we see that the classical scholars are the first to make use of new computer technology in the humanities. But Perseus is definitely a product which can and must be made available to other humanist milieus – such as archeology, history, art history and comparative literature.

### ***A SUCCESSFUL CONFERENCE – BUT WHERE WERE THE HUMANISTS??***

---

There were many other things also taking place at the conference: stands exhibiting diverse CD-ROM and video disc productions, and, naturally, a number of presentations which I have not touched on here. Personally, I missed the opportunity for discussion after each presentation, but the tight program did not allow for that.

Those responsible for program and organisation deserve unreserved praise for a varied and well composed conference – an exciting encounter between industry and academic research. The progress of the agenda might be read as a good old-fashioned, linear narrative; a rosy and optimistic start (Brown-Martin), problems and objections underway (Brustad and Rannem), and, not least, a happy ending (Mylonas/the Perseus project).

My biggest objection to the conference is concerned neither with the program nor with my personal enrichment, but with the conference fee: NOK 2,000 (in addition to expensive lunches and dinners) for a 2–3-day conference is prohibitively expensive for a student – poor or not. Perhaps for this reason almost nobody from the local humanities milieus participated. Considering the fact that the topic of the conference was very relevant for research as well as teaching in the humanities, this is deplorable, whether the cause is a lack of interest among the humanists, too little marketing or too high a fee. Here we have to look sharp – as was pointed out in the discussion – multimedia is a humanist concern.



# KOMITÉINNSTILLING OM NY TEKNOLOGI PÅ HF-FAKULTETET VED UiB

---

---

I september 1990 nedsatte Rådet for Det historisk-filosofiske fakultet ved Universitetet i Bergen en komite som skulle avgi en innstilling om bruken av ny teknologi innenfor de humanistiske fag. Komiteens leder ble professor *Ståle Dyvik*.

Komiteens mandat var å kartlegge dagens situasjon og formulere mål, utvikle en strategisk plan og foreslå virkemidler for fremtidig utnyttelse av edb. I mandatet lå også at edb-seksjonene fremtidige funksjon og organisasjonsform skulle utredes. Komiteen la frem sin innstilling 3. desember 1990.

Edb innenfor humanistiske fag betraktes av noen som et fagnøytralt hjelpemiddel som ikke blir "humanistisk" før teknologien blir tatt i bruk innenfor disse fagene. Andre hevder at fordi edb er et virkemiddel som innvirker sterkt på valg av tilnæringsmåte og metode, representerer det noe kvalitativt nytt. Komiteen tar ikke stilling til denne problemstillingen, men uttaler at alt peker i retning av å legge mest mulig av edb-relatert forskning, undervisning og formidling til fagmiljøene, og avgrense den rene edb-ekspertisen til det som kan kalles fellesfunksjoner og støttfunksjoner. Det blir presisert i innstillingen at man hele tiden befinner seg i skjæringspunktet mellom fag og teknologi; mellom brukerinteressene ved instituttene og ren edb-ekspertise.

Komiteen henvendte seg først til de forskjellige fagmiljøene for å kartlegge behov og planer.

Følgende hovedpunkter er trukket frem i innstillingen:

- *Databaser*. Her tas med både bruk og opprettelse av databaser som er av interesse for alle instituttene
- *Publisering*. Flere institutter vil skaffe seg kompetanse på "desktop-publishing"
- *Datastøttet undervisning*. Herunder kommer språklaboratorier og interaktiv video
- *Billedbehandling*.
- *Nettverk*. Over hele fakultetet er der interesse for å bedre kommunikasjonen internt og eksternt
- *Forskningsprosjekt*. Her nevnes konkret maskinoversettelse og metoder for tekstanalyse

- *Administrative behov.*
- *Studenter.* Studentene har signalisert behov for utvidet studietilbud i edb ved fakultetet

Etter kartleggingen formulerte komiteen sine konklusjoner i 4 strategiske punkt (forkortet og omskrevet her):

1. Edb-bruken ved fakultetet bør ikke isoleres i en egen avdeling.
2. HF-fakultetet bør ha egen edb-ekspertise uavhengig av instituttene.
3. En edb-avdeling skal være tjenesteytende og formidlende og må til enhver tid kunne forholde seg til de aktuelle behov hos brukerne.
4. En edb-avdeling må ikke dublere fagkunnskap som er lett tilgjengelig utenfor fakultetet.

HF/UiB har siden 1979 hatt en egen edb-seksjon. Komiteen mener at edb-seksjonen i stor utstrekning fungerer etter de ovennevnte prinsipielle retningslinjene. Seksjonens konsulenter har gitt kurs, individuell veiledning og praktisk bistand til ansatte og studenter. De yter også prosjekthjelp f.eks ved oppbygging av databaser. I tillegg har seksjonen tatt seg av driftsfunksjoner og enklere teknisk vedlikehold og også tilrettelagt nettverk og foretatt oppgradering av maskiner.

Ved seksjonen har det siden 1986 vært undervist i delfaget edb for humanister.

Komiteen mener at seksjonen bør fortsette som et tjenesteytende organ for fakultetet, men med et styre som har bredere representasjon fra brukermiljøene. Likeledes bør delfaget edb for humanister opprettholdes, og seksjonen bør overta ansvaret for språklaboratoriene fra Institutt for fonetikk og lingvistikk.

Komiteen viser til kartleggingen av instituttenes behov og ønsker, og nevner omtrent de samme punktene som prioriteringsområder for edb-seksjonen, men uten å rangere dem innbyrdes. Komiteen gjør det klart at seksjonen må styrkes på medarbeidersiden.

Samarbeidet mellom HF-fakultetet og NAVF's edb-senter for humanistisk forskning har vært omfattende og nyttig. Komiteen mener det er ønskelig at Senteret får fortsette sin virksomhet etter noenlunde samme retningslinjer som tidligere. Det anbefales at edb-seksjonen søker å få en enda bedre arbeidsdeling med Senteret. Spesielt nevnes utprøving av programmer, og at man må kalle inn ekspertise fra og samarbeide med Senteret innenfor høyt spesialiserte felt som f.eks. multimedia. Senteret nevnes også som en ressurs innenfor opplæringsarbeidet.

Det historisk-filosofiske fakultet, Universitetet i Bergen. *Implementering og utnytting av ny teknologi på HF-fakultetet : Edb-seksjonens plassering og oppgaver.* Desember 1990.

# DISKUSJONSSEMINAR – INTERAKTIV VIDEO I SPRÅKOPPLÆRING

---

---

*Signe Marie Sanne*

Fredag 14. desember 1990 ble det i Høyteknologisenteret i Bergen arrangert et diskusjonsseminar som tok for seg interaktiv video i språk-opplæring. Seminaret ble arrangert av førsteamanuensis/forsker-NAVF *Signe Marie Sanne* i regi av NAVFs edb-senter for humanistisk forskning. Seminaret var annonsert i universitets- og skoleorganer, og det deltok rundt 20 språklærere. Det var ingen forutsetning å ha edb-kunnskaper for å delta, målsetningen var for det første å gjøre mediet – interaktiv video – kjent for lærerne, for det andre å diskutere hvilken utnyttelse det er fornuftig å gjøre av dette nye mediet.

Som utgangspunkt for diskusjonen ble Sannes prosjekt for italiensk-opplæring grundig demonstrert (se HD nr. 3-90). Siden dette systemet primært er en arbeidsstasjon hvor hver elev/student arbeider alene, var dessuten cand.mag. *Ingrid Hatlestad* spesielt invitert for å snakke om lærerens muligheter til å kontrollere elevens ytelser. Hun viste eksempler fra egenutviklede edb-programmer hvor programmet førte resultatene av elevaktivitetene i en loggbok, hvor noe var tilgjengelig for bare læreren og noe kunne sees av eleven selv.

I løpet av dagen fikk deltagerne informasjon om forskjellige anvendelser av en eksisterende videoplate, fra den rimeligste løsning med bare vanlig fjernkontroll, via en anvendelse med lypenn/strekkoder, en toskjermsløsning og endelig en énskjermsløsning med video, tekst og grafikk på samme skjerm. Det ble også orientert om europeisk og amerikansk videoformat, og muligheten for å skaffe videoplater som kan egne seg til bruk i språkopplæring.

Det meste av diskusjonen dreide seg om utnyttelsen av videomaterialet, hvorvidt læreren burde konsentrere seg om spørsmål og aktiviteter knyttet til språklig opplæring, eller burde gi spørsmål om elementer i kommunikasjons-situasjonen, ekstralingvistiske faktorer, kulturelle faktorer

o.a. For en dokumentarfilm ville det være aktuelt å mate inn tekster som kunne ytterligere utdype det som ble vist på videoen.

Man kom også inn på hva slags type aktiviteter som kan knyttes til et slikt opplegg: Utfyllingsoppgaver, flervalgsspørsmål, ordning i riktig rekkefølge av ord/setninger, rekonstruksjon av en tekst som først vises. For yngre elever ville det være aktuelt å utnytte språkmaterialet i spill, kanskje av typen Hangman, eller i kryssord. Det er lærerens fantasi som setter grenser for nye aktiviteter og nye måter å utnytte videoplaten på.

Det ble også diskutert hvorvidt det er ønskelig at programmet gir eleven anledning til muntlig produksjon, f.eks. ved innspilling av elevens stemme og sammenligning med en innspilt stemme, eller ved at eleven kan overta en av aktørenes replikker. Teknisk sett representerer dette idag ikke noe problem.

Deltagerne så fordelene ved å knytte skriftlige aktiviteter til videomaterialet, med den øyeblikkelige feedback som datamaskinen gir, og med forsterkning i form av visning av nettopp den talesekvensen eleven arbeider med. De mente også at det var fordel å ha både manus, grammatikk og ordbok tilgjengelig for et museklikk til ethvert tidspunkt, og mente at eleven ville benytte seg av dette i langt større grad enn tilsvarende papirutgaver som var tilgjengelig på pulten.

# LESERUNDERSØKELSE OM HUMANISTISKE DATA

---

---

Som våre lesere vil huske, gjennomførte vi i fjor høst en spørreundersøkelse. 474 enkeltabonnenter på HD fikk tilsendt et skjema, der vi spurte hva slags informasjon leserne hadde mest interesse av og hva de mente om tidsskriftet. Vi fikk dessverre bare 114 utfylte skjema i retur. Det gir en svarprosent på ca 24, som er for lavt til å være representativt.

Men det er også andre grunner til at det er vanskelig å trekke vidtrekkende konklusjoner på grunnlag av de innkomne svarene. Svarene tyder på at leserne representerer et bredt spekter av faglige interesser, og at de har ulike følelser for *Humanistiske Data*, som får, mildt sagt, blandet kritikk.

Nedenfor skal vi gjengi noen spørsmål og svar som vi synes det gir mening å analysere, på tross av den lave svarprosenten:

Spørsmål: *Hvilke emner og typer stoff er du mest interessert i?*

| Svaralternativene:                      | antall svar | prosent av avgitte svar |
|---|-------------|-------------------------|
| Edb i humaniora i Norden:               | 69          | 63%                     |
| Teoretiske artikler om edb i humaniora: | 65          | 60%                     |
| Prosjektpresentasjoner:                 | 62          | 57%                     |
| Programvareanmeldelser:                 | 59          | 54%                     |
| Edb og humaniora i resten av verden:    | 50          | 46%                     |
| Edb i undervisning:                     | 46          | 42%                     |
| Produktnyheter:                         | 45          | 41%                     |
| Edb i kultur- og forskningsformidling:  | 38          | 35%                     |
| Bokanmeldelser:                         | 35          | 32%                     |
| Optisk lagring:                         | 33          | 30%                     |
| Multimedia:                             | 33          | 30%                     |

|                         |    |     |
|-------------------------|----|-----|
| Senterets eget arbeid:  | 30 | 28% |
| Tekniske artikler:      | 21 | 19% |
| Edb i biblioteksarbeid: | 11 | 10% |

De fem kategoriene leserne var mest interessert i, kan klassifiseres som "sekkeposter" som kan omhandle ulike emner og ulike fag. Først på en 6. plass kommer et forholdsvis mer avgrenset emne – "Edb i undervisning". De mer spesifikke alternativene har fått en gradvis lavere skåre.

Enkelte av dem som svarte, ønsker mer stoff om bestemte fagfelt: lingvistisk orientert tekstanalyse ved hjelp av tekstbase-program, databaser til bruk i oversettelse, maskinoversettelse, vurderinger av edb og filosofi, edb til hjelp ved handicap/synshandicap og kommunikasjon.

Denne listen ville sikkert ha vært lengre hvis vi hadde fått inn flere svar. Den synes å beskrive dagens situasjon innen edb-bruk i humaniora: Edb-bruken øker, og humanister har fått et uttalt behov for utstyr og programvare som kan løse spesifikke faglige oppgaver.

Vi tolker de innkomne svarene og ønskene slik at HD fortsatt bør orientere og informere om et bredt spekter av feltet edb og humaniora. Vi vil bringe presentasjoner av utstyr og programvare, og rapporter fra prosjekter og arrangementer. Vi vil også prøve å få frem artikler o.l. om de etterlyste fagfeltene. Så langt om HDs faglige innhold.

## ***MEN SÅ VAR DET FORMEN –***

---

Her var lesernes dom svært variert. Noen synes bladet er bra, andre synes det er kjedelig, med gammeldags lay-out og "feil" typografi. Flere ettelyser debatt, og at redaksjonen skal vurdere og ta stilling til programvare og utstyr og ikke bare presentere nyvinningene ukritisk. Dessuten blir tekniske beskrivelser ofte for detaljerte, og dermed utilgjengelige, skriver noen.

Mange lesere betegner HD som et et enestående tiltak – i Norden. Vi har ikke så mange lesere utenfor Norge, men påfallende mange av disse har benyttet anledningen til å gi uttrykk for at HD har en viktig funksjon ved å bringe stoff som ikke er tilgjengelig fra andre kanaler.

Vi i redaksjonen føler at en slik funksjon forplikter oss til å ta hensyn til våre leseres ønsker om forbedringer, men vi kan ikke love å rette på alt som blir kritisert. Vi vil gjøre redaksjonen mer *synlig*, for å gi HD en noe klarere profil, men vi kan ikke love å vurdere kritisk alle presentasjoner av nye program og metoder. Disse vil nok fortsatt måtte stå for forfatterens regning.

Når det gjelder *debatt* i HDs spalter, er vi avhengig av å få inn stoff med spissformulerte meninger. Men det er absolutt mye i feltet edb og humaniora å diskutere, og føler vi oss kallet, skal vi uttrykke oss så klart at noen kan bli provosert til å protestere.

Vi bad også om råd angående bladets format og antall utgivelser pr år, fordi vi har syslet med planer om å legge om til nytt format og hyppigere utgivelser, slik at nyhetene er litt ferskere når de kommer leserne i hende.

Svarene på disse spørsmålene var enda vanskeligere å tolke i noen bestemt retning enn svarene på spørsmål om innhold, så her vil vi prøve rent praktisk hvordan leserne reagerer på et nytt produkt. Neste HD vil få et annet format og utseende, og vi vil gjerne ha reaksjoner fra leserne på det "nye" HD.

Vi takker alle som tok seg tid til å svare på spørreundersøkelsen. Vi har absolutt en viss medfølelse med den som skrev på baksiden av arket: "Unnskyld, men min tid tillater ikke utfylling av denne side". Spørreundersøkelser hører tross alt med til den type post som *kan* gå i papirkurven.



# KONFERANSER

---

---

## **NASJONALT SEMINAR OM TEKSTKODING I BERGEN, 20-21. JUNI 1991**

---

NAVFs edb-senter for humanistisk forskning og Wittgensteinarkivet ved Universitetet i Bergen inviterer til et 2 dagers seminar om metoder og teknikker for koding (også kalt merking eller tagging) av maskinleselige tekster for analyse og presentasjon forskningsarbeid.

Målgruppen er humanistiske forskere og fagmedarbeidere som arbeider med koding av tekster eller med å utvikle systemer for slik bruk. Også andre interesserte er velkomne.

På seminaret vil bl.a. følgende emner bli tatt opp:

- \* SGML (Standard Generalized Markup Language). Internasjonal standard for oppbygging av tekstkodesystemer (ISO 8879-1986).
- \* TEI (Text Encoding Initiative). Internasjonalt samarbeidsprosjekt for standardisering av tekstkoding i humaniora.
- \* MECS (Multi-Element Code System). Tekstkodesystem under utvikling ved Wittgensteinarkivet ved Universitetet i Bergen.
- \* Koding for korpusstudier, med vekt på ikke-trykte kilder.
- \* Koding i terminologisk arbeid.
- \* Hjelpemidler under koding av tekst.
- \* Analyse- og presentasjonsredskaper for bruk av kodet tekst.

Vi inviterer nå til deltakelse på seminaret, som vil gi rik anledning til faglige diskusjoner og demonstrasjoner.

**Du inviteres med dette også til å anmelde et foredrag/innlegg innenfor seminarets emneområde.**

Seminaravgiften på kr. 500,- inkluderer lunsj og forfriskninger.

Avgiften bes innbetalt innen 15. juni til postgirokonto 0802 3384567 eller bankgirokonto 3625.88.53657. Merk giroen "Tekstkodingsseminar".

Vi vil arrangere en uformell sammenkomst om kvelden 20. juni.

Påmelding til seminaret og av foredrag/innlegg ønskes snarest mulig og ikke senere enn 10. juni til: *Øystein Reigem, NAVFs edb-senter for humanistisk forskning, Postboks 53 – Universitetet, N-5027 Bergen. Tlf: + 47 (0)5 - 21 29 54/55/55. Fax: + 47 (0)5 - 32 26 56. E-mail: fafor@nobergen.earn, eller Claus Huitfeldt, Wittgensteinarkivet, Universitetet i Bergen, Strømggt. 53, N-5007 Bergen. Tlf: + 47 (0)5 - 21 29 50. Fax: + 47 (0)5 - 32 27 57. E-mail: huitfeldt@cc.uib.no.*

## **CIDOC ANNUAL MEETING 25-30 MAY 1991**

---

*CIDOC* er *ICOM's Documentations Committee*. Den årlige konferansen har "pre- og post-meetings" i tillegg til hoveddelen. Disse vil bli holdt på *The National Museum, Documentation Department i København*. Hoveddelen av konferansen vil bli holdt ombord i *King of Scandinavia* som vil gå tur/retur København-Oslo.

Programmet veksler mellom foredrag, demonstrasjoner og "working groups" og konsentreres om edb-anvendelser for informasjonshåndtering og gjenstandsregistrering ved museer og innenfor arkeologi. En del tid er også viet mulighetene for å koble dokumentasjonssystemer og bibliografiske databaser sammen med databaser for registrering av museums-gjenstander, arkeologiske funn etc.

Flere opplysninger fås fra: *The National Museum, Documentation Department, Ny Vestergade 11. back building, 2nd floor, DK-1471 Copenhagen K, Denmark*

## **NATIONAL ARCHAEOLOGICAL RECORDS CONFERENCE 30 MAY - 1 JUNE 1991**

---

Denne konferansen blir arrangert i tilknytning til *CIDOC Annual Meeting* som er nevnt ovenfor, og blir også holdt på *The National Museum, Documentation Department, København*.

Tema er bruk av ny teknologi ved registrering av arkeologiske gjenstander og data, og det vil bli lagt vekt på både teoretiske og praktiske aspekter. Der vil bli foredrag og presentasjoner fra mange europeiske land, men hovedtyngden av foreleserne kommer fra de nordiske land.

Frist for registrering er 15. april, og adressen er den samme som for *CIDOC Annual Meeting*. (Se ovenfor.)

## **INDIVID OCH STAD - LONGITUDINELLA ANALYSER I URBANHISTORISK FORSKNING**

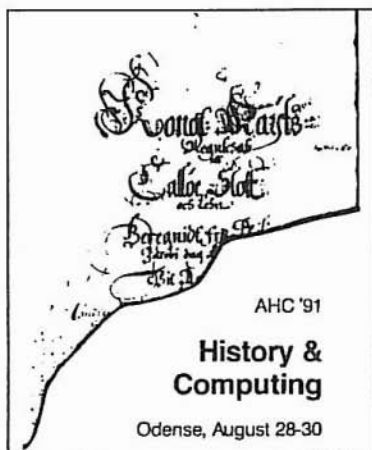
---

Dette er et kurs innefor rammen av Nordiska forskarutbildningsakademins kurser, og arrangeres i tiden 5-16. august. Arrangør er Stockholms Historiska Databas i samarbeid med Stockholms Universitet.

Tema som vil bli tatt opp er migrasjon, dødelighet, fattigdom og familie- og husstandsstrukturer. Der blir forelesninger og gruppearbeid, og deltakerne skal også gjennomføre selvstendige prosjekt med utgangspunkt i data som finnes i SHD - Stockholms Historiska Databas.

Kurset gis for 20 deltakere fra de nordiske land. Reise og oppholds-utgifter dekkes av Nordisk Ministerråd. Kursavgiften er SEK 500.-

For flere opplysninger, kontakt: *Stockholms Historiska Databas, S:t Eriksgatan 121, S-113 43 Stockholm. Tlf: + 46 8 333810* eller *Gunnar Thorvaldsen, Riksarkivet, boks 10 Kringsjå, 0807 Oslo 8. Tlf: + 47 (0)2 23 74 80. E-mail: Gunnar.Thorvaldsen@kari.uio.no.*



***ASSOCIATION FOR HISTORY AND COMPUTING – 6TH INTERNATIONAL CONFERENCE, ODENSE, DENMARK, AUGUST 28-30, 1991***

---

ACH er en internasjonal organisasjon som arbeider for å fremme bruken av edb på alle nivå av historiske studier. Årets konferanse arrangeres av *Dansk Data Arkiv i Odense*.

*Demografiske studier* og *store databaser* blir viktige emner ved konferansen, og arrangørene inviterer også forskere til å forelese over emner som:

- Standardisering og utveksling av maskinleselige data i de historiske disipliner
- Dataanalyse og -presentasjon
- Analyse av historiske begivenheter
- Tekstanalyse
- Simulering og modellering
- Edb-støttet undervisning
- Sosial og økonomisk historie
- Kvantitative metoder

For ytterligere informasjon om konferansen, kontakt: *Hans Jørgen Marker, Danish Data Archives, Munkebjergsvænget 48, DK-5230 Odense*

M, Denmark. Tlf: + 45 66 15 79 20. Fax: + 45 66 15 83 20. E-Mail: DDAHM@NEUVM1 (earn).

## **ICHIM '91**

---

*International Conference on Hypermedia and Interactivity in Museums.* Pittsburg, Pennsylvania 14–16. oktober 1991.

Konferansen, som er den første i sitt slag, har til hensikt å fremme interessen for bruk av interaktive media i museer. Målgrupper er museumsfolk av alle kategorier og personer som utvikler programmer innenfor multimedia og hypermedia. Det blir veksling mellom foredrag og demonstrasjoner og der blir også en salgsmesse i tilknytning til konferansen.

De som har liten kjennskap til de nye mediene, kan delta i en "Pre-Conference Workshop" 13. oktober. Flere opplysninger fås fra: ICHIM '91, Archives & Museum Informatics, 5501 Walnut Street, Suite 203, Pittsburg, Pennsylvania 15232-2311 USA. Tlf.: 412-683-9775 Fax: 412-683-7366

## **DET DIGITALE BILDE: ESTETISKE FAG OG INFORMASJONSTEKNOLOGI**

---

Konferanse på Bergen Airport Hotel 17–19 oktober 1991, arrangert av Kunsthistorisk institutt, Universitetet i Bergen, NAVFs edb-senter for humanistisk forskning og IBM Bergen Scientific Centre.

I løpet av det siste ti-året har informasjonsteknologien utviklet muligheter for å lagre, analysere og presentere bilder, film og video ved hjelp av datateknikk. Dette er et forskningsfelt i rask utvikling, og teknologien gir stadig flere muligheter for rask tilgang til nye typer informasjon.

På konferansen vil denne type humanistisk fagarbeid bli presentert gjennom en rekke foredrag og presentasjoner av fremtredende forskere fra inn- og utland.

Blant prosjekt som blir presentert:

- \* Project Emperor I (Kinas første keiser og hans terrakottakrigere og -hester)
- \* The Museum Education Consortium Hypermedia Project (impresjonisme og post-impresjonisme)
- \* Det nederlandske Åpne universitets prosjekt for visuell kunst (undervisning i kunsthistorie)
- \* The Brandywine Project (billed-database over kunstneren Andrew Wyeth)

- \* EFs prosjekt Narcisse (vitenskapelig databank over kunstverk i europeiske museer) og Vasari (digital analyse av kunstverk)
- \* Italienske kunst- og kulturhistoriske prosjekt
- \* Perseus (klassisk gresk litteratur og arkeologi)
- \* Hypermedia-prosjekt i egyptisk kulturhistorie
- \* Hypermedia i bruk ved Munch-museet, Oslo
- \* FotoMac – et program for bilderegistrering
- \* Interaktiv video i filmanalyse
- \* Digitalt basert stilforskning
- \* Bruk av infrarød bestråling ved digital analyse av maleri.

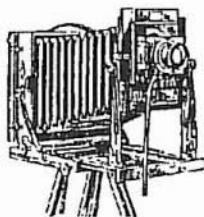
Konferansen vil være av interesse for:

Fagpersonale og studenter innen kunst- og kulturhistoriske fag, film, teater og mediefag, museums- og biblioteksfag, pedagogiske utviklingsmiljøer, forlagsbransjen og informasjonsteknologiske utviklingsmiljøer.

Henvendelser rettes til: *NAVFs edb-senter for humanistisk forskning, Postboks 53 Universitetet, N-5027 Bergen, Tlf: + 47 (0)5 21 29 55/56. Fax: + 47 (0)5 32 26 56. E-mail: imageconf-91@navf-edb-h.uib.no (Kari Sørstrømmen / Kjell Morland)*

eller til:

*Britt Kroepelien, Høyteknologisenteret i Bergen, Tlf: + 47 (0)5 54 42 28*



## **FOTOKONFERANSE I SANDEFJORD**

*Fotografiet: Medium – Middel – Mål. Landskonferanse Park Hotel, Sandefjord, 15–17 november 1991.*

Det er klart for ny fotokonferanse i Sandefjord. Temaene for konferansen er konsentrert om bruk og formidling av foto i dag – i motsetning til landskonferansen i 1989 hvor hovedvekten lå på arbeid med bevaring av eldre materiale. Emnet blir behandlet fra ulike synsvinkler som bruk av foto i samtidsdokumentasjon, i bok, på film og i utstilling, som formidler og som selvstendig uttrykk med egenverdi. Foredragsholderne

vil bli hentet fra et bredt spekter av Nordens fagekspertise, og i tillegg til plenumsforedrag og debatt blir det også utstillingsmesse med demonstrasjon av teknisk utstyr.

Arrangører er Norsk kulturråd, Sekretariatet for fotoregistrering, NAVFs edb-senter for humanistisk forskning og NKKM.

Fullstendig program sendes ut i august/september. Flere opplysninger fås fra: *Norske Kunst- og Kulturhistoriske Museer, Ullevålsvn. 11, N-0165 Oslo 1. Tlf: + 47 2 20 14 02. Fax: + 47 2 11 23 37.*

## ***NORDISKE DATALINGVISTIKKDAGER***

---

blir arrangert ved Universitetet i Bergen 28.–30. november 1991.

Arrangører er *edb-miljøene ved Det historisk-filosofiske fakultet ved Universitetet i Bergen* og *NAVFs edb-senter for humanistisk forskning.*

Målsettingen for årets nordiske forskersymposium/konferanse er å formidle kunnskaper og kontakter mellom datalingvister i Norden, og å presentere pågående forskning innenfor fagfeltet.

Arrangørene har valgt å fokusere på disiplinene *maskinoversettelse* og *datamaskinell leksikografi/terminologi.*

Blant foreleserne:

*Bente Maegaard, Center for Sprogteknologi, København*

*Fred Karlsson, Universitetet i Helsinki*

*Helge J. Dyvik, Universitetet i Bergen*

Forskere innbys til å gi demonstrasjoner eller forelese over emner som f.eks.:

- automatisk syntaktisk og/eller semantisk analyse
- tekstgenerering
- diskursanalyse
- dialogsystemer
- talesyntes/gjenkjenning
- dokumentbasert informasjonshenting eller
- modellering av språkforståelse

Påmeldingsfrist for foredrag/demonstrasjoner er 15. mai.

Påmeldingsfrist for deltakere: 15. oktober

For ytterligere informasjon, kontakt: *Amanuensis Ivar Utne, Nordisk institutt, Universitetet i Bergen, Tlf: + 47 (0)5 21 2404 / 21 2970* eller sekretariatet: *NAVFs edb-senter for humanistisk forskning, Postboks 53 Universitetet, N-5207 Bergen. Tlf: + 47 (0)5 21 29 55/56. Fax: + 47 (0)5 32 26 56. E-mail: dataling-91@navf-edb-h.uib.no (Kari Sørstrømme / Kjell Morland)*

# KONFERANSEKALENDER

Redaksjonen kan være behjelpelig med mer informasjon om de fleste av arrangementene som er nevnt nedenfor.

Ytterligere opplysninger finnes dessuten i den elektroniske meldingstjenesten som kan nås med brev til NAVFSERV@NORA.NAVF-EDB-H.UIB.NO og SEND KONFERANSER KONF som eneste linje i brev.

## JUNI

- 7-8 **International Workshop on Man-machine Interaction in Live Performance**, Computer Music, Computer Graphics, Robot Art..., Pisa, Italia. Informasjon: Leonello Tarabella, Computer Music Dept. of CNUCE/CNR, via S.Maria 36 - 56126 Pisa, Italia. Tlf: + 39 50 593276. Fax: + 39 50 576751. E-mail: music5@icnucevm.cnuce.cnr.it.
- 14-15 **STELLA Symposium**. The Use of Computers in English. First Announcement and Invitation to Contribute. Informasjon: Mary Pat Gibson, STELLA Symposium, 6 University Gardens, University of Glasgow, Glasgow G12 8QQ. Fax: + 041 330 4808. E-mail: stella@uk.ac.glasgow.vme.
- 16-19 **NordDATA '91**. Oslo Spektrum/Universitetet Blindern. Informasjon: Univers Congress Service. Postboks 54 Blindern, N-0313 Oslo 3, Norge. Tlf: + 47 2 45-32-44. Fax: + 47 2 45-32-39.
- 17 **Lexical Semantics and Knowledge Representation**. University of California Berkeley, California, USA. Informasjon: Peter Norvig, 573 Evans Hall, Berkeley, CA 94720, USA. Tlf: + 1 415 642-9533. E-mail: norvig@teak.berkeley.edu.
- 17 **Reversible Grammar in Natural Language Processing**. University of California Berkeley, California, USA. Informasjon: Peter Norvig, 573 Evans Hall, Berkeley, CA 94720, USA. Tlf: + 1 415 642-9533. E-mail: norvig@teak.berkeley.edu.
- 18 **Workshop on Evaluation of Natural Language Processing Systems**. University of California Berkeley, CA. Informasjon: Jeanette G. Neal, Ph.D. (Committee Chair) Calspan Corporation. P.O. Box 400, Buffalo, NY 14225, USA. Tlf: + 1 716 631-6844. Fax: + 1 716 631-6722. E-mail: neal@cs.buffalo.edu.



- 18-21 **29th Annual Meeting of the Association for Computational Linguistics.** University of California Berkeley, CA, USA. Informasjon: Don Walker (ACL), Bellcore, MRE 2A379, 445 South Street, Box 1910, Morristown, NJ 07960-1910, USA. Tlf: + 1 201 829-4312. E-mail: walker@flash.bellcore.com.
- 24 **ICLP Workshop.** Advanced Logic Programming Tools and Formalisms for Language Processing. INRIA, Paris-Rocquencourt, Frankrike. Informasjon: Patrick Saint-Dizier, IRIT Universite Paul Sabatier, 118 route de Narbonne F-31062 Toulouse, Frankrike. Tlf: + 33 61 556344. E-mail: stdizier@irit.irit.fr.
- 25-27 **The MultiMedia '91,** The First European Multimedia Event, Olympia 2, London. Informasjon: Blenheim Exhibitions & Conferences Ltd. Blenheim Online, Blenheim House, Ask Hill Drive, Pinner Middlesex, HA5 2AE, UK. Tlf: + 44 081 868 4466. Fax: + 44 081 868 9933.
- 25-28 **Third International Conference on Artificial Intelligence and Law.** St. Catherine's College, Oxford, UK. Informasjon: Carole Hafner, College of Computer Science, Northeastern University, Boston, MA 02115 USA. Tlf: + 1 617 437-5116. Fax: + 1 617 437-5121. E-mail: hafner@corwin.ccs.northeastern.edu.

## **JULY**

---

- 2-4 **MT SUMMIT III,** Conference on Machine Translation. Washington, DC . Informasjon: MT Summit III, Organizing Committee Center for Machine Translation, Carnegie Mellon University Pittsburgh, PA 15213-3890 USA. Tlf: + 1 412 268-6591. Fax: + 1 412 268-6298. E-mail: mtsummit@cs.cmu.edu (arpanet).
- 2-5 **1st Workshop of the Principia Cybernetica Project.** Computer-supported Cooperative Development of an Evolutionary-systemic Philosophy. Free University of Brussels, Belgia. Informasjon: Francis Heylighen, PESP, V.U.B., Pleinlaan 2, B-1050 Brussels, Belgia. Tlf: + 32 2 641-25-25. Fax: + 32 2 641-22-82. E-mail: Z09302@BBRBFU01 (bitnet).
- 8-21 **Summer School in Computational Linguistics Formal and Computational Models of Meaning.** Charles University Prague, Czechoslovakia. Informasjon: Dr. Eva Hajicova, MFF UK - Linguistics, Malostranske n. 25, CS - 118 00 Prague 1, Czechoslovakia.
- 10-12 **12th International SCCAC (Society for Conceptual and Content Analysis by Computer) Meeting,** Mannheim/Germany. Concepts and Computers.

- 14-15 **AAAI-91 Intelligent Multimedia Interfaces Workshop** (to be determined) Anaheim, CA, USA. Informasjon: Mark Maybury (chair) The MITRE Corporation, Bedford, MA 01730. E-mail: maybury@linus.mitre.org.
- 15-18 **Multi-Media Information**. The Second International Information Research Conference. Churchill College, Cambridge. Informasjon: Karen Merry, Multi-Media Information Conference, The British Library R&D Department, 2 Sheraton Street, London W1V 4BH, UK. Tlf: + 44 071 323 7050
- 20-21 **ASL/LSA – Conference on Logic and Linguistics**. University of California at Santa Cruz, USA. Informasjon: Richard Oehrle, Department of Linguistics, Douglass 200E, University of Arizona, Tucson, AZ 85721. E-mail: oehrle@ccit.arizona.edu.

## AUGUST

---

- 4-7 **The International Conference on the Learning Sciences**. Informasjon: The International Conference on the Learning Sciences, Northwestern University, 1890 Maple Avenue, Suite 150 Evanston, Illinois 60201, USA. Fax: + 1 708 491 5258. E-mail: icls@aristotle.ils.nwu.edu.
- 5-16 **Individ och stad – longitudinella analyser i urbanhistorisk forskning**. Et tvärvetenskaplig forskarkurs anordnad av Stockholms Historiska Databas i samarbete med Stockholms Universitet. Se utfyllende omtale under vignetten konferanser.
- 10-12 **The Sixth International Conference on Computing and Philosophy** (6th CAP). New Haven, Connecticut at Southern Connecticut State University. Informasjon: Robert J. Cavalier, CDEC, Carnegie Mellon University, Pittsburg, PA 15213. E-mail: rc2z+@andrew (bitnet) eller rc2z+@andrew.cmu.edu (internet).
- 12-23 **The Third European Summer School in Language, Logic and Information**, Universitaet des Saarlandes Saarbruecken. Informasjon: Hans Uszkoreit/Maike Paritong. Tlf: + 49 681 302-3714/4115. Fax: + 49 681 302-4351. E-mail: ess@coli.uni-sb.de.
- 14-17 **International Association for the Semiotics of Law**. 7th International Colloquium. Utrecht, The Netherlands. Informasjon: Prof. Dr. Paul J. van den Hoven, Vakgroep Nedelands, Rijksuniversiteit Utrecht, 3512 JK Utrecht, Nederland. Tlf: + 31 30 394093/394000. Fax: + 31 30 333380.
- 21-23 **Nordisk Konferens för Datorstödd Utbildning inom Universitet och Högskolor**. Tekniska högskolan Otnäs, Finland. Informasjon:

Nordisk Konferens för Datorstödd Utbildning/Sekretariat/Sutinen  
Tekniska högskolan, Fortbildningscentralen, Norrskensvägen 6,  
SF-02100 Esbo, Finland.

Tlf: + 358 90 4511. Fax: + 358 90 451 4068.

E-mail: TAK-JPE@FINHUT.HUT.FI.

- 24-30 *International Joint Conferences on Artificial Intelligence*. IJ-  
CAI-91. Sydney, Australia. Informasjon: Dr. Donald E. Walker,  
Secretary-Treasurer, IJCAI-91, Bellcore, MRE 2A379, 445 South  
Street, Box 1910, Morristown, NJ 07960-1910, USA.

Tlf: + 1 201 829-4312.

E-mail: walker@flash.bellcore.com (internet).

- 26-30 *AIBI-3: 3rd International Conference of Association Interna-  
tionale Bible et Informatique*. Tübingen, Tyskland. Informasjon:  
Prof. Dr. Harald Schweitzer, Reutlinger Strasse 12 D-7400 Tü-  
bingen, Tyskland.

Tlf: + 49 70761 295248. Fax: + 49 7071 295404.

E-mail: USBA001@DTUZDV2 (earn).

- 28-30 *Association for History and Computing*. 6th International Con-  
ference. Odense, Denmark. Informasjon: Hans J. Marker, Danish  
Data Archives, Munkebjergvænget 48, DK-5230 Odense M, Dan-  
mark. Tlf: + 45 66 157920. Fax: + 45 66 158320.

E-mail: DDAHM@NEUVM1 (earn).

## SEPTEMBER

---

- 9-12 *UMIST, Intelligent Computer-Aided Language Learning Work-  
shop Centre for Computational Linguistics*. Informasjon: Chris  
Bowerman. Tlf: + 44 061 200 3089. Fax: + 44 061 228 7040.  
E-mail: CHRIS@UK.AC.UMIST.CCL eller Christoph Zaehner,  
Tlf: + 061 200 3081. E-mail: CHRISTOPH@UK.AC.UMIST.CCL.  
Adresse: Centre for Computational Linguistics, UMIST, PO Box  
88, Manchester, M60 1QD, England.

- 18-20 *Conference on CALL and HYPERMEDIA*, University of Exeter.  
Informasjon: Keith Cameron, Department of French, The Uni-  
versity, Exeter, EX4 4QH, England.

E-mail: cameron@uk.ac.exeter.

- 23-27 *First Quantitative Linguistics Conference (QUALICO)*. Univer-  
sity of Trier, Tyskland. Informasjon: R. Kochler, University of  
Trier, P.O.Box 3825, D-5500 Trier, Tyskland.

- 29-1 *Using Corpora*, 7th Annual Conference of the University of  
Waterloo Centre for the New OED and Text Research, St.  
Catherine's College Oxford, England. Informasjon: Prof. Frank

Tompa, UW Centre for the New OED and Text Research, Davis Centre, University of Waterloo, Waterloo, Ontario, Cannada N2L 3G1. E-mail: newoed@uwaterloo.ca.

## OKTOBER

---

- 4-6 *Applied Linguistics at Michigan State*, Theory Construction and Methodology in Second Language Acquisition Research. Informasjon: Alan Beretta/Susan Gass, Department of English, 201 Morrill Hall, Michigan State University, E. Lansing, MI 48824-1035.  
Tlf: + 1 517 353-0800. Fax: + 1 517 336-1149.  
E-mail: 21910mgr@MSU eller 21003smg@MSU (bitnet).
- 14-16 *ICHIM '91*, International Conference on Hypermedia and Interactivity in Museums. Sheraton Station Square Pittsburg, Pennsylvania, USA. Se fylldig omtale under vignetten konferanser.
- 17-19 *Det digitale bilde*, Estetiske fag og informasjonsteknologi, Bergen. Arrangører: Kunsthistorisk institutt, Universitetet i Bergen, IBM Scientific Centre og NAVFs edb-senter for humanistisk forskning. Se utførlig omtale under vignetten konferanser.
- 24-27 *Second European Conference of Music Analysis*. Trento, Italia. Informasjon: The Accademia Filarmonica Trentina, via Oriola 12, 38100 Trento, Italia.  
Tlf: + 39 461 238008. Fax: + 39 461 238166.

## NOVEMBER

---

- 3-4 *CIMI, Computer Interchange of Museum Information*, Santa Monica, CA, USA. Informasjon: John Perkins, CIMI Project Manager, 5659 Merkel St., Halifax, NS, B3K 2J1, Canada.  
Tlf: + 1 902 454 4077. Fax: + 1 902 453 6153.
- 13-15 *International Symposium on Artificial Intelligence Applications in Informatics*. Cancun, Mexico. Informasjon: Hugo Terashima, Centro de Inteligencia Artificial, ITESM. Ave. Eugenio Garza Sada 2501, Col.Tecnologico C.P. 64849 Moterrey, N.L. Mexico.  
Tlf: + 52 83 58-2000. Fax: + 52 83 58-1400.  
E-mail: ISAI@tecmyvm (bitnet)  
eller ISAI@tecmyvm.mty.itesm.mx.
- 28-30 *Nordiske Datalingvistikkdager*, Universitetet i Bergen. Arrangører: Fagmiljøer ved HF-fakultetet ved Universitetet i Bergen og NAVFs edb-senter for humanistisk forskning. Informasjon:

NAVFs edb-senter for humanistisk forskning, Postboks 53, Universitetet, N-5027 Bergen.  
Tlf: + 47 5 212954/55/56. Fax: + 47 5 322656.  
E-mail: dataling-91@navf-edb-h.uib.no.

## **DESEMBER**

---

- 3-6 *Third Symposium Social Communication*, Santiago de Cuba. Informasjon: Eloína Miyares Bermúdez, Secretaria Ejecutiva Delegación Territorial ACC, Avda. Manduley No. 308, esq. 13, Aptdo. 4065. Vista Alegre, Saniago de Cuba. CP 90400.  
Tlf: + 53 4-1458/4-2787/4-2209/4-2760.
- 15-18 *Hypertext '91* Conference, San Antonio, Texas. Informasjon: General Chair John J. Leggett, Hypertext '91 Conference, Hypertext Research Lab, Department of Computer Science, Texas A&M University, College Station, TX 77843 USA.  
Tlf: + 1 409 845 0298. Fax: + 1 409 847 8578.  
E-mail: leggett@bush.tamu.edu.
- 16-18 *CATH '91* Computers and Teaching in the Humanities, University of Durham. Informasjon: Audrey McCartan, Caedmon Building, School of Education, Leazes Rd., Durham, DH1 1TA, England.

## **1992**

---

### **APRIL**

---

- 1-3 *3rd Conference on Applied Natural Language Processing*, Trento, Italia. Informasjon: Tullio Grazioli/Oliviero Stock, I.R.S.T. 38050 Povo (Trento), Italia.  
Tlf: + 39 461 814444. Fax: + 39 461 810851.  
E-mail: interne@irst.it.
- 5-9 *ALLC-ACH 1992*, Christ Church, Oxford, England. Informasjon: ALLC-ACH92, Centre for Humanities Computing, Oxford University Computing Service, 13 Banbury Road, Oxford, OX2 6NN. Tlf: + 44 0865 273200. Fax: + 44 0865 273275.  
E-mail: ALLCACH@VAX.OX.AC.UK.
- 7-10 *EP '92*. International Conference on Electronic Publishing Document Manipulation and Typography, Lausanne. Informasjon: EP92, Swiss Federal Institute of Technology, IN Ecubiens, CH-1015 Lausanne, Sveits.  
Tlf: + 41 2169 32575. Fax: + 41 2169 35263.  
E-mail: ep92@eldi.epfl.ch.

## ***JULI***

---

- 23–28 ***COLING-92***. Fourteenth International Conference on Computational Linguistics. Nantes, Frankrike. Informasjon: Prof. A. Zampolli, Università di Pisa, ILC, via della Faggiola 32, I-56100 Pisa, Italia. Tlf: + 39 50 560481. Fax: + 39 50 589055.

## ***AUGUST***

---

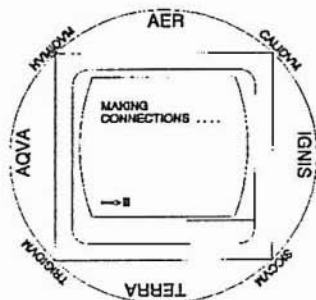
- 9–14 ***XVth International Congress of Linguists***, Quebec City, Canada. Informasjon: CIL92, Department de Langues et Linguistique, Université Laval, Quebec City, (Que.), G1K 7P4, Canada. Tlf: + 1 418 656-5353. Fax: + 1 418 656-2019. E-mail: CIPL92@LAVALVM1 (bitnet).

# FRA ACH/ALLC '91

---

---

*Per Vestbøstad/Jostein H. Hauge*



ACH/ALLC '91

March 17-21, 1991  
Arizona State University  
Tempe, AZ

## "MAKING CONNECTIONS" I ARIZONA

---

---

*Humanistiske Data* har hatt en samtale med Per Vestbøstad som for kort tid siden kom tilbake fra Arizona State University etter å ha deltatt på konferansen "Making Connections" som ble holdt i dagene 17-21. mars.

Konferansen ble arrangert av Association for Computers and the Humanities og Association for Literary and Linguistic Computing.

Per Vestbøstad forteller at forventningene til konferansen var store etter deltakelse på konferansen for to år siden i Toronto, hvor ikke minst demonstrasjonsdelen av konferansen var imponerende. På den internasjonale konferansen i Arizona var det ca. 230 deltakere, herav bare ca. 40 deltakere fra Europa – noe som trolig skyldes den politiske



situasjonen og advarslene som var gitt mot flyreiser i tiden forut for konferansen, sier Vestbøstad.

I forhold til Toronto var demonstrasjonsdelen av opplegget en skuffelse, ettersom denne delen var trengt sammen til noen få timer og foregikk i trange klasserom. De faglige presentasjonene var imidlertid svært omfattende og dekket et bredt spekter av emner, særlig innenfor det en tidligere kalte datamaskinell språkbehandling, som bl.a. omfatter datamaskinell tekstanalyse og leksikografi.

Vestbøstad forteller at stor oppmerksomhet ble viet det internasjonalt drevne utviklingsarbeidet som er knyttet til tekstkoding (Text Encoding Initiative) hvor det bl.a. er en norsk gruppeformann, professor Stig Johansson, Universitetet i Oslo. (Se egen omtale på s. 121)

Faren er at man – slik målsettingen også er – har tenkt på alle finurligheter i forbindelse med koding av tekster i skriftlig og talt form i så stor detalj at den som trenger enkle løsninger på sitt tekstproblem gjerne går seg vill i alle detaljene, sier Vestbøstad, som legger til at et norsk system for tekstkoding utviklet innenfor Det norske Wittgenstein-prosjektet i Bergen også fikk positiv omtale på konferansen.

Vestbøstad mener ellers at interessen for bygging og bruk av store tekst- og språkdata-baser er sterkt stigende i mange land, f.eks. i USA og Canada. Datamaskinell tekstanalyse bygget på store corpora synes nå å være en vitenskapelig stuere aktivitet, men nå er det ikke snakk om million-corpora, men om å legge til rette tekstsamlinger i størrelsesorden 40–50 millioner ord. På spørsmål om det er store forskjeller på den type arbeid som ble presentert på denne internasjonale konferansen og det som foregår på feltet i vårt land, sier Vestbøstad at det neppe er store forskjeller i kvalitet, men heller i prosjektstørrelse og bredde på aktivitetene.

Av fagfelt som hadde egne sesjoner på konferansen, nevner Vestbøstad automatisk innholdsanalyse, semantikk, databaser som utnytter kunnskapsbaserte metoder, ekspertsystemer og leksikografi og metaforstudier. På flere av disse feltene er det heller liten aktivitet i Norge, sier Vestbøstad.

Men ikke alle synes at metodeutviklingen gir resultater fort nok. På konferansen ble et hovedforedrag holdt av en av nestorene på feltet, Professor Martin Kay fra Xerox Park i Palo Alto, California. Han undret seg over at en stor del av de temaene og problemene som ble tatt opp, allerede var formulert på 70-tallet. En av grunnene kan kanskje være, sier Vestbøstad, at det nå er en ny generasjon forskere som tar fatt. De vil da gjerne med stort pågangsmot løfte på steiner som også generasjonen før har løftet på uten at den nye generasjonen kanskje alltid vet om det, sier Vestbøstad. (Se egen omtale av prof. Kays foredrag på s. 122)

Hovedgevinsten med å overvære en slik konferanse, mener Vestbøstad, ligger i å kunne "ta pulsen på" hele fagfeltet akkurat nå og oppnå nye faglige kontakter. Og på dette felt skjer det i alle fall noe nytt, forteller Vestbøstad: Når du melder deg på en slik konferanse, vil du heretter automatisk bli medlem av det elektroniske universitet: Nesten alle deltakere oppgir nå E-mail-adresser. At dette nettverket fungerer fra første stund, fikk Vestbøstad også erfare: Før han var kommet hjem, var det allerede en – elektronisk – melding i posten fra en av deltakerne som han ikke hadde hatt direkte kontakt med på konferansen.

Vestbøstad forteller til slutt at konferansen arrangementsteknisk ble fint gjennomført, og at det forelå godt utarbeidede sammendrag til nesten alle foredragene. Det siste har redaksjonen tenkt å la leserne av *Humanistiske Data* få glede av. På sidene 125–132 finner du en oversikt over konferanseforedragene. Dersom du kopierer listen og krysser av for de sammendrag som du ønsker kopier av, og sender kopien til oss, vil vi sende deg tekstene i retur – om ikke ønsket ditt er altfor omfattende.

Neste ACH/ALLC-konferanse blir holdt i Oxford i England i 1992.

## **TEI**

---

---

Hovudet på TEI sine brevark forklarar tiltaket som eit "Initiative for Text Encoding Guidelines and a Common Interchange Format for Literary and Linguistic Data".

Text Encoding Initiative er styrt av dei tre internasjonale organisasjonane Association for Computers and the Humanities (ACH), Association for Computational Linguistics (ACL) og Association for Literary and Linguistic Computing (ALLC), og støtta av mange fleire organisasjonar, som alle er opptekne av å fremja tekstorientert forskning.

Og tekst skal her forståast i svært vid meining: TEI tek sikte på å laga retningslinjer for merking av alt frå ordlister og ordbøker via faglitteratur og avistekst til poesi og talt språk.

Legg elles merke til at dette ikkje er nok eit gluft kodesystem for merking av tekst, men overordna retningslinjer for kva særdrag ved ein teksttype som bør merkjast for at tekstane skal kunna brukast mest mogleg allsidig i ulike slag forskning. Utveksling av maskinlagra tekstar mellom forskarar har ofte vore hemma av at den einskilde forskaren sine snevre interesser har prega den merkinga som er lagt inn i teksten, slik at lite eller ingenting av dette kan nyttast til å studera andre fenomen.

Eit akseptert sett av retningliner vil gjera det lettare for forskarar og andre interesserte å utveksla bearbeidde tekstar og såleis byggja vidare på kvarandres innsats.

Hausten 1990 la TEI fram eit utkast (versjon 1.0) til slike retningliner. Her vert det gjeve råd om korleis ein kan handtera fenomen som diakritiske teikn, teikn i ikkje-romanske alfabet, spesialsymbol o.s.b., fotnotar, tekstkritisk apparat, paralleltekst og andre ikkje-linjære innslag i tekstar som skal omformast til ein linjær datastrøm. Den logiske oppdelinga av tekstar t.d. i bok, kapitel, avsnitt eller i akt, scene og replikk, informasjon om opphavet til teksten, kven som er ansvarleg for innlagd merking, utgjevar m.v.

Utkastet tilrår også eit utvekslingsformat for tekst som skal vera medium- og programvare-uavhengig.

Utkastet har no vore ute til vurdering minst eit halvår. Ein av krumtappane i arbeidet med utkastet, C.M. Sperberg-McQueen (Univerity of Illinois at Chicago) oppsummerte reaksjonane slik på ACH/ALLC-konferansen i Tempe, Arizona: Dei fleste meiner utkastet er rimeleg velegna til å byggja vidare på, sjølv om mange har merknader til detalj-løysingar. Einskilde forskarar har meint at tilrådingane er for strenge, andre at dei er for vide og upresise. Han venta at mange detaljar i utkastet kom til å verta endra, men såg førebels reaksjonane som ei stadfesting av at dei var på rett veg med dei store retninglinene. I tida fram til dei endelege retninglinene skal publiserast, truleg i juni 1992, skal retninglinene testast på realistiske mengder med varierte tekstar, parallelt med at nye utkast kjem til frå dei arbeidsgruppene som ikkje har levert tilrådingar enno.

Eit eksemplar av utkastet kan tingast kostnadsfritt hjå: *Oxford University Computing Service, 13 Banbury Road, Oxford OX2 6NN, UK*. Ytterlegare eksemplar må betalast med £20,- pr. stk.

## **"MARTIN KAY BACK IN THE HUMANITIES"**

---

---

Forventningane i salen vart utløyst i eit latterbrøl då den velkjende datalingvisten Martin Kay, æresdoktor ved Göteborgs Universitet og professor i lingvistik ved Stanford University, innleia åpningsforedraget ved ACH/ALLC-konferansen slik: "It is good to be back in the Humanities after thirty years in the Engine room"! Replikken spela på dei mange åra han har vore forskar ved Xerox sitt forskingssenter i Palo Alto, CA. No har vel Martin Kay hatt eit bein i kvar leir mesteparten av tida som er gått sidan han etablerte seg som datalingvist på 1960-talet.

Han har i det minste halde fleire kurs i datalingvistik i Europa i denne perioden. I 1973 var han t.d. hovudforelesar på Nordisk sommarkurs i Bergen.

Tittelen på hans "Keynote speech" var "Remarkable and Unremarkable Language".

Han opna med nokre muntre, sjølvvironiske utsegner, der han såg seg sjølv som sermonimeister og forsamlinga som "dei truande". Vidare oppsummerte han datalingvistikken sine bragdar dei siste tredivje åra med å peika på at dei konkrete resultatane avgrensa seg til programvare for stavings- og grammatikk-kontroll. Det manglar såleis ikkje på utfordringar i faget.

### ***KVA SKIL LITTERATURVITSKAP FRÅ SPRÅKVITSKAP?***

---

Kay peikte på at vitskapen til vanleg er ute etter å finna generelle prinsipp.

Litteraturvitskapen derimot studerer det avvikande, leitar etter skilnader.

Lingvistane leitar etter generelle trekk ved språket. Dei søker kjernen, det robuste i språket som ikkje så lett vert endra.

Litteraturforskarane samanliknar to eller fleire forfattarar, verk, genrar eller periodar, medan lingvistane samanliknar det som finst med det som kunne ha funnest. "Vanlege" lingvistar studerer trekk som har med språkproduksjon å gjera, medan datalingvistane er meir innretta mot å analysere språket og mottaksprosessen, altså språkforståinga.

### ***KVA HØYRER MED I EIT SPRÅK?***

---

Kay spurde korleis folk veit kva som høyrer med i språket deira. Denne avgrensinga er svært problematisk å få formalisert, slik at det kan nyttast i datamaskinprogram. Legg merke til at programma for grammatikk-kontroll ikkje kontrollerer om grammatikken er korrekt, men berre peikar på meir eller mindre opplagde feil.

### ***Å INFORMERA FOLK ER EIN KOMPLISERT PROSESS.***

---

Når me informerer folk, gjev me dei bitar til ein modell som me ønskjar at dei skal byggja i medvitet. For å illustrera kor samansett denne prosessen er, ba Kay tilhøyrarane samanlikna desse to utsegna:

- "Eg er nett komen attende frå Saudi-Arabia og har mest gløymt korleis godt øl smakar!" (Infomasjon: Eg har vore lenge utan øl!)
- "Eg er nett komen attende frå Tjekkosllovakia og har mest gløymt

korleis godt øl smakar!" (Informasjon: Ølet i Tjekkosllovakia er dårleg!)

Med fleire slike døme viste Martin Kay at språket ikkje nokon enkel og stabil kode, men ein system av sjølvmodifiserande kodar. Konklusjonen hans var følgjande: "What we are doing all the time is to use language in a very remarkable way".

# ACH/ALLC '91

## "MAKING CONNECTIONS"

---

---

### COMPUTER SCIENCE AND LINGUISTIC ANALYSIS

---

*Chair: Robert Oakman, South Carolina*

*Kip Canfield, Maryland, Baltimore:*

**OBJECT-ORIENTED DATABASE DESIGN FOR RESEARCH AND EDUCATION IN TEXTUAL STUDIES**

*Keith Devlin, Colby College:*

**THE LOGIC OF INFORMATION** Edward A. Fox, Virginia Polytechnic:  
**CODER: A RETRIEVAL AND HYPERTEXT SYSTEM USING SGML AND A LEXICON**

### ANALYSIS OF LINGUISTIC INTENT

---

*Chair: Sylvia Candelaria de Ram, New Mexico State*

*Terumasa Ehara and Tsuyoshi Morimoto ATR, Kyoto:*

**CONTENT AND STRUCTURE OF THE ATR BILINGUAL DATABASE OF SPOKEN DIALOGUES**

*Kevin Donaghy, Rochester Institute of Technology:*

**IMPERATIVES, SPEECH ACTS AND PRACTICAL ARGUMENTS**

*Harry C. Bunt, ITK - Tilburg:*

**THEORY BUILDING ON THE COMPUTER**

### THE PRACTICING HUMANIST AND THE COMPUTER

---

*Chair: Eric Dahlin, UC Santa Barbara*

*Peter Serdiukov, Kiev Pedagogical Institute of Foreign Languages:*

**CALL IN THE USSR**

*Deborah Wilde, Getty Art History Information:*

**THE ULTIMATE CONNECTION: SCHOLARS AS END-USERS**

### EXPERT SYSTEMS AND THE LEXICON

---

*Chair: Gordon Dixon, Manchester Polytechnic*

*Donalee H. Attardo, Purdue:*

**FRAMEBUILDER: A TOOL FOR COMPUTATIONAL LEXICOGRAPHY**

*Brainimir Boguraev and Mary S. Neff, IBM, Yorktown Heights, NY:*

**TEXT REPRESENTATION, DICTIONARY STRUCTURE, AND LEXICAL KNOWLEDGE**

*Elisabetta Marinai, et.al., CNR, Pisa:*  
**FIRST PROTOTYPE OF A SYSTEM FOR THE SEMI-AUTOMATIC LINKING  
AND MERGING OF MONO- AND BILINGUAL LDBS**

#### **ANALYSIS OF STYLE**

---

*Chair: Joseph Rudman, Carnegie Mellon*  
*Douglas Biber, et.al., Southern California:*  
**MULTI-DIMENSIONAL ANALYSES OF AUTHOR'S STYLE: SOME CASE  
STUDIES FROM THE 18TH CENTURY**

*David Holmes, Bristol Polytechnic:*  
**VOCABULARY RICHNESS AND THE BOOK OF MORMON: A STYLOMETRIC  
ANALYSIS OF MORMON SCRIPTURE**

*Estelle Irizarry, Georgetown:*  
**ONE WRITER, TWO AUTHORS: RESOLVING THE POLEMIC OF LATIN  
AMERICA'S FIRST NOVEL**

#### **EXTRACTING THE SOCIAL CONTEXT OF TEXT**

---

*Chair: Vicki Walsh, UCLA*  
*Marilyn Deegan, Oxford:*  
**CULTURAL THEORY AND COMPUTING: A MODEL FOR THE STUDY OF  
ANGLO-SAXON SOCIETY**

*Rosanne G. Potter, Iowa State:*  
**MODERN BRITISH LITERATURE, READER RESPONSE CRITICISM, AND  
GENDER DIFFERENCES**

#### **LANGUAGE INSTRUCTION**

---

*Chair: Charles Bush, BYU*  
*Lori Levin, et.al., Carnegie Mellon:*  
**CAPTURING CONTEXT IN A COMPUTER-BASED FOREIGN LANGUAGE  
ASSISTANT**

*Gregory Lessard and Michael Levison, Queen's University:*  
**COMPUTER-AIDED ANALYSIS AND MODELLING OF SECOND LANGUAGE  
PERFORMANCE ERRORS**

*Christophe Fouqueré, Paris-Nord:*  
**EVIDENCE FOR PREFERENTIAL ANALYSIS**

#### **PUBLISHERS, INTELLECTUAL PROPERTY RIGHTS, AND THE DEVELOP- MENT OF ELECTRONIC TEXT**

---

*Chair: Michael Neuman, Georgetown*  
*Dennis Karjala, ASU:*  
**ELECTRONIC TEXTS, COPYRIGHT AND THE LAW**

*Mark Rooks, InteLex, Pittsboro, NC:*  
**COPYRIGHT AND THE DEVELOPER OF ELECTRONIC TEXT**

*Darrell Bock, Dallas Theological Seminary:*  
**ELIMINATING CONCERNS AND ASSESSING THE COSTS**



*Eric Calaluca, Chadwyck-Healey, New York, Ny:*  
**MARKETING ELECTRONIC TEXT**

---

### **UNDERSTANDING PLOT**

*Chair: Joel Goldfield, Institute for Academic Technology, Chapel Hill, NC  
Clifford W. Anderson and George E. McMaster, Brandon University:*  
**CONNECTED EMOTION IN SIMPLIFIED TEXT: DID THEY REALLY RUIN  
PETER RABBIT?**

*John Reeves, UCLA:*  
**COMPUTER MODEL OF THEMATIC STORY UNDERSTANDING FROM  
MORAL REASONING**

*Kathryn Sutherland, Oxford:*  
**WAITING FOR CONNECTIONS: HYPERTEXTS, MULTILOTS, AND THE  
ENGAGED READER**

---

### **COMPLEX TEXT RETRIEVAL**

*Chair: Nancy Frishberg, IBM, Yorktown Heights, NY  
Tom Horton, Florida Atlantic:*  
**TEXT RETRIEVAL OF PASSAGES BASED ON WORD CO-OCCURRENCES**

*Takahiro Nakamura and Satoshi Aisaka, Computer Applications, Tokyo:*  
**KEYWORD EXTRACTION IN JAPANESE BY FUZZY INFERENCE**

*Susan Siegfried, Getty Art History Information:*  
**MATCHING PERSONAL NAMES IN THE HUMANITIES**

---

### **TOOLS FOR LITERARY ANALYSIS**

*Chair: Paul Fortier, Manitoba  
Peter Robinson, Oxford:*  
**A NEW PROGRAM FOR INTERACTIVE COLLATION OF LARGE MANU-  
SCRIPT TRADITIONS**

*Heyward Ehrlich, Rutgers, and George Vallasi, Chernow Editorial, New York, NY:*  
**THE JAMES JOYCE TEXT MACHINE**

*Michael Hilton, South Carolina:*  
**THE URICA! II INTERACTIVE COLLATION SYSTEM**

---

### **PLENARY SESSION #2**

*Chair: David Chisholm, Arizona  
Ralph Griswold, Arizona:*  
**THE PROGRAMMING LANGUAGE CONNECTION**

---

### **STATUS REPORT ON THE TEI**

*Chair: Susan Hockey, Oxford  
Lou Burnard, Oxford and Michael Sperberg-McQueen, Illinois at Chicago*

## **ARTIFICIAL INTELLIGENCE APPROACHES TO THE STUDY OF METAPHOR**

---

*Chair: Mary Dee Harris, Language Technology, Washington, DC*

*Bipin Indurkha, Boston U:*

**METAPHOR AS CHANGE OF REPRESENTATION**

*James Martin, Colorado:*

**A COMPUTER UNDERSTANDING OF CONVENTIONAL METAPHORIC LANGUAGE**

*Dan Fass, Simon Fraser:*

**A COMPUTER METHOD FOR RECOGNIZING METAPHORS IN SENTENCES**

## **METRICS**

---

*Chair: Roy Flannagan, Ohio University*

*Malcolm Hayward, Indiana University of Pennsylvania:*

**A CONNECTIONIST COMPUTER MODEL OF POETIC METER**

*David G. Novick and Thomas A. Doehne, Oregon Graduate Institute:*

**AUTOMATED POETRY CLASSIFICATION**

*David Chisholm and Royce Robbins, Arizona:*

**FRAME: A COMPUTER PROGRAM FOR THE STUDY OF PHONOLOGICAL EQUIVALENCE IN LITERARY LANGUAGE**

## **VARIATION IN SPEECH**

---

*Chair: Ian Lancashire, Toronto*

*Junko Hosaka, et.al., ATR, Kyoto:*

**CLASSIFICATION OF S-POSTPOSITIONS IN SPOKEN LANGUAGE: TOWARD SPEECH RECOGNITION**

*John M. Kirk and William A. Kretzschmar, Georgia:*

**THE ANALYSIS AND INTERPRETATION OF DIALECT DATABASES BY INTERACTIVE MAPPING**

*Martin Dickey and Leonard Faltz, ASU:*

**STATISTICAL ANALYSIS OF PHONETIC DIALECT AND GENDER DIFFERENCES IN A CD-ROM SPEECH CORPUS**

## **SYNTACTIC ANALYSIS**

---

*Chair: Wendy Wilkins, ASU*

*Rodolfo Delmonte, Venice, and Dario Bianchi, Parma:*

**COMPUTING DISCOURSE ANAPHORA FROM GRAMMATICAL REPRESENTATIONS**

*Palmira Marrafa, Lisbon:*

**ON SECONDARY PREDICATION IN PORTUGUESE: CONSTITUENCY AND COMPUTABILITY**

*Xiaojin Yu and Robert Oakman, South Carolina:*

**CAPTURING SYNTACTIC DEPENDENCIES IN BUILDING LOGICAL SENTENCE FORMS FOR MACHINE TRANSLATION**

## **PERSPECTIVES ON THE TEI**

---

*Chair: Elaine Brennan, Brown*

*Panelists: Elaine Brennan, Brown, Steve Siebert, Dragonfly Software, New York, NY, Sperling Martin, Consultant, Gaithersberg, MD*

## **PSYCHOLOGICAL RESPONSES TO COMPUTER-MEDIATED INSTRUCTION**

---

*Chair: Marianne Gaunt, Rutgers*

*Jeutonne Brewer, UNC Greensboro, and Boyd H. Davis, UNC Charlotte:*

**REPETITION AND POLITENESS IN ELECTRONIC CLASSROOM COMMUNITIES**

*Karen Smith and Barbare Hoffman Maginnis, Arizona:*

**COMPUTER-MEDIATED COMMUNICATION ENVIRONMENTS FOR EDUCATION: RESEARCH AND METHODOLOGY**

*Kathleen Skubikowski and John Elder, Middlebury:*

**COMPUTERS AND THE SOCIAL CONTEXTS OF WRITING**

## **MACHINE TRANSLATION: EAST - WEST**

---

*Chair: Maureen Schmid, Tandem Translation Project, Tempe, AZ*

*Hsin-Hsi Chen, National University of Taiwan:*

**THE TRANSFER OF PREPOSITIONAL PHRASES IN AN ENGLISH-CHINESE MACHINE TRANSLATION SYSTEM**

*One-Soon Her, et.al, Executive Communication Systems, Provo, UT:*

**THE TREATMENT OF ENGLISH IDIOMS IN THE LFG-BASED ECS MACHINE TRANSLATION SYSTEM**

## **ARCHIVES AND DOCUMENT RETRIEVAL**

---

*Chair: Randall Jones, BYU*

*James E. Tierney, Missouri-St. Louis:*

**A CD-ROM SUBJECT INDEX TO PRE-1800 BRITISH PERIODICALS**

*Patricia Galloway, Mississippi Department of Archives:*

**HISTORICAL MAPS AND RUBBER SHEETS: FINDING NATIVE AMERICAN HISTORY ON THE GROUND**

## **APPRECIATING MONUMENTS OF FRENCH LITERATURE**

---

*Chair: Glyn Holmes, University of Western Ontario*

*Paul Fortier and Carl J. Schwarz, Manitoba:*

**PROCRASTEAN ANALYSIS OF THEMATIC STRUCTURES IN ANDRE GIBE'S L'IMMORALISTE**

*Gregory Lessard and Jean-Jacques Hamm, Queen's University:*

**COMPUTER-AIDED ANALYSIS OF REPEATED STRUCTURES: THE CASE OF STENDHAL'S ARMANCE**

## **STATISTICAL METHODS IN COMPUTATIONAL RESEARCH**

---

*Chair: Ken Church, AT&T Bell Laboratories, Murray Hill, NJ*

*Mark Liberman, U of Pennsylvania:*  
**HOW MANY WORDS DO PEOPLE KNOW?**

*Martin Kay, Xerox PARC, Palo Alto, CA:*  
**TEXT-TRANSLATION ALIGNMENT**

*Bob Mercer, IBM, Yorktown Heights, NY:*  
**A STATISTICAL APPROACH TO MACHINE TRANSLATION**

*Ken Church, AT&T Bell Laboratories, Murray Hill, NJ:*  
**USING STATISTICS IN LEXICAL ANALYSIS**

---

#### **DEFINING A NATIONAL LITERARY CORPUS**

*Chair: Allen Renear, Brown*

*Keiko Kitamura and Hisashi Yasunaga, National Institute of Japanese Literature, Tokyo:*

**DATA BASE DELIVERY FOR JAPANESE LITERATURE BY CD-ROM**

*Andrea de Leeuw van Weenen Leiden:*

**DEVELOPMENTS AT THE LEIDEN ARMENIAN DATA BASE**

*Sune Magnberg, Stockholm:*

**CORPUS BASED RESEARCH ON MODELS FOR PROCESSING UNRESTRICTED SWEDISH TEXT**

---

#### **A SPECIAL SESSION**

*Concordanze della Lingua Italiana Poetica dell'OttolNovecento (CLIPON), Italian National Council for Research (CNR)*

*Chair: Luciano Farina, Ohio State*

*Giuseppe Savoca, Catania:*

**WORDS BY UNGARETTI AND BY MONTALE**

*Alida D'Aquino:*

**TOWARDS A DICTIONARY OF D'ANNUNZIO'S POETRY**

*Sebastiano Catrona:*

**COMPUTING SOLUTIONS FOR DEALING WITH ITALIAN LITERARY TEXTS**

*Amalia Mannino:*

**ABOUT SOME FORMAL AND SEMANTIC FIELDS IN LEOPARDI'S CANTI**

---

#### **AUTOMATING CONTENT ANALYSIS**

*Chair: David Bantz, Chicago*

*Nancy M. Ide, et.al., Centre National de la Recherche Scientifique (CNRS):*

**CAUGHT IN THE WEB OF WORDS: USING NETWORKS GENERATED FROM DICTIONARIES FOR CONTENT ANALYSIS**

*Burghard Rieger, Trier:*

**AUTOMATIC WORD MEANING REPRESENTATION AND CONTENT-DRIVEN INFERENCE**

*David Hunter, Texas at Arlington, and Rita D'Arcangelis, Mary Washington:*

**A PARTITIONED RULE-BASED APPROACH TO CONTENT ANALYSIS OF SPOKEN LANGUAGE**

## **TEXTS, CONCORDANCES AND TEXTUAL INFORMATION**

---

**Chair: David Owen, Arizona**

*Syun Tutiya, Chiba:*

**ARCHAEOLOGY OF TEXTUAL INFORMATION**

*Alastair McKinnon, McGill:*

**THE MULTIDIMENSIONAL CONCORDANCE: A NEW TOOL FOR LITERARY RESEARCH**

## **MANAGING LEXICAL KNOWLEDGE BASES**

---

**Chair: Leonard Faltz, ASU**

*Patrizia Cotoneschi, Biblioteca di Ingegneria Florence, and Monica Monachini, CNR, Pisa:*

**AN EMPIRICAL EXPERIENCE ON THE UTILIZATION OF THE ITALIAN REFERENCE CORPUS IN MEANING ANALYSIS**

*Sumali Pin-Ngern Conlon and Martha Evens, Mississippi:*

**BUILDING A NOUN LEXICAL DATABASE TO SUPPORT NATURAL LANGUAGE PROCESSING APPLICATIONS**

*Patrick Saint-Dizier, U Paul Sabatier, Toulouse:*

**A DENOTATIONAL SEMANTICS FOR SYNONYMS AND OPPOSITES**

## **EDITING AND TEXT MARKUP**

---

**Chair: George Logan, Queen's University**

*David R. Chesnutt, South Carolina:*

**THE PAPERS OF HENRY LAURENS: A TEST CASE FOR THE TEI GUIDLINES**

*Hans van Halteren, Nijmegen:*

**EFFICIENT STORAGE OF AMBIGUOUS STRUCTURES IN TEXTUAL DATABASES**

*Michael Sperberg-McQueen, Illinois at Chicago:*

**THE VALIDATED-OR VIOLATED?-TEXT: ISSUES IN SPECIFYING DOCUMENT STRUCTURES**

## **PLENARY SESSION #3**

---

**Chair: Dan Brink, ASU**

*Helen Agüera, National Endowment for the Humanities, Washington, DC:*

**A CONNECTION BETWEEN PAST AND FUTURE: NEH FUNDING OF HUMANITIES COMPUTING PROJECTS**

## **LEXICAL, TEXTUAL, AND SOFTWARE RESOURCES**

---

**Chair: Don Walker, Bellcore, Morristown, NJ**

*Mark Liberman, U of Pennsylvania:*

**THE NEED FOR OPEN LEXICAL AND TEXTUAL RESOURCES**

*Nicoletta Calzolari, CNR, Pisa:*

**THE REUSABILITY OF LEXICAL AND TEXTUAL RESOURCES**

*Yorick Wilks, New Mexico State:*

**THE ROLE OF A CONSORTIUM FOR LEXICAL RESEARCH**

*Antonio Zampolli, CNR, Pisa:*

**DETERMINING WHAT LANGUAGE DATA ARE AVAILABLE IN MACHINE-READABLE FORM**

*Elizabeth Hinkelman, Chicago:*

**ACCESSIBILITY AND QUALITY OF SOFTWARE RESOURCES**

**BUILDING INTELLIGENT ENVIRONMENTS FOR HUMANISTS**

---

*Chair: Peter Lafford, ASU*

*Donalee H. Attardo, et.al., Purdue:*

**THE SMEARR SEMANTIC DATABASE: AN INTELLIGENT AND VERSATILE RESOURCE FOR THE HUMANITIES**

*Walter Daelemans, Tilburg, Nederland:*

**AN OBJECT-ORIENTED LINGUISTIC TOOLBOX**

**TOOLS FOR LINGUISTIC ANALYSIS**

---

*Chair: Kevin Roddy, UC Davis*

*Gregory Crane, Harvard:*

**GENERATING AND PARSING CLASSICAL GREEK**

*Vyacheslav Ivanov, Moscow:*

**COMPUTER-ASSISTED ANALYSIS OF A BILINGUAL, HATTIC-HITTITE TEXT**

**CREATING NATURAL LANGUAGE CORPORA**

---

*Chair: Nancy Ide, Vassar*

*Jeremy Clear, Oxford UP:*

**CRITERIA FOR CONSTRUCTING CORPORA**

*Beatrice Santorini, Pennsylvania:*

**BUILDING VERY LARGE NATURAL LANGUAGE DATABASES: SOME METHODOLOGICAL CONSIDERATIONS**

*Nicoletta Calzolari, CNR, Pisa:*

**KNOWLEDGE EXTRACTION FROM CORPORA**

**INSTRUCTION IN STYLE**

---

*Chair: Don Nilsen, ASU*

*Donald Ross, Minnesota, and David Hunter, Texas at Arlington:*

**M-EYEBALL: AN INTELLIGENT SYSTEM FOR STYLISTIC DESCRIPTIONS AND COMPARISONS**

*Bettina Harriehausen-Mühlbauer, IBM, Tyskland:*

**THE COMPUTER AS "TEACHER" FOR GRAMMAR AND STYLE ERRORS**

*Julie Payette and Graeme Hirst, Toronto:*

**COMPUTER-ASSISTED INSTRUCTION IN SYNTACTIC STYLE**

# NYTT NUMMER AV OLUFF

---

---

OLUFF er en interessegruppe organisert av NAVFs edb-senter for **humanistisk forskning** ved Universitetet i Bergen. Interessegruppens fulle navn er *Interessegruppe for optisk lagringsteknologi i undervisning, forskning og formidling*. Siden opprettelsen i 1988 har interessefeltet blitt utvidet til å omfatte multimedia, fjernundervisning, OCR, m.m.

Medlemmer av OLUFF får ved innmelding tilsendt en utskrift av medlemsregisteret, og får på den måten en oversikt over de andre medlemmene, med en kort beskrivelse av deres interesser, prosjekter og produkter. Tre ganger i året får medlemmene tilsendt et meldingsblad. Medlemmene kan også bruke interessegruppen til selv å informere andre medlemmer, enten ved en presentasjon i meldingsbladet, eller via medlemsregisteret.

OLUFFs primære målgruppe er universitets- og høyskolemiljøene i Norge og øvrige humanistiske fag- og forskningsinstitusjoner, som arkiv, museer og bibliotek. Derfor vil OLUFF legge særlig vekt på anvendelser innenfor kulturhistorisk dokumentasjon, språkfag og undervisning. Men medlemskap er åpent for alle, uavhengig av fag- og institusjonstilknytning, og i dag har gruppen et stort antall medlemmer utenfor humaniora. Derfor dekkes også andre anvendelsesområder.

Avhengig av om man er ansatt i en kommersiell bedrift eller om man er en person med tilknytning til universitet eller annen offentlig virksomhet, koster et årsmedlemskap enten kr 400 eller kr 125. Ønsker du også å bli medlem, ring, fax, skriv eller send e-post til redaksjonen, så får du tilsendt innmeldingsskjema.

Redaktørene av OLUFF og Humanistiske Data samarbeider, og som OLUFF-medlem er du velkommen med bidrag til begge publikasjoner. Stoff av rent teknisk karakter passer best i OLUFF.

Redaksjonens adresse er:

## OLUFF

NAVFs edb-senter for humanistisk forskning

Postboks 53 Universitetet

N-5027 Bergen

Tlf.: + 47 5 212954/55/56. Fax: + 47 5 322656

Nettadr.: FAFOR@NOBERGEN.EARN (bruker: Øystein Reigem)

Portacom (UiO): Espen S. Ore



|  |    |
|--|----|
| Nettverksdistribusjon av OLUFF-stoff . . . . .   | 5  |
| Vertsmaskiner . . . . .  | 5  |
| Distribusjonsliste HUMEDB . . . . .  | 6  |
| Filtjener på NORA . . . . .  | 6  |
| FotoMac videreutvikles, <i>Espen S. Ore</i> . . . . .  | 7  |
| «Databas Norden» – nordiske kommunedata på CD-ROM,<br><i>Jostein Ryssevik</i> . . . . .                  | 10 |
| Rogaland Mediesenter, TV-VEST, Elektronisk Universitet Norge<br>og Ullandhaug Opplæringscenter . . . . . | 13 |
| Apple Multimedia Ullandhaug, <i>Dan Dyrli Daatland</i> . . . . .   | 15 |
| Fjernundervisningsprosjekt om lese- og skrivevansker,<br><i>Dan Dyrli Daatland</i> . . . . .             | 18 |
| Multimediaprojekt for videregående skole,<br><i>Jostein Saakvitne og Øystein Reigem</i> . . . . .        | 24 |
| Medielaboratoriet ved Nasjonalbibliotekavdelinga i Rana, <i>Øystein Reigem</i> . . . . .                 | 27 |
| Multimedialaboratoriet ved AVH, Universitetet i Trondheim, <i>Julie Feilberg</i> . . . . .               | 29 |
| Kompresjonsstudier i Grieg-prosjektet, <i>Øystein Reigem</i> . . . . .                                   | 31 |
| Optisk lagring av dokumenter i offentlige institusjoner<br>i USA, <i>Øystein Reigem</i> . . . . .        | 33 |
| Moderne folkesang med LaserKaraoke, <i>Øystein Reigem</i> . . . . .                                      | 34 |
| Firma- og produktpresentasjon . . . . .  | 36 |
| CD-ROM, Norge AS, <i>Trond Hammernes</i> . . . . .   | 36 |
| Multimedja – til informasjon, nyhetsformidling og opplæring,<br><i>Jan Roar As</i> . . . . .             | 37 |
| CD-WORD – flerspråklig ordbok på CD-ROM . . . . .  | 42 |
| Grensesnitt . . . . .  | 45 |
| EyesCream – verktøy for bygging av brukergrensesnitt,<br><i>Øystein Reigem</i> . . . . .                 | 45 |
| Input er ikke bare tastatur og mus . . . . .   | 48 |
| Framtidige brukergrensesnitt . . . . .   | 48 |
| Video- og lydkort, <i>Øystein Reigem</i> . . . . .   | 49 |
| DVA-4000-kortet – en multimediaplattform for tre<br>maskintyper, <i>Skule Sarmo</i> . . . . .            | 50 |
| Computer Eyes fra Digital Vision – rimelig digitaliseringskort . . . . .                                 | 51 |
| Videokort fra SuperMac . . . . .   | 51 |
| Brikke kan gjøre avanserte videokort rimeligere . . . . .  | 52 |
| Tre kort fra New Media Graphics: video/lyd, kompresjon,<br>høy oppløsning . . . . .                      | 52 |
| RadiusTV – videokort for Mac . . . . .   | 53 |
| TARGA og andre video- og grafikkort fra Truevision . . . . .   | 53 |
| Stillbildekompresjon med C-Cube . . . . .  | 54 |
| Videoredigering på Amiga med Video Toaster . . . . .   | 54 |
| Digital audio . . . . .  | 54 |

|   |           |
|---|-----------|
| Oversikt over videokort . . . . .   | 55        |
| Videoproduksjon med Video F/X . . . . .   | 55        |
| Multimedia + AI + transputer = Picture Book Professional . . . . .  | 56        |
| Lyd til Animator med Gloria . . . . .   | 56        |
| Videodigitalisering for PC og P5/2 med SuperVIA . . . . .   | 56        |
| Vanlig video ut fra Mac med Neotech Color Video Encoder . . . . .   | 57        |
| <b>CD-nyheter, Øystein Reigem . . . . .</b>   | <b>57</b> |
| CD-ROM-artikler/tester . . . . .  | 57        |
| Pioneer DRM-610 CD-ROM-skifter . . . . .  | 57        |
| United States History på CD-ROM . . . . .   | 58        |
| Ny versjon av Microsoft CD-ROM Extensions for DOS . . . . .   | 59        |
| Microsofts multimediaspesifikasjoner . . . . .  | 59        |
| CD-RDx . . . . .  | 59        |
| Hitachi CD-ROM-spiller raskest? . . . . .   | 59        |
| Skrivbar CD-ROM . . . . .   | 59        |
| Kort fra CD-ROM XA-utvikling fra Sony . . . . .   | 60        |
| CD-ROM på spillmarkedet . . . . .   | 60        |
| Tidsskrifter på CD-ROM . . . . .  | 61        |
| Utvikling av CD-I undervisningsprogrammer . . . . .   | 61        |
| <b>Nyheter lagring . . . . .</b>  | <b>61</b> |
| Overskrivbare digitale (magneto-optiske) plater på<br>markedet i Norge – ytterligere opplysninger . . . . . | 61        |
| Magneto-optiske selger bra . . . . .  | 62        |
| Store magnetiske harddisker på Mac . . . . .  | 62        |
| Flerfunksjonsstasjoner igjen . . . . .  | 62        |
| IBM 3 1/2" magneto-optisk . . . . .   | 62        |
| Floptical . . . . .   | 63        |
| Holografisk minne . . . . .   | 63        |
| <b>Ymse . . . . .</b>   | <b>64</b> |
| Farger . . . . .  | 64        |
| Fraktalkompresjon, Øystein Reigem . . . . .   | 65        |
| Tester av OCR-systemer, Øystein Reigem . . . . .  | 65        |
| Storyboard Live! – multimediaprogramvare fra IBM . . . . .  | 66        |
| Ask LCD . . . . .   | 66        |
| DynaVision (eks Dynabook) . . . . .   | 66        |
| 4B 6F 72 74 74 79 70 65, Øystein Reigem . . . . .   | 67        |
| «Bølgetter», Øystein Reigem . . . . .   | 67        |
| <b>Tidsskrifter, bøker, rapporter . . . . .</b>   | <b>69</b> |
| Erling Maartmann-Moe: «Multimedia» – ny bok . . . . .   | 69        |
| Hypermedia – nytt tidsskrift . . . . .  | 70        |
| Image Systems – nytt tidsskrift . . . . .   | 72        |
| Interactive Media International – nytt tidsskrift . . . . .   | 73        |
| Ny CD-ROM undersøkelse . . . . .  | 74        |
| <b>Konferanser . . . . .</b>  | <b>75</b> |

# CHArt – FORENING OG TIDSSKRIFT FOR KUNSTHISTORIE

---

---

*Øystein Reigem*

*Computers and the History of Art* (CHArt) er en internasjonal forening som ble stiftet i 1985, i stor grad med medlemmer fra Storbritannia. Formålet var å koordinere og utvikle bruken av datasystemer for innlegging av kunsthistorisk informasjon og forskning innen kunsthistorie. CHArt avholder en årlig konferanse, og fram t.o.m. 1989 utga foreningen et meldingsblad (the CHArt Newsletter). F.o.m. 1990 er dette avløst av et ordinært tidsskrift (the CHArt Journal).

Vi skulle gjerne ha anmeldt dette bladet her og nå, men på grunn av en misforståelse et eller annet sted i systemet, har ikke vårt abonnement trådt i kraft. Vi har fått tak i det aller første nummeret av tidsskriftet, nr 1-90, og dette er i stor grad viet 1989-konferansen, som ble holdt i London 7–8.12.89. Fem av de åtte foredragene på konferansen er kommet med i bladet. *Øystein Reigem* refererte fra konferansen i HD 2-90. Der dekket han i tillegg til noen av foredragene en meget interessant demonstrasjon fra VASARI-prosjektet (Visual Arts System for Archiving and Retrieval of Images) – et EEC-prosjekt med lagring av høyoppløselige bilder (malerier). Denne demonstrasjonen/prosjektet er ikke beskrevet i bladet, men CHArt lover en full rapport fra prosjektet i et senere nummer.

Her er innholdsfortegnelsen for bladet:

Busch, Joseph A.: *Applied research sponsored by the Getty Art History Information Program*

Tanya Szrajber, *Oriental and Japanese Records, CDMS, British Museum: CDMS: the computerisation of the British Museum collection*

Leung, C.H.C., Page, S., og Mannock, K.L.: *Knowledge-based pictorial database management*

Saunders, David: *The role of computers in environmental data-logging*

Gastkemper, F.H.D.: *Computer-aided learning and interactive video in distance education*

Miles, Jean: *Making an index with WordPerfect*

Good, Michael: *Computerizing The Buildings of England: Conflict or Continuity?*

Book review

*Computers and the History of Art*, edited by Anthony Hamber, Jean Miles and William Vaughan. Reviewed by David Lee

De som ønsker mer informasjon, kan kontakte CHArt for et abonnement, evt medlemskap (som gir gunstigere abonnement og deltakelse på konferansen). For Europa er kontakten for abonnement: *P.O Box 90, Reading, Berkshire, U.K RG1 8JL. Tlf: + 44 (0)73 456 0080. Fax: + 44 (0)73 456 8211.*

USA, Japan, Asia og Australia har egne kontaktadresser.

Bladet sier ingenting om hvor medlemskap bestilles fra. Da konferansen ble avholdt, var adressen *Computers and the History of Art (CHArt), 43 Gordon Square, London WC1.*

# DATASTØTTA FORSKING I GERMANSK SPRÅK OG SKANDINAVISK MIDDELALDERSPRÅK

---

---

*Per Vestbøstad*

**"INTERNATIONAL ANNUAL NEWSLETTER FOR  
GERMAN AND MEDIEVAL SCANDINAVIAN COMPUTER  
RESEARCH"**

---

*International Annual Newsletter for German and Medieval Scandinavian  
Computer Research* (nr. 16, hausten 1990) gjev mellom anna informasjon  
om:

***Komplett datakatalog for middelaldermanuskript i  
matematiske emner:***

---

*Menso Folkerts* og *Andreas Kühne* ved Inst. für Geschichte der Natur-  
wissenschaften der Universität München (Postfach 260102, D-8000 MÜN-  
CHEN 26) har laga utvalde opplysningar om alle middelaldermanuskript  
som gjeld aritmetikk, geometri, algebra, trigonometri og beslekta emner  
i ein database ved hjelp av *Nota Bene v. 3.0* (akademisk tekstbehandlar).  
Databasen dekkjer tida mellom 500 og 1500 e. Kr.

***Prosjekt innanfor moderne germansk filologi:***

---

- \* Bibliografi over tysk litteratur som er omsett til engelsk i det 20. hundreåret – med kryssreferansar.
- \* *PRISEK* – eit maskinlagra leksikon til bruk ved kvantitative analysar av stil og innhald i litterære verk av *Goethe, Novalis, Rilke* og *Bachmann*.
- \* Datastøtta undersøking av tale og skriftspråk i Austerrrike.

- \* Ein konkordans over *Freuds* "Gesammelte Werke".
- \* Ein konkordans over *Kafkas* verk og dagbøker(?).
- \* Ein konkordans over tyske salmar frå det 16. og 17. hundreåret.
- \* Eit automatisk lemmatiseringsprogram for *Schillers* poesi.

### ***Datastøtta opplæring i tysk***

---

- \* *GERDA* – eit program for interaktiv språklæring på Prime- og Macintosh-maskiner.
- \* *CALIS* – eit liknande program for MS-dos-maskiner.
- \* Grammatikk- og vokabularøvingar både for Macintosh- og MS-dos-maskiner.
- \* Planar for "a multi-volume analytic dictionary of frequency and range of contemporary spoken German".
- \* Ein hypermedia databank over tysk litteratur berekna både for undervisning og forskning. Skal innehalda bibliografiar og vert truleg publisert på CD-ROM.
- \* Forfattarspråk for språkøvingar for studentar (CAPE og TenCORE for MS-dos-maskiner).

### ***Skandinavisk middelalderspråk***

---

Her er det berre ein notis, nemleg om *Concise Dictionary of North European Mythology and Antiquities*: 2.500 namn frå nord-europeisk mytologi definerte og kryssrefererte.

### ***Maskinlagra tekstar, elektroniske diskusjonsgrupper, newsletter, m.m.***

---

Eit grundig oversyn over institusjonar som lagrar og distribuerar maskinlagra tekstar, elektroniske diskusjonsgrupper, newsletter, m.m. Oversynet ser ut til å dekkja Europa, Israel og Nord-Amerika.

### ***Samarbeid med "Germanic Notes and Reviews"***

---

Frå og med 1992 vil *International Annual Newsletter for German and Medieval Scandinavian Computer Research* verta publisert i tidsskriftet *Germanic Notes and Reviews* (GNR), som Univ. of Minnesota og Bemidji State University gjev ut to gonger årleg. GNR kan tingast ved å senda US \$ 9.- (\$ 11.- fra 1992) til: *Professor Dr. Evelyn Scherabon Firchow, Director, Computer Clearing House Project in German and Medieval Scandinavian, Folwell Hall, University of Minnesota, Minneapolis, Minnesota, 55455 USA.*

***GNR NR 3/4 1990 INNEHELD FØLGJANDE ARTIKLAR:***

---

Doris T. Wight: Masochism, Mourning, Melancholia. A Freudian interpretation of E. T. A. Hoffman's Tale, "The Mines of Falun".

Dieter Saalman: The Experience of Exile: Prolegomena For a Deconstructive Analysis of the Exilic Domain Within the Purview of Contemporary German-German Literature.

Anne L. Critchfield: Forum Stadtpark Graz: 30 Years of Tradition.

Ahuva Belkin: "Ugliness without Pain": The Peasant as the Theatrical Fool in German Graphic Art of the Renaissance.

Book Reviews.



# ARTIKLER/RAPPORTER I HUMANISTISKE DATA 1990

---

---

## SPRÅK OG LITTERATUR

---

### 1-90

Bjørn Sørensen: Instruction or Learning: Challenges for Interactive Video.

Signe Marie Sanne: Interaktiv video fra amerikanske miljøer.

Ivar Utne: Datalingvistik og datastøtta leksikografi/terminologi.

Knut Vikør: Seminar on Bilingual Computing in Arabic and English.

Signe Marie Sanne: Interactive Media '89.

Signe Marie Sanne: Interaktiv video og CD-ROM i Italia.

### 2-90

Erik Meistrup: Fremmedsprogsundervisning og de elektroniske medier.

Unni Hovstad: Da Miss Brill og Uncle Ernest avla eksamen i engelsk linjefag.

Espen Ore: Klassisk litteratur på CD-ROM.

Kristin Natvig og Signe Marie Sanne: Seminar om interaktiv video i språkopplæring.

Signe Marie Sanne: CALICO '90.

### 3-90

Signe Marie Sanne: Interaktiv video i italiensk språkopplæring.

Olle Eriksen: Glosebok for døve.

Unni Hovstad: Sans Passeport.

Lou Burnard: The Text Encoding Initiative: A Progress Report.

Dagfinn Worren: Leksikografiske databasar.

Espen S. Ore og Kristin Natvig: Årets ALLC/ACH konferanse.

Knut Hofland: The 11th ICAME Conference.

## **MUSEUMSFAG**

---

### **1-90**

Tine Wanning: Den første interaktive videoplade på Nationalmuseet i København.

Jostein H. Hauge: Museene som informasjons- og kunnskapskilde.

## **FILOSOFI**

---

### **1-90**

Solveig Bøe: Intensjonalitet i mennesker og maskiner.

### **2-90**

Kristin Natvig: Wittgenstein lever i Bergen.

## **HISTORIE**

---

### **1-90**

Alf Christophersen: DIS Norge – Forening for databehandling i slektsforskning.

### **2-90**

Sheila Anderson: An Historical Data Archive for the UK: A Feasibility Study.

Øystein Reigem: CHArt 1989 Conference.

### **3-90**

Dan Dyrli Daatland: Nye måter å formidle okkupasjonshistorien på.

## **PEDAGOGIKK**

---

### **1-90**

Mette Strømnes: Datamaskiner – Et spennende og nyttig hjelpemiddel i pedagogisk-psykologisk arbeid.

Erik Olsen: Bedre muligheter for vurdering av pedagogisk programvare i skolen.

Erik Olsen: Utprøving av pedagogisk programvare.

### **2-90**

Erik Meistrup: Undervisningskulturelt bredbåndsnettverk til voksenundervisning.

## **MUSIKK**

---

### **1-90**

Kristin Natvig: Nasjonalt musikkteknologisk senter.

## **BIBLIOTEKSTJENESTER**

---

### **2-90**

Hans Martin Fagerli: Med de nasjonale edb-tjenestene inn i 90-åra.

### **3-90**

Karl Erik Andersen: Medielaboratoriet ved Nasjonalbibliotekavdelinga i Rana (NBR).

## **HUMANISTISK INFORMATIKK**

---

### **1-90**

Asbjørn Brændeland: Et forsvar for programmeringsundervisning i edb for humanister.

Jan Oldervoll: Semesteremne + Pascal = Tøv?

Kristin Natvig: Grunnfag i humanistisk informasjonsbehandling.

Susan Kruse: Humanities Computing at King's College London.

### **2-90**

Espen J. Aarseth: Informatica Rhetorica.

### **3-90**

Jostein H. Hauge: Scholarship and Technology in the Humanities.

Gunnar Liestøl: Mot en digitalisert verden.

## **MULTIMEDIA**

---

### **1-90**

Espen S. Ore: Fotomac.

Kristin Natvig og Signe Marie Sanne: Konferanse om optiske medier og multimedia.

Dan Dyrli Daatland: Konferanse om bruk av CD-ROM.

### **3-90**

Espen S. Ore: Programmering av interaktiv video.

## **PROGRAMVARE**

---

### **1-90**

Asbjørn Hoem: Dr. Strong Abroad.

### **2-90**

Olav G. Karlsen: MEMORY, CLOZE, GRAMENG, ENGELSK, BAS-GRAM.

Signe Marie Sanne: MacNorsk – Norwegian Education Stack.

Oddvar Johan Jensen: AskSam – Et nyttig verktøy.

Kirsti Rye Ramberg: Nedlastbar skrift, skrifteitorer og skrifteffekter.

### **3-90**

Svein Arne Rasmussen: En kort orientering om Winix.

## **DIVERSE**

---

### **1-90**

Karin Nordenstam: 7. Nordiske IoD.

### **2-90**

Axel Baudouin: Edb-innføring, mennesker og makt: Lever vi fortsatt under føydalitet?

Vibeke Vestby: National Computer Graphics Association's årlige konferanse.

### **3-90**

Ivar Gubberud: Databaser: Nødvendige for videre utvikling.

# NYTT I BIBLIOTEKET

---

---

- Brisis, Kararzyna de & Manders, Ingunn 1989, *A retrieval language for Nordic databases: language requirements for the IANI facility*, Bibliotekcentralens forlag, Ballerup.
- Germundson, Lars & Olsson, Bjørn 1990, *Bilden i tryck*, Spektra, Halmstad.
- Himes, Andrew et.al. 1990, *Inside SuperCard: the complete guide for Macintosh developers and advanced users. Including version 1.5*, Microsoft Press, Redmond, Washington.
- Karel Babcicky et.al. 1989, *Sort-hvitt bildebehandling med standard datateknologi: en utredning angående muligheter og begrensninger*, Grafisk Institutt, Oslo.
- Kiil, Alf 1987, *Arkivkunnskap: Statsarkiva*, Universitetsforlaget, Oslo.
- Lunde, Øivind et.al. 1990, *Arkivutvalgets rapport*, Riksantikvaren, Oslo.
- Miall, David s. 1990, *Humanities and the computer*, Clarendon Press, Oxford.
- Norton, Peter 1990, *Inside the IBM PC and PS/2. 3rd ed. (Rev.ed. of: Inside the IBM PC. Rev. and enl. 1986)*, Bradly, New York.
- Quarterman, John S. 1990, *The Matrix: computer networks and conferencing systems worldwide*, Digital Press, Bedfors, Massachusetts.
- Smith, Wm Flint 1989, *Modern technology in foreign language education: applications and projects*, Modern Textbook Co., Lincolnwood, Illinois.
- UNINETT: *hva og hvorfor 1990*, UNINETT, Trondheim.
- University of London Audio-Visual Centre 1989, *The IMAGE 89 video. IMAGE: the international meeting on museums & art galleries image databases London, 18-20 May 1989*, London.
- White, Jan 1990, *Color for the electronic age*, Watson-Guptill, New York.

# ÅRSRAPPORTER FRA SIF OG HUM-DOK

---



Sentral for informasjon om forskningsprosjekter (SIF) og Humanistisk forskningsdokumentasjonstjeneste (HUM-DOK), presenterer seg med årsrapporter for 1990 i et hendig format. Dokumentasjonstjenestene for pågående forskning synes nå for alvor å ha funnet et marked: Bruken av databasen har økt med nesten 100% fra 1989 til 1990.

Dokumentasjonstjenestene har hittil fått sine opplysninger fra saksbehandling i forskningsrådene. I løpet av kort tid regner en også med at data fra forskning finansiert ved alle fire universitetene, vil bli lagt inn i den felles databasen. Da burde databasen få enda større nytteverdi for forskere, planleggere og journalister.

Den humanistiske forskningsdokumentasjonstjenesten har også skiftet navn. Det tidligere SIF-H er nå blitt HUM-DOK, som vel egentlig er selvforklarende, slik at det ikke skulle ta for lang tid for brukerne å venne seg til det nye navnet.

# *PROGRAM I BRUK VED SENTERET*

---

---

*Arne Hallheim*

Dette er en ny spalte i bladet. Her vil vi gi en oversikt over nye programmer som vi anskaffer og som vi etter hvert opparbeider kompetanse på. Siden dette er første gang, vil vi denne gangen ta med de viktigste programmene som vi bruker ved Senteret. Det vil siden komme nærmere omtale av en del av de programmene som er listet opp i denne oversikten.

## *MACINTOSH*

---

### **4th Dimension v2.1.1 norsk**

Beskrivelse: Databaseprogram; håndterer tekst og bilder, relasjoner.

### **Aldus Persuasion v2.0**

Beskrivelse: Presentasjonsprogram

### **Authorware Professional**

Beskrivelse: Multimedia forfatterverktøy. Inkluderer Farallon Mac-Recorder hardware for digitalisering av lyd.

### **DOS Mounter v2.01**

Beskrivelse: PC/MS-DOS Desktop Utility. Tillater montering av DOS-disketter som om de var Mac-disketter.

### **Excel v2.2**

Beskrivelse: Regneark-program med forretningsgrafikk og database.

### **HyperCard v2.0**

Beskrivelse: System for arkivering, henting og presentasjon av informasjon og utvikling av programvare.

### **HyperCard CD Audio Toolkit**

Beskrivelse: Funksjoner for styring av CD-ROM-spiller fra HyperCard.

### **Illustrator v3.0**

Beskrivelse: Farge illustrasjons/tegne-program. Lager fargeseparasjoner på Postscript-skriver. Adobe Type Manager v2.0 følger med.

**MacRecorder Sound System v2.0.3**

Beskrivelse: Digitalisering/behandling av lyd.

**PhotoShop**

Beskrivelse: Bildebehandlingsprogram for farge/punktgrafikk, lager fargeseparasjon på Postscript-skriver.

**SUM II**

Beskrivelse: Disk(ett) hjelpeprogram. Redder data fra ødelagte filer m.m.

**Soft PC v1.3**

Beskrivelse: DOS-emulator (tillater kjøring av DOS-programmer (f.eks. DataFlex)) for Macintosh II og SE/30, emulerer 8088, CGA og DOS v3.3.

**Soft PC, EGA/AT Option Module v1.4**

Beskrivelse: Tilleggsmodul til Soft PC som emulerer 80286/80287 og EGA.

**SuperCard v1.5**

Beskrivelse: Alternativt HyperCard. Støtter alle kortstørrelser og 8-bits farger. Alle objekter kan koples til et script.

**The Voyager VideoStack v2.0**

Beskrivelse: HyperCard Tool Kit for interaktiv video. Styrefunksjoner for videospillere. Kan også brukes med SuperCard.

**Think Pascal v3**

Beskrivelse: Pascal-kompilator med debugging-verktøy.

**Type Manager v2.0**

Beskrivelse: "Display Postscript". Bygger opp skjerm/skrift-fonter basert på outlines.

**Word v4.0**

Beskrivelse: Tekstbehandling. Enklere kontorsetting.

**The Voyager CD AudioStack**

Beskrivelse: Styring av lyd-CD'er og lyddelen av "mixed-mode" CD'er fra HyperCard.

***PC/MS-DOS:***

---

**Advanced Revelation v2.0**

Beskrivelse: Databasesystem med fri feltlengde, gode indekseringsrutiner og god tekstbehandling. Muligheter for repeterende felt og underfelt. Her i landet brukt til biblioteksystem og museums kataloger.

**Brief v3.0**

Beskrivelse: Basic Reconfigurable Interactive Editing Facility (editor). Best egnet for skriving av programmer (kildekode). Kompilator-støtte. Mange muligheter for automatisering av finn&bytt (makro-program-



mer), formatering (innrykk). Kan behandle flere filer samtidig, størrelsen begrenset til diskplass. Linjelengden begrenset til 512 tegn. Bonusdiskett med 'Best of Brief macros' følger med.

#### **DESQview-386 v2.3**

Beskrivelse: Multitasking-system for DOS-baserte programmer. DESQview v2.3 og QEMM386 v5.1 utgjør DESQview-386 v2.3. Inkluderer også Manifest v1.01.

#### **DataFlex v2.3 norsk**

Beskrivelse: Databasesystem, brukes til museums- og bibliotekskataloger.

#### **LapLink v3.00a**

Beskrivelse: Filoverføring mellom to serielt/parallelt oppkoblede PC'er. Universalkabel medfølger. Automatisk installering på "remote" PC gjennom seriekabelen.

#### **Lbase v5.0**

Beskrivelse: Fritekstsøkesystem med innebygde muligheter for logging. Kan brukes mot TLG CD-ROM.

#### **Manifest v1.01**

Beskrivelse: Oversikt over hardware-ressurser (memory, I/O-adresser osv.). Bundlet med DESQview-386 v2.3.

#### **Micro OCP**

Beskrivelse: Batch-orientert ordliste og konkordansprogram.

#### **NetRoom v1.12**

Beskrivelse: Memory Management. Kan i utgangspunktet laste nettverksdrivere/TSR til high memory eller expanded memory på alle maskiner med ext/exp RAM.

#### **Nota Bene v4.0**

Beskrivelse: Akademisk tekstbehandling med tekstdatabase. Enklere kontorsetting. Gode indekserings- og fotnotefunksjoner.

#### **PC-NFS v3.0.1**

Beskrivelse: Network File System, nettverksoperativsystem med UNIX-maskin som server. Deling av filer (felles data) og skrivere mellom brukere i nettverket.

#### **PC-NFS LifeLine v1.0**

Beskrivelse: Tilleggsmodul til PC-NFS med elektronisk post og backup.

#### **Paradox v3.5**

Beskrivelse: Database-system. SQL Link gjør Paradox klar for bruk mot SQL-server.

#### **QEMM386 v5.1**

Beskrivelse: RAM-håndteringsprogram. Konverterer ext.memory til exp.mem. Støtter XMS (kan brukes istedenfor HIMEM.SYS) og EMS 4.0. Kan laste TSR (residente) programmer over 640k-grensen (loadhi).

OPTIMIZE finner de beste kombinasjonene av TSR-loadhi. Levert sammen med DESQview-386 (som ett produkt).

#### **SpinRite II v1.1**

Beskrivelse: Harddisk-vedlikehold (lav-nivå reformaterer uten å ødelegge data, endrer inter-leave, tester sikkerheten for data (lesbarheten)).

#### **Tact v1.2**

Beskrivelse: Shareware-produkt utviklet ved University of Toronto's Computing Services (UTCS). Tekst-indeksering og søk.

#### **Turbo Pascal Professional v6.0**

Beskrivelse: Programmerings-miljø. Professional består av Turbo Pascal v6.0, Turbo Assembler v2.01, Turbo Debugger v2.01 og Turbo Profiler v1.01. Nytt i TP60 er Turbo Vision, et bibliotek for å lage vindusbaserte applikasjoner.

#### **Vedit Plus v3.31**

Beskrivelse: Editor, passer best for programutvikling (skrivning av kildekode). Kan arbeide med flere filer samtidig, størrelsen på filene bare begrenset av diskplass. Makromuligheter, klipp og lim (med 36 registre tilgjengelig for mellomlagring), kompilator-støtte, pop-up ASCII-tabell, inntil 100 siste endringer kan angres. Hurtig lesing av filer fra disk/eksekvering av makroer, tar liten plass (i RAM).

#### **Ventura Publisher Profesional Ext v2.0**

Beskrivelse: Kontorsetting/desktop publishing, basert på GEM grensesnitt. Utnytter utvidet minne (expanded).

#### **Windows v3.0, norsk**

Beskrivelse: Grafisk bruker-grensesnitt. Multitasker Windows-programmer (og DOS-programmer på 386/486-maskin). Gir Macintosh-liknende arbeidsmåte.

#### **Word Perfect Office v1.0**

Beskrivelse: Kontorstøttesystem med bl.a. møteplanlegger/utstyrsreservasjon, oppstarting av programmer fra meny.

#### **WordCruncher v4.3**

Beskrivelse: Tekstindeksering og søkeprogram, konkordans.

#### **WordPerfect v5.1 norsk**

Beskrivelse: Tekstbehandlingsprogram, enkel kontorsetting. Mange tastetrykk å huske for enkle funksjoner.

#### **askSam v4.1**

Beskrivelse: Søking i fritekst/database-program med hypertekstmuligheter.

#### **dBase IV v1.1**

Beskrivelse: Databasesystem

## **WINDOWS**

---

### **Corel Draw v2.0**

Beskrivelse: Farge tegne/illustrasjonsprogram. Tilleggsmoduler for autotracing, font-konvertering og filhåndtering. 6MB clip-art bibliotek. Meget godt program, enkelt å bruke. På topp i tester sammen med Designer.

### **Designer v3.01**

Beskrivelse: Farge tegne/illustrasjonsprogram. Skalerbare out-line fonter. SlideShow-utility for presentasjon. Autotrace (også farge-bitmap). Kan gjøre "alt", ikke så lett å lære som CorelDraw. På topp i tester sammen med CorelDraw.

### **Excel v3.00**

Beskrivelse: Regnearkprogram med forretningsgrafikk og database. Q+E (Querying & Editing) er en egen modul for manipulering/oppdatering av også dBase, MS SQL Server, Oracle og OS/2 EE databaser.

### **PowerPoint for Windows v2.0**

Beskrivelse: Presentasjonsprogram for produksjon av overheads, slides, slideshow, handouts. Kan også kjøre presentasjonen direkte på maskinen.

### **ToolBook v1.0**

Beskrivelse: Software Construction Set for Windows. Bygge egne applikasjoner uten programmering. Minner om HyperCard.

### **Word for Windows v1.10**

Beskrivelse: Tekstbehandling. Grafisk brukergrensesnitt. Enklere kon-torsetting. Egner seg godt til lengre tekster. Krever kraftig maskin med min. 2MB RAM.

## **Annet:**

---

### **3+Share v1.6**

System: 3Com-servere, personlige datamaskiner. Beskrivelse: Nettverksoperativsystem for deling av filområder (felles data) og/eller skrivere, elektronisk post. Knytter sammen Macintosh og PC.

### **EtherShare v1.3.3**

System: Sun 386/i

Beskrivelse: Mac'ene kan bruke Sun386/i som fil/utskrift-server.

# TACT

## ET PROGRAM FOR INDEKSERING OG SØKING I TEKST

---

---

*Knut Hofland*

TACT er et gratisprogram for MS-DOS som er utviklet ved *University of Toronto*. På samme måte som *WordCruncher*, som det er naturlig å sammenligne med, består TACT av to deler, MAKBAS for indeksering og TACT for søking. *WordCruncher* er et kommersielt program som er omtalt i HD nr 3/89.

MAKBAS starter med teksten og krever at brukeren i 3 skjermbilder gir opplysninger om alfabetet, spesielle tegn og strukturelle forhold. Denne informasjonen blir bevart i en egen parameterfil *?.MKS*, se figur 1.

|                            |   |
|----------------------------|---|
|                            | MAKBAS Version 1.2 (June 1990)  |
| Files:                     | Input File: DUKKTEST.TXT<br>Setup File: DUKKTEST.MKS<br>Database: DUKKTEST.TDB<br>DB Title: Et dukkehjem, Henrik Ibsen<br>Starting Ref: |
|                            | Temporary File Area:  |
|                            | Copyright (c) 1990 by John Bradley, Lidio Presutti  |
| F1: Help                   |   |
| F2: Alphabet               |   |
| F3: Special Characters     | F9: RUN, creating a database  |
| F4: Structural Information | F10: QUIT without a run   |

Figur 1.

I spesifikasjonen av alfabetet angir en de tegn som et ord kan bestå av og sorteringsrekkefølgen for disse, se figur 2. Inntil 6 tegn kan

grupperes som et symbol. Diakritiske tegn kan inkluderes i et ord og kan enten tas hensyn til eller ikke ved sortering. TACT bruker et internt kodesett (8 bit) ved representasjon av teksten og kan f.eks. legge inn fullt gresk tegnsatt i øvre del av tabellen. Med programmer som *Duke Language Toolkit* kan disse også vises på skjermen. Bokstaver med aksenter lagres som to tegn. Hvert tegn har et internt navn. Ved beskrivelsen av alfabetet kan en bruke disse navnene på tegn som normalt ikke finnes på tastaturet. I en tabell er det definert hvilke taster som brukes for å angi et tegn og hvorledes dette skal vises på skjermen. Standardtabellen bruker f.eks. ALT A for inntasting av en a med acute aksent. Dette kan imidlertid forandres til at tegnet kan skrives med dødtast for aksent og deretter bokstaven. En kan ikke bruke ALT og siffer fra det numeriske tastatur. Tilsvarende tabeller kan brukes ved utskrift for å konvertere tekst til spesielle skrivere.

| LETTERS, DIACRITICS, and CONTINUATION Character  |   |
|--|---|
| Alphabet   |   |
| a b c d e f g h i j k l m n o p q r s t u v w x y z æ ö å  |   |
| 0 1 2 3 4 5 6 7 8 9 -  |   |
| Diacritics Retained  |   |
| \Acute\ \Grave\ \Circumflex\ \Cedilla\ \Umlaut\ "'   |   |
| Diacritics NonRetained   |   |
| Continuation Sequence:   | _ |
| To edit or add material, move to line using UP or DOWN arrow Keys & type.<br>When finished, push the ENTER/RETURN key. |   |
| Letters and Word Symbols   |   |

Figur 2.

I avdelingen for spesialtegn kan en angi om det er tekst som ikke skal indekseres og hvilke tegn som omslutter denne type tekst. Videre kan det angis sekvenser av tegn som skiller ord. En definerer også her hvilke tegn som omslutter referanseopplysninger i teksten.

De siste opplysningene som må gis, er om strukturelle forhold i teksten, se figur 3. Programmet skiller mellom referanse, merkelapp og tellere. Det er to innebyggete tellere, en for ord og en for linjer. Referanser har et navn og en verdi, f.eks. <side 1> og blir ikke indeksert. En spesiell referansetype som kan angis er den som WordCruncher bruker slik som f.eks. !BVildanden. Merkelapper har bare verdi og kan indekseres om en ønsker. Disse kan f.eks. brukes til replikknavn i skuespill, f.eks. {PEER GYNT}. Normalt vil TACT bruke referanse

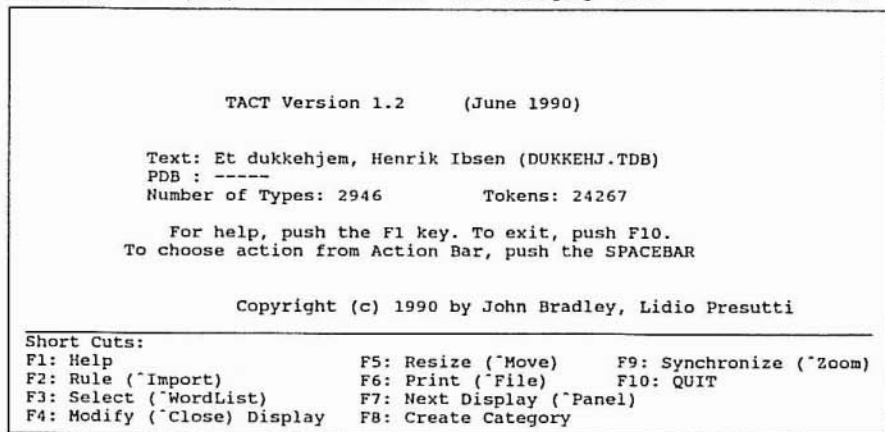
til linjenummer i tekst- eller indeksvinduer, men under indeksering kan en angi at egne referanser skal brukes som standard.

|   |        |            |        |          |          |   |
|---|--------|------------|--------|----------|----------|---|
| Type:   | Empty: | Reference: | Label: | Counter: | BYU:     | X |
| TACT ID: akt  |        |            |        |          |          |   |
| Text ID: K                      Numeric: Yes: _ No: X                         |        |            |        |          |          |   |
| No of chars to keep: 9999   |        |            |        |          |          |   |
| -----   |        |            |        |          |          |   |
| Type:   | Empty: | Reference: | Label: | Counter: | BYU:     | X |
| TACT ID: side   |        |            |        |          |          |   |
| Text ID: S                      Numeric: Yes: _ No: X                         |        |            |        |          |          |   |
| No of chars to keep: 9999   |        |            |        |          |          |   |
| -----   |        |            |        |          |          |   |
| Type:   | Empty: | Reference: | Label: | X        | Counter: | _ |
| TACT ID: rep replikk  |        |            |        |          |          |   |
| Open Bracket: {                      Close Bracket: }                         |        |            |        |          |          |   |
| Suppress Text: Yes: _ No: X                      Word Separator: Yes: X No: _ |        |            |        |          |          |   |
| Numeric: Yes: _ No: X                      No of chars to keep: 9999          |        |            |        |          |          |   |
| -----   |        |            |        |          |          |   |
| F3: Special Characters                      F9: RUN, creating a database      |        |            |        |          |          |   |
| F4: Structural Information                      F10: QUIT without a run       |        |            |        |          |          |   |

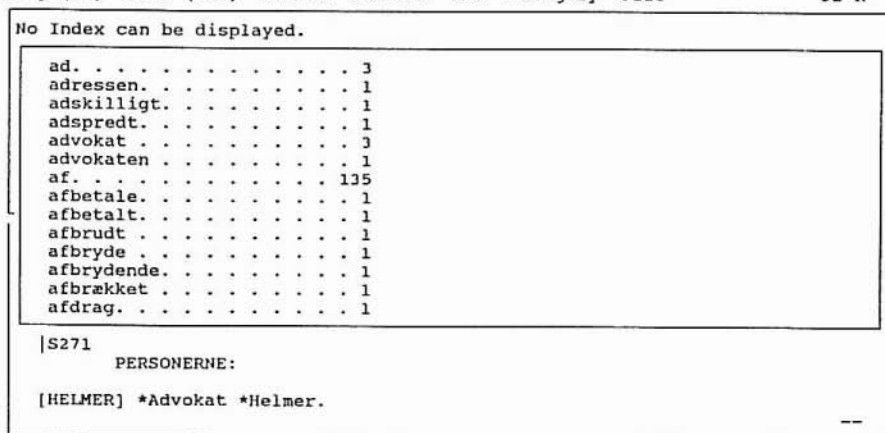
Figur 3.

Etter at alle opplysningene er gitt, kan en starte selve indekseringen og MAKBAS viser hvor stor andel av teksten som til en hver tid er behandlet. Hvis en forandrer parameterne må en indeksere teksten på ny. Men en får ikke gjøre dette dersom det allerede eksisterer en indeks. Da må en avslutte MAKBAS og slette denne filen i DOS. Parametrene blir heller ikke lagret dersom en avslutter programmet uten å indeksere. Dette virker noe tungvint. Dersom tekstfilen er for stor, må en dele denne opp i mindre deler og indeksere disse hver for seg. Inntil 4 deler kan deretter flettes sammen med programmet MERGEBAS. Både MAKBAS og MERGEBAS kan kalles opp i BAT filer, slik at en kan automatisere indekseringen av store tekster. Den optimale størrelsen på en tekstfil er ca. 200 KB.

I motsetning til WordCruncher samler MAKBAS tekst og indeks i én fil, kalt tekstdatabase. For en råtekst på 156 KB (Ibsens "Et dukkehjem"), utgjør denne filen 444 KB, dvs. en økning på ca. 185% (mot 75% for WordCruncher). Indekseringen av tekst går noe hurtigere ved TACT enn ved WordCruncher. På en standard PC tar indeksering av en fil på 156 KB 20:18 minutter, mot 31:24 for WordCruncher. Tilsvarende tall for en rask AT maskin er 6:22 og 13:48 (7:12 for WordCruncher ved bruk av disk cache) og for en 386-maskin 2:35 og 8:48 (2:36 med disk cache). MAKBAS gjør bruk av all ledig hukommelse opptil 640 KB og en bør derfor indeksere uten residente programmer eller nettverksdrivere.



Figur 4.



Figur 5.

Åpningsbildet for søkeprogrammet TACT er vist i figur 4. Øverst er det en meny og nederst en snarvei til noen av undermenyene. En kan starte med å bla i ordlisten se figur 5. I denne kan en plukke ut ord som en vil se i kontekst, og deretter åpne et KWIC-indeksvindu og et tekstvindu. Disse vinduene kan plasseres på samme skjerm bilde, se figur 6. Innenfor vinduene kan en bla med piltastene. Det er også mulig å se hvorledes forekomstene av en gruppe ord fordeler seg etter en referansekategori eller merkelapp, se figur 7. De forskjellige vinduer

kan skrives til fil eller skriver, men dette er dessverre ikke mulig med ordlisten. Et eget vindu (Collocation) gir ord som er brukt i nærheten av en angitt gruppe ord. Dette vinduet kan brukes til å analysere ordsammenstillinger, og vinduet er sortert etter et statistisk mål (Z-score) som ordner de mest signifikante ordsammenstillinger øverst i vinduet. I alt er det fem display-vinduer: INDEX, KWIC, TEXT, DISTRIBUTION OG COLLOCATION.

Help (F1) eXit (F10) Select Current New cAtegory File 66 K

```

▶ vidunderlige (9)
(2 331 NORA) forstå det? Det er jo | det >vidunderlige, som nu vil ske.
(2 331 LINDE) ske. | (LINDE). [NORA] Det >vidunderlige? | (NORA).
(2 331 NORA) | (NORA). [LINDE] Ja, det >vidunderlige. Men det er
(2 336 NORA) her, at | gå og vente på det >vidunderlige. | (LINDE).
(3 346 HELMER) skuldre, - om denne >vidunderlige | nakkebøjning, -
(3 361 NORA) Det var iaften, da det | >vidunderlige ikke kom; for da
(3 361 NORA) jeg innså jo nok, at det >vidunderlige kommer | ikke
(3 361 NORA) viss på: nu kommer | det >vidunderlige. Da Krogstads
(3 362 NORA) for dine? - *Det var det | >vidunderlige, som jeg gikk og
Index Display
(LINDE). [NORA] Det skal jeg nok. Men jeg forstår ikke alt
dette.
(NORA). [LINDE] Å, hvor skulde du kunne forstå det? Det er jo
▶ det vidunderlige, som nu vil ske.
(LINDE). [NORA] Det vidunderlige?
(NORA). [LINDE] Ja, det vidunderlige. Men det er så forførdeligt,
Kristine, - det *må ikke ske, ikke for nogen pris i
verden.
(LINDE). [NORA] Jeg vil gå lige hen og tale med Krogstad.
(NORA). [LINDE] Gå ikke hen til ham; han gør dig noget ondt.
(2 331 NORA)
Text Display

```

Figur 6.

Help (F1) eXit (F10) Select Current New cAtegory File 90 K

```

(??? 269 ???)
??? █ = 13, Total: 1543
BARNEPIGEN 11
BØRNENE 2
HELMER 396
KROGSTAD 167
LINDE 145
NORA 731
RANK 89
Distr Display

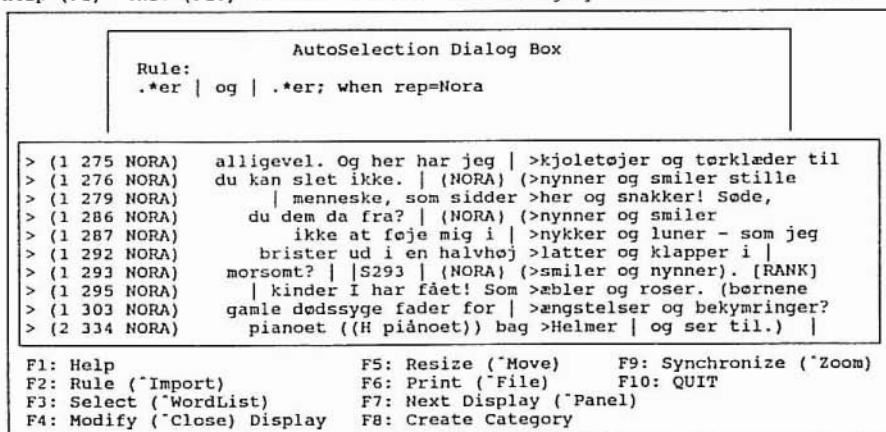
```

Figur 7.



Det er to ulike søkemuligheter. Én går via ordlisten, som vi har sett. En annen inngang til materialet er ved hjelp av søkemønstre som kan bestå av regulære uttrykk. I en posisjon kan en enten ha et bestemt tegn, en gruppe tegn eller et vilkårlig tegn. Ord eller uttrykk kan kombineres til fraser eller en kan spesifisere maksimal avstand mellom disse. Uttrykkene kan også kombineres med referansevariablene eller frekvensopplysninger slik at søkemønsteret kan bli svært komplekst. Søkemønstre kan lagres på egne filer. Denne søkemuligheten er mye kraftigere enn den forholdsvis enkle søkemulighet som er i WordCruncher, se figur 8.

Help (F1) eXit (F10) Select Current New cAtegory File 232 K



Figur 8.

I tillegg til tekstdatabasen kan en opprette en personlig database. I denne kan en definere kategorier av grupper av ord og på denne måten lage en tesaurus. Det er også mulig å foreta en manuell homograf-separering ved å splitte forekomster av et ord til to forskjellige kategorier.

TACT inneholder en mulighet til å lagre tastetrykk på en fil (script) og senere få utført disse. I denne filen kan det legges inn kommentarvinduer og pauser. Hastigheten til utførelsen av tastetrykkene kan også angis. Dette er en god mulighet til å lage pedagogiske opplegg rundt en tekst, se figur 9.

Et separat program, COLLGEN, kan generere sammenstillinger av ord som opptrer flere ganger (kollokasjoner). Det er mulig å angi maksimalt antall ord i sammenstillingen og minimumfrekvens.

Søkingen i ordlisten tar noe lenger tid en med WordCruncher, hvor oppslaget er mer eller mindre umiddelbart. Versjon 1.2 går også ut i feil pga. for lite hukommelse dersom en prøver å se på kontekster til

et ord eller en gruppe som har mer en ca. 3.000 forekomster, bruker en komplisert søkemaske eller prøver å finne for lange ordsammenstillinger. Programmet kan ikke gjøre bruk av ekstra hukommelse utover 640K. Dette er imidlertid feilsituasjoner som utviklerne jobber med. Programmet får også problemer dersom et ord har mer en 65.000 forekomster, da blir frekvensopplysningene i ordlisten feilaktige og etterfølgende ord i ordlisten får feil kontekst.

Help (F1) eXit (F10) Select Current New cAtegory File 146 K

|   |  |   |
|---|--|---|
| <p>TACT</p> <p>Text: Et du<br/>PDB : ----<br/>Number of T</p> <p>For help,<br/>To choose act</p> <p>Cop</p> <hr/> <p>Short Cuts:<br/>F1: Help<br/>F2: Rule (^Import)<br/>F3: Select (^WordList)<br/>F4: Modify (^Close) Display</p> | <p>Velkommen til en styrt gjennomgang av programmet TACT brukt på Ibsens "Et dukkehjem".</p> <p>Hver gang du ser en slik melding, trykk en vilkårlig tast for å fortsette, F10 avslutter og ESC avbryter.</p> <p>Se øverst i høyre hjørne. Dette tallet angir ledig hukommelse. Du bør ha ca. 200 K hvis ikke avslutt og ta vekk residente program og drivere.</p> | <p>AR</p> <p>io Presutti</p> <hr/> <p>ynchronize (^Zoom)<br/>QUIT</p> |
| <p>F7: Next Display (^Panel)<br/>F8: Create Category</p>  |  |   |

Figur 9.

Til TACT medfølger det en manual på 177 sider og en eksempeldatabase som brukes som illustrasjon i manualen. Programmet virker gjennomtenkt og har kvaliteter som går ut over kommersielle programmer som OCP og WordCruncher. I senere versjoner blir forhåpentligvis programmet mer robust mht. brukergrensesnitt og administrasjon av hukommelse.

TACT distribueres fra Senteret med manual til selvkost for 250 kr. inkl. porto. Beløpet kan betales til postgirokonto 0802 3384567 eller bankgiro 3625.88.53657 merket TACT. Programmet kan også hentes via anonym FTP fra maskinen nora.navf-edb-h.uib.no (129.177.24.42) i katalogen /pub/pc/tact.

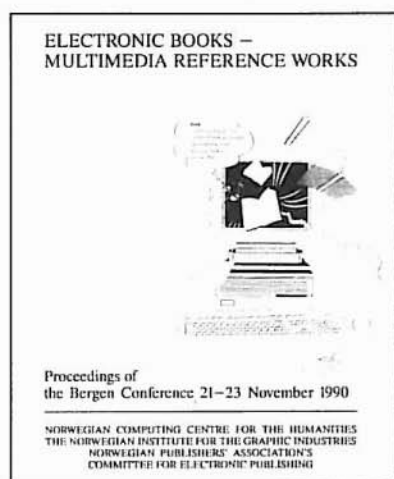
# KONFERANSERAPPORT

---

---

## **ELECTRONIC BOOKS – MULTIMEDIA REFERENCE WORKS: PROCEEDINGS OF THE BERGEN CONFERENCE 21-23 NOVEMBER 1990**

---



Konferansen "Elektroniske bøker – multimedia oppslagsverk" ble arrangert av NAVFs edb-senter for humanistisk forskning, Grafisk Institutt, og Den norske forleggerforenings komite for elektronisk publisering.

17 foredragsholdere fra inn- og utland presenterte forskningsprosjekter og diskuterte hva de elektroniske mediene kan komme til å bety for forfattere, utgivere og brukere. Foredragene omhandlet teknologiske, økonomiske og juridiske aspekt ved de nye mediene, og et hovedtema var de pedagogiske utfordringene som ligger i å tilrettelegge tekst, bilder og

lyd for brukerne. Flere av foredragsholderne tok opp sentrale begrep som "interaktivitet" og "linearitet/ikke-linearitet".

Nå foreligger det en fyldig rapport som inneholder de fleste av konferansens foredrag, en ny artikkel og en anmeldelse av konferansen. Fordi mange av foredragene ble holdt på engelsk og emnet har interesse ut over Skandinavia, har redaktørene valgt å gjøre rapporten tilgjengelig på engelsk, med engelske sammendrag av skandinaviske foredrag.

Rapporten *Electronic Books – Multimedia Reference Works: Proceedings of the Bergen Conference 21-23 November 1990* er på 195 sider, koster kr. 175 + porto, kr 20,-, og kan bestilles fra: NAVFs edb-senter for humanistisk forskning, Boks 53 Universitetet, N-5027 Bergen. Tlf: + 47 (0)5 21 29 54/55/56. Fax: + 47 (0)5 32 26 56.

Innholdslisten er gjengitt på neste side.

**ELECTRONIC BOOKS – MULTIMEDIA REFERENCE  
WORKS: PROCEEDINGS OF THE BERGEN  
CONFERENCE 21-23 NOVEMBER 1990**

---

**Contents**

|   |    |
|---|----|
| Preface <i>Jostein H. Hauge</i> .....   | 3  |
| A Taste of the Electronic Library <i>Espen Aarseth</i> .....  | 7  |
| Multimedia CD – The Enabling Technology for Third Wave<br>Computing <i>Graham Brown-Martin</i> .....  | 15 |
| Elektroniske bøker og multimedia – temaer og<br>problemstillinger/Electronic Books and Multimedia – Themes<br>and Problems <i>Øyvin Rannem</i> .....  | 17 |
| Portable Electronic Publishing <i>Bela Hatvany</i> .....  | 31 |
| Redaksjon og design av multimedia-produkter – LaserAtlas<br>over Møre og Romsdal/Editing and Designing Multimedia<br>Products – LaserAtlas of the Norwegian County Møre og<br>Romsdal <i>Knut Brustad</i> .....   | 35 |
| What Publishers Need to Know about the Hypermedia<br>Textbook: The Example of the Intermedia Emblem Web<br><i>George P. Landow and Gary S. Weissman</i> .....   | 63 |
| Interactivity as a New Phenomenon in Literature and the<br>Presentation of Information <i>Tor Edvin Dahl</i> .....  | 83 |
| Informasjonsarkitektur – Informasjons-boomens kritiske<br>suksessfaktor? En nøkkeldisiplin innen IT, med boklige<br>tradisjoner/Information Architecture – the Crucial Success<br>Factor of the Information Boom? A Key Discipline within IT,<br>with Literary Traditions <i>Petter Henriksen</i> ..... | 85 |
| Electronic books – a New Dimension in Publishing<br><i>Jens Erlandsen</i> .....   | 99 |

|  |     |
|--|-----|
| Producing Multimedia Titles: a Closer Look at the Design Phase<br><i>Patrizia Ghislandi</i> .....  | 101 |
| CD-ROM – En begränsad nischmarknad eller ett potentiellt verktyg för breda användargrupper?/CD-ROM – A Narrow Market Area or a Potential Tool for wide User Groups?<br><i>Tomas Bjöersdorff</i> .....  | 113 |
| Multipublisering – informasjonsprodusentens muligheter og begrensninger i fremtidens teknologi og marked/<br>Multipublishing – the Information Possessor's Options and Information Strategies within the Future Technology and Market<br><i>Erling Maartmann-Moe</i> ..... | 121 |
| Multimedia Compact Discs and their Design<br><i>Dick Fletcher</i> .....  | 135 |
| Perspektiver på hypermedia og sakprosaens posisjon/Hypermedia and the Status of Prose<br><i>Gunnar Liestøl</i> .....   | 143 |
| Intellectual Property and Multimedia Products – the Experience of SIULLEQ/the Greenland Project<br><i>Peter Olaf Looms</i> .....   | 157 |
| Samarbete mellan förlag och forskningsmiljö för att bygga upp kompetens inom elektronisk publicering/Co-operation between Publishers, Graphic Industry and Research to Develop Competence within Electronic Publishing<br><i>Kari Marklund</i> .....                       | 167 |
| The Perseus Project: Ancient Greece in Texts, Maps and Images<br><i>Elli Mylonas</i> .....   | 173 |
| Program .....  | 189 |
| Liste over demonstrasjoner/List of Demonstrations .....  | 191 |
| Mediedekning/Media Coverage .....  | 195 |

# ENGLISH COMPUTER CORPORA

## MATERIAL AVAILABLE THROUGH ICAME

---

---

*The Norwegian Computing Centre for the Humanities* (NCCH) functions as secretariat and clearing house for machine readable English texts. The secretariat publishes *ICAME Journal*, which brings articles and information on corpus linguistic topics. The information given below is gleaned from *ICAME Journal*. More information about ICAME and the text material is available from NCCH.

The following material is currently available through the *International Computer Archive of Modern English* (ICAME):

- **Brown Corpus, untagged text format I** (available on tape or diskette): A revised version of the Brown Corpus with upper- and lower-case letters and other features which reduce the need for special codes and make the material more easily readable. It contains approximately a million words of printed text (500 text samples of about 2,000 words). A number of errors found during the tagging of the corpus have been corrected. Typographical information is preserved; the same line division is used as in the original version from Brown University except that words at the end of the line are never divided.
- **Brown Corpus, untagged text format II** (tape or diskette): This version is identical to text format I, but typographical information is reduced and the line division is new.
- **Brown Corpus, KWIC concordance** (tape or microfiche): A complete concordance for all the words in the corpus, including word statistics showing the distribution in text samples and genre categories. The microfiche set includes the complete text of the corpus.
- **Brown Corpus, WordCruncher version** (diskette): This is an indexed version of the Brown Corpus. It can only be used with WordCruncher (for MS-DOS). See the article by Randall Jones, *ICAME Journal* 11, pp. 44-47.
- **LOB Corpus, untagged version, text** (tape or diskette): The LOB Corpus is a British English counterpart of the Brown Corpus. It contains approximately a million words of printed text (500 text samples of about 2,000 words). The text of the LOB Corpus is not available on microfiche.

- **LOB Corpus, untagged version, KWIC concordance** (tape or microfiche): A complete concordance for all the words in the corpus. It includes word statistics for both the LOB Corpus and the Brown Corpus, showing the distribution in text samples and genre categories for both corpora.
- **LOB Corpus, tagged version, horizontal format** (tape or diskette): A running text where each word is followed immediately by a word-class tag (number of different tags: 134).
- **LOB Corpus, tagged version, vertical format** (available on tape only): Each word is on a separate line, together with its tag, a reference number, and some additional information (indicating whether the word is part of a heading, a naming expression, a quotation, etc).
- **LOB Corpus, tagged version, KWIC concordance** (tape or microfiche): A complete concordance for all the words in the corpus, sorted by key word and tag. At the beginning of each graphic word there is a frequency survey giving the following information: (1) total frequency of each tag found with the word, (2) relative frequency of each tag, and (3) absolute and relative frequencies of each tag in the individual text categories.
- **LOB Corpus, WordCruncher version** (diskette): This is an indexed version of the tagged LOB Corpus (horizontal format). It can only be used with WordCruncher (for MS-DOS).
- **London-Lund Corpus, text, original version** (computer tape or diskette): The London-Lund Corpus contains samples of educated spoken British English, in orthographic transcription with detailed prosodic marking. It consists of 87 'texts', each of some 5,000 running words. The text categories represented are spontaneous conversation, spontaneous commentary, spontaneous and prepared oration, etc.
- **London-Lund Corpus, KWIC concordance I** (computer tape): A complete concordance for the 34 texts representing spontaneous, surreptitiously recorded conversation (text categories 1-3), made available both in computerized and printed form (J. Svartvik and R. Quirk (eds.) *A Corpus of English Conversation*, Lund Studies in English 56, Lund: C.W.K. Gleerup, 1980).
- **London-Lund Corpus, KWIC concordance II** (computer tape): A complete concordance for the remaining 53 original texts of the London-Lund Corpus (text categories 4-12).
- **London-Lund Corpus, supplement** (diskette): The remaining 13 texts of the 100 spoken texts collected and transcribed at the Survey of English Usage, University College London. See the presentation by S. Greenbaum, *ICAME Journal* 14 (1990), pp. 108-110.

- **Melbourne-Surrey Corpus** (tape or diskette): 100,000 words of Australian newspaper texts (see the article by Ahmad and Corbett, *ICAME Journal* 11, pp. 39-43).
- **Kolhapur Corpus** (tape or diskette): A million-word corpus of printed Indian English texts. See the article by S.V. Shastri, *ICAME Journal* 12, pp. 15-26.
- **Kolhapur Corpus, WordCruncher version** (diskette): This is an indexed version of the Kolhapur Corpus. It can only be used with WordCruncher (for MS-DOS).
- **Lancaster/IBM Spoken English Corpus** (tape or diskette): A corpus of approximately 52,000 words of contemporary spoken British English. The material is available in orthographic and prosodic transcription and in two versions with grammatical tagging (like those for the LOB Corpus). There is an accompanying manual. See further *ICAME Journal* 12, pp. 76-77.
- **Polytechnic of Wales Corpus** (tape or diskette): Orthographic transcriptions of some 61,000 words of child language data. The corpus is parsed according to Hallidayan systemic-functional grammar. There is no prosodic information. See further *ICAME Journal* 13 (1989), p. 20ff, and the presentation of the edited version of the corpus in this issue.

Most of the material has been described in greater detail in previous issues of our journal. Prices and technical specifications are given on the order forms which accompany the journal. *Note that tagged versions of the Brown Corpus cannot be obtained through ICAME. The same applies to audio tapes for the London-Lund Corpus, the Lancaster/IBM Spoken English Corpus, and the Polytechnic of Wales Corpus.*

There are available printed manuals for the LOB Corpus (the original manual and a supplementary manual for the tagged version). Printed manuals for the Brown Corpus cannot be obtained from Bergen. Some information on the London-Lund Corpus is distributed together with copies of the text and the KWIC concordance for the corpus. Users of the London-Lund material are also recommended to consult J. Svartvik (ed.). *The London-Lund Corpus: Description and Research*, Lund University Press, 1990.

A manual for the Kolhapur Corpus can be ordered from: S.V. Shastri, Department of English, Shivaji University, Vidyanagar, Kolhapur-416006, India. The price of this manual is US \$15 (including airmail charges). Payment should be sent along with the order by cheque or international postal order drawn in favour of The Registrar, Shivaji University, Kolhapur.



# SUMMARIES

---

---

## *LITT OM BILDER OG DATAMASKINER OG LAGRINGSPLASS*

---

### *PICTURES, COMPUTERS AND STORAGE SPACE*

---

In this article the author discusses some of the possibilities and limitations for storing and using digitized black-and-white pictures. It is necessary to establish precisely how one plans to make use of such a picture base, since digitized images demand large storing capacity. When a picture is scanned into the computer, it is converted into a number of points, each of which is given a specific grey tone value. This information is stored as a number, and if the scanner distinguishes between 16 different tones of grey, we need 4 bits of storage capacity to store the information about each point.

However, 16 grey tones would still make a rather crude image, with sharp distinctions, very unlike the even gradations we are accustomed to in photographs. Twice the storage capacity, 8 bits of information per point, would give us 256 grey tones, which is sufficient for all practical purposes. The size of each point – or the number of points per square inch – is also of importance. A good standard computer screen today contains about 70 points per inch. This means that one picture of 8 by 11 inches with 256 grey tones demand a storage capacity of 3,449,600 bits or 431,200 bytes. That sounds like a lot, but if we double the number of points per square inch, we quadruple the digitized information. This picture would then demand more than 1.6 megabytes of storage space.

The practical problems of storage capacity demand a thorough analysis of how the picture base is to be used – now and in the future. If the pictures are to be used as a catalog and viewed on the computer screen, a solution of 70 points per inch will do nicely. If the pictures are to be printed, or enlarged after scanning, one may want a higher resolution. For analogous/digital conversion the author also applies Nyquist's theorem of sampling, which predicts that for the best results you should scan the picture at twice the solution that you plan to use. To print the picture with 70 points, scan it with 150 points per square inch.

The results are shown in the photo illustrations that accompany the article.

The author also shows how the digitized picture can be manipulated. Making rather unsophisticated use of available software, he has increased the contrasts of the photograph by making use of a higher number of grey tones than the picture originally contained (Fig. 2), or by taking out parts of the picture and putting in a different face. A nice piece of forgery, which the author says took 30 seconds to accomplish. What if he really wanted to falsify the photograph? He concludes that the new possibilities may discredit the photograph as evidence material.

## ***FOTOMAC VIDEREUTVIKLES***

---

### ***FOTOMAC IN A NEW AND IMPROVED VERSION.***

---

Fotomac is a picture storing program for Macintosh, developed largely by the author. A prototype of the program has earlier been described in HD. An early version of Fotomac has been tested at Oslo University Library for storing and cataloging a large collection of pictures taken by or showing Fridtjof Nansen. In the test period the library built up a catalog of some 400 pictures. On the basis of the test results, the 4,000 finest pictures, about one half of the Nansen collection, will be digitized and stored together with relevant information. This work is carried out in cooperation between Oslo University Library and NCCH.

In this article the author presents Fotomac, version 1.0, developed in Hypercard 2.0, which is considerably improved over earlier versions:

- Catalog cards may be of any size, up to 1280 by 1280 points.
- The system for displaying grey tones is incorporated in Hypercard and is more robust than the system previously used.
- Fotomac employs pre-defined print-out forms.
- The new version has improved menu control.

Fotomac v. 1.0 separates the data base completely from the user screen. The search panel was separated from the data base in the earlier versions, too, but now there is a separate form for data input and alterations. This provides improved safety for the stored data, and means that data need not even be stored in Hypercard: the data base may reside in a Unix server, while the user would be presented with the same Hypercard data/picture cards as today. It will also be possible to use logical operators and linked searches. Fotomac will enable searches for standard operators linked by *and*, *not* and *or*. *Larger/smaller than* searches will also be implemented, which will be useful when the date of a picture is of importance.

***DOKUMENTASJONSPROSJEKTET VED HF-FAKULTETET,  
UNIVERSITETET I OSLO***

---

***DOCUMENTATION PROJECT AT THE FACULTY OF THE  
ARTS, UNIVERSITY OF OSLO***

---

The Faculty of the Arts at the University of Oslo, has recently decided to carry out the first year of a six-year project of documentation. The goal is to collect all the material of the various archives and collections within the faculty in one data base. The archives will be transferred to modern electronic media, and will be made available to all users.

The project has been in the pipeline for some years, and in April 1990 the author was given the task of surveying all the archives and collections within the Arts Faculty, establishing the character, physical condition and size of the various archives in preparation for a comprehensive plan for the project. In February 1991 the decision was made to go ahead with the project.

The project is large: The goal is to register and make electronically available the information today registered in some 14 million documents and catalog cards. Registering the data is alone estimated to take several hundred man years. Also in professional terms it is a demanding undertaking. The material includes for instance the collections of the University Museum of National Antiquities, folk music collections and the archives of the lexicographic departments.

The article describes the various archives and collections, and presents examples of the types of cards and documents involved. They contain various kinds of information collected by both professionals and laymen over several decades, and to do the material justice will be an exacting task. The archives of the Department of Dialectology, for instance, contain voice recordings together with phonetic transcriptions, which may now be entered in a combined sound/text data base, enabling researchers not only to study the dialects but also to compare recordings and transcriptions.

Another large archive is the collection of some 30,000 photographs at the Department of History. These may now be digitized and combined with textual information, creating a valuable resource for research and education.

The author also discusses various technical solutions. The standard set-up is to store all data in one central computer unit which serves the users and runs all searches. This is a tested and practicable solution, but may easily become a bottleneck at peak user periods, since not just data but also displays have to be generated by the central unit.

Local computer power is only used to gain access to the central base.

The author suggests that a better alternative would be to install a common communications program both in the central unit and in the user computers. Searches and administration would still be carried out by the central unit, but displaying and organizing the data would be done by the local unit. Such a distribution of labour will radically diminish the network traffic. It is further possible to distribute the various data bases – the actual discs – to a number of physical locations, while the whole base could still be organized as one logical unit, and be experienced as such by the user, even though one single search might be drawing on data stored in several locations. With this kind of network, a breakdown in one part would leave the rest of the network intact.

The purpose of the project is, ultimately, to make all of the material more easily accessible to the users. At the present stage the main objective is to register the information faithfully in order to preserve the desired information of the original material. For years to come the material will be of interest to researchers, teachers and students, and will be accessible both in Norway and in other parts of the world via international electronic networks.

#### ***MULTIMEDIELABORATORIET VED AVH, UNIT***

---

#### ***THE MULTIMEDIA LABORATORY AT THE FACULTY OF ARTS, UNIVERSITY OF TRONDHEIM.***

---

In January 1990 one of the three language laboratories at the Faculty of Arts of The University of Trondheim was replaced by a multimedia laboratory. The equipment consists of 9 IBM PS/2, model 55SX X61, 1 IBM PS/2, model 70 121, and 1 IBM AT with a frame grabber for interactive video. There are also 5 Macintosh computers: 1 Mac II, 1 Mac IICx, 1 Mac SE/30, 1 Mac II fx 4/160 and 1 Mac II fx/80.

The two largest Macintosh machines have video frame grabbers and are connected to a scanner and to CD-ROM and video disc players. Four of the IBM machines have MIC-4000 frame grabbers for real time digital video and are connected to Sony Laservision LDP 3600D video disc players. In addition one of the IBM machines is connected to a Pioneer LD4100 video disc player. All computers have ethernet cards and are connected to the University's Vax machine via the local network.

The laboratory also has a CD-player for video and sound, assorted sound equipment, a VHS video player for digital picture, and satellite receiver and dish in order to take in international television transmissions.

The laboratory has been selected as an Apple Multimedia Centre.

The multimedia laboratory is used for educational purposes by several departments at the Arts Faculty, but it is primarily intended as a research laboratory for staff and post-graduates at the Arts Faculty. The laboratory is intended to "stimulate, initiate, coordinate and assist the use of multimedia in education and research" in the humanities.

Hence, the lab is today used for projects, courses and seminars for staff, graduate and post-graduate students. It is important to raise the level of competence in all of these groups. The author includes the course plan for the spring of 1991, to indicate the variety of courses possible in the new surroundings.

### ***DISTRIBUSJON AV INFORMASJON FRA SENTERET VIA ELEKTRONISKE NETT***

---

### ***ELECTRONIC INFORMATION DISTRIBUTION FROM THE NORWEGIAN COMPUTING CENTRE FOR THE HUMANITIES (NCCH)***

---

The Norwegian Computing Centre for the Humanities runs several electronic network services. The host machine for these electronic services is a Sun 3/60 called NORA, or in full, NORA.NAVF-EDB-H.UIB.NO, with IP address 129.177.24.42.

There is also a HELMER: a Sun 386i, with IP address 129.177.24.41. Helmer:NAVF-edb-h is an Apple Share file server in the *EtherMac* zone at the University of Bergen. This service is only accessible via Macintosh nets connected to the broadband network at the University of Bergen. There are two volumes that are open to everybody: GjesterLes and GjesterSkriv (GuestsRead/GuestsWrite). Helmer informs about the services available on *Les Meg* (Read Me) in GjesterLes.

NCCH also makes use of NOBERGEN.BITNET which is the IBM 4381 at the Bergen University. Here NCCH has installed the file server FAFSRV@NOBERGEN.BITNET in order to service people who are interested in machine readable English texts from ICAME (*International Computer Archive of Modern English*). This information is also found on NORA, which will gradually replace FAFSRV.

HUMEDB is an electronic mail address which contains a number of other addresses. All incoming mail goes out to all the addresses on the list (about 80 by March 1991). HUMEDB is dedicated to the topic of computing in the humanities, and anybody may be entered in the mailing list, manually, by the administrator. The full name of the mailing list is HUMEDB@NAVF-EDB-H.UIB.NO, and the address of the

administrator is **HUMEDB-ADM@NAVF-EDB-H.UIB.NO**.

NCCH will inform about its own activities on HUMEDB, and also send out other information. The participants are encouraged to make use of this e-mail service for opinions, queries and answers.

Those who wish to make use of this service need an electronic address in a host computer connected to Uninett, Internet, Bitnet, Janet, etc. In many institutions it is possible to read and write electronic mail on your own PC or Mac, using an "invisible" Unix machine as mailbox. This service is free of charge.

File servers on NORA contain general information about and from NCCH, articles, information from other sources and software. There are two ways to gain access to the NORA files: The first one is through the file transfer program FTP. The machine permits "anonymous FTP".

Access: **FTP NORA.NAVF-EDB-H.UIB.NO**

User: **Anonymous**

Password: **Your name**

Commands:

|                                      |                               |
|--------------------------------------|-------------------------------|
| <b>CD PUB</b>                        | To go to main catalog         |
| <b>DIR</b>                           | To see file list              |
| <b>CD &lt;dir. name&gt; or CD ..</b> | To change directory           |
| <b>GET &lt;file name&gt;</b>         | To fetch a file               |
| <b>BINARY</b>                        | To set binary file conversion |

Those who wish to make use of this service need access to a machine which has a so-called TCP/IP connection. This may be a Macintosh or a PC with NCSA Telnet/FTP, HyperFTP or a similar program installed. Or it may be a host machine accessed via serial or modem connection. The files have then first to be transferred to the host machines via FTP and then on to the local PC or Mac via Kermit or a similar program.

The second manner of access is mail based file transfer. **NAVFSERV@NORA.NAVF-EDB-H.UIB.NO** is a mail based file server on NORA. Some of our readers will be familiar with **INFO@UNINETT.NO**, which uses the same commands as NAVFSERV. Parts of the information available through FTP will be available through NAVFSERV, but binary files cannot be taken in directly through NAVFSERV. For more information, send e-mail, Subject: *HELP*.

**The file servers today contain the following main catalogs:**

|             |   |
|-------------|---|
| ICAME       | Information related to International Computer Archive of Modern English |
| INFO        | Information on texts, projects, bibliographies                          |
| KONFERANSER | Updated list of conferences   |
| MAC         | Macintosh software  |
| NAVF-EDB-H  | General information about and from NCCH (in Norwegian)                  |
| HUMEDB      | Backlog from HUMEDB (in Norwegian)                                      |
| OLUFF       | Articles from the periodical OLUFF (in Norwegian)                       |
| HD          | Articles from Humanistiske Data (in Norwegian)                          |
| NCCH        | Information on NCCH in English  |
| NETTINFO    | Information about network services                                      |
| PC          | Software for PC   |
| UNIX        | Software for Unix computers   |

**Coordinator for the electronic services:**

**Knut Hofland,**  
**KNUT@NAVF-EDB-H.UIB.NO, and**  
**FAFKH@NOBERGEN.EARN**  
**NAVFS edb-senter for humanistisk forskning**  
**Postboks 53 Universitetet**  
**N-5027 Bergen**  
**Telephone + 47 (0)5 21 29 54/55/56**  
**Telefax + 47 (0)5 32 26 56**

***MOZART OG BEETHOVEN I HYPERCARD***

---

***MOZART AND BEETHOVEN IN HYPERCARD***

---

CDs may be controlled by a computer in the same way as a video disc. *Warner New Media* and *The Voyager Company* have published a number of products combining CD-sound, HyperCard computer displays



and Macintosh sound. NCCH has aquired the first of these products from each publishers: Mozart's *The Magic Flute* and Beethoven's 9th Symphony. The products naturally share a good many features in common, but there are also differences in the way the material is presented.

*The Magic Flute* comes on three CD-ROMs and a diskette. The CDs contain the installation program and HyperCard stacks, and extra sound tracks used for musical examples for the running commentary that accompanies the opera. Besides there are the MIDI-data for those who have CD-players with MIDI output, and there are graphics for CD+G-players. The graphics can be shown on a television screen connected to the CD player, but to date the author has not seen a CD+G-player and has no idea of how this works. The HyperCard stacks (both those on the CDs and the separate one from the diskette, also have built-in control codes for a video disc player, which may be used for connection to the video disc version of Bergman's filmed *Magic Flute*.

The installation and use of this product appears an unnecessary cumbersome operation: You need to copy about 15 MB of HyperCard stacks from the CD's on to your hard disc – less if you don't wish to play the whole opera at once. If you have the storage space, fine, so far. However, you also have to use a specific home stack which contains the necessary drivers for the CD-player. The directions suggest that you give your original home stack a new name and call the one that comes with *The Magic Flute* "Home". A better solution is to copy the HyperCard program into the folder where the rest of the Magic Flute stacks are and use it with the dedicated home stack there. However, if you have several copies of HyperCard on the disc, you have to be careful in opening stacks by double-clicking on them. You have no guarantee which version of HyperCard will be started....

The author finds that even this is a problem he can live with, but he does wish HyperCard 2.0 had existed when the *Magic Flute* stacks were written, so that there would be no need for a dedicated home stack.

Once the program is started, you may choose between four different sets of running commentary: English or German libretto, description of the action or music commentary. The options are available while the opera is playing, too. While the CD is playing, buttons appear indicating hypertextual links to more background information. Since the CDs are specially made, commentary and other music examples are added, including the dubious pleasure of Florence Foster Jenkins attempting the dramatic heights of the aria of the Queen of the Night, a recording from a 1940's concert in Carnegie Hall.

Beethoven's 9th symphony: this product consists of one ordinary CD



and two HyperCard stacks. In this case there is no need for a special home stack, but the author reports certain problems with the fonts and in playing the recorded digitized sound, when he tried to convert the HyperCard stacks to version 2. He found it better to keep a version 1.2.5. on his hard disc.

In *The Magic Flute* the material is presented such that you are supposed to listen to the music in proper sequential order, and then make use of the hypertextual links when there are interesting new routes to follow. This is only one of the options in Beethoven's 9th Symphony – the option "A close reading", which incidentally has a user interface that functions very much like that of Mozart's *Magic Flute*. The other options are reading an essay on Beethoven's life and music, or another essay on music theory and an analysis of the 9th Symphony.

The program surrounding Beethoven's 9th Symphony is largely created by one man alone – musicologist Robert Winter, UCLA. The author finds that this presentation is simpler, more direct and more effective than *The Magic Flute*, which is created by a team.

Now, how useful are these for musical education? The author admits to being an amateur, a music-lover who has enjoyed both products, and feels certain that his music teachers in Upper Secondary school would have found these programs excellent teaching aids.

### ***SEMINAR OM INFORMASJONSTEKNOLOGI I DE HUMANISTISKE FAG – TENDENSER OG UTVIKLINGSBEHOV***

---

### ***INFORMATION TECHNOLOGY FOR THE HUMANITIES: TRENDS AND NEEDS. SEMINAR IN BERGEN 5 OCTOBER 1990***

---

This seminar was part of the ongoing evaluation of the Norwegian Computing Centre for the Humanities (NCCH). Director Hauge welcomed the ca thirty participants and expressed the hope that the seminar would lead to a clearer understanding of the situation for computing in the humanities in Norway as well as of the needs for services from NCCH. The participants came from the computer sections and various departments of the arts faculties of the four Norwegian universities, and had very different ways of understanding the situation of computing in the humanities.

Computing in the humanities may be in its infancy: there is still a certain inertia among teachers and researchers when it comes to making

use of computer methods, but as the equipment becomes faster and more powerful, as well as cheaper, smaller and lighter, more and more people have started using computers for other jobs than just word processing. Cost is still an obstacle, since these milieus traditionally have demanded little in the way of equipment. However, one of the participants was convinced that the various departments and subjects would soon demand modern laboratories for research and teaching, and that the need for equipment would increase rapidly.

A number of general questions relating to computing in the humanities were raised: How far can humanists influence the development of machinery and software? Traditionally the complaints have concerned the lack of funding; now it appears to be a larger problem that humanists have no say in the development or definition of standards. The development is driven by commercial interests, and teachers and researchers in the humanities have to adapt the available equipment to their needs, rather than participate in developing humanist systems. The universities and the NCCH were challenged to address this problem.

Another problem is recruiting. It was considered important that NCCH make its existence more strongly felt, for instance by demonstrating machinery and software to stimulate interest for making use of the technology in research and education. But both students and staff need "crash courses" in computing. There are students who are interested in using computer methods for their theses, and professors who are reluctant to accept new methods simply because they don't know enough about them. The average staff member at the universities has a computer which he uses for word processing and very little beyond that. He makes his transparencies by hand and knows no program for computer assisted teaching in his field. This person needs to be made aware of the possibilities of computing. Another speaker felt that the students were overrated: many were complete illiterates in computing. They would need education and training in computer use, too, and NCCH was in a position to do something for both groups.

Several of the participants felt that NCCH ought to become a national competence center within multimedia (including hypertext/hypermedia/interactive video) and picture processing. It was also suggested that NCCH do more within the field of sound and speech processing. On the other hand, pedagogy might not be a primary field for NCCH work.

There was general agreement about the role of NCCH in information dissemination. Several participants were interested to learn that NCCH has started a more active use of electronic information services. More people now have the necessary equipment and are in the position to make use of these services, which will make it possible to mail queries and answers to "public" electronic mailing lists.

It was also suggested that scholars should be given the opportunity to work at NCCH for a period, to learn about computing and bring the competence back to their work places.

Some people felt NCCH should do more to develop products and services, such as digital maps, data bases, bibliographic bases and Norwegian computer texts.

In the years to come, the humanities will develop a need for machinery and equipment comparable to that of the natural sciences. The students must also be included in this development: they will need education and access to equipment. One participant wanted development work done on a "history work station", where the researcher could work with different types of data and different types of software – better software than what is currently on the market. Hypermedia programs might do part of the job for the historian, provided the software for material analysis is improved. There should be systems that will tolerate a high degree of deviation: today's systems are not sufficiently tolerant for historical material.

Language teaching is a field where it is possible to foresee fundamental changes due to the new computer technology. One of the participants believed that within a ten-year period working methods and evaluation forms will have changed radically, and liberated teachers from many routine tasks. Perhaps it will be possible for students simply to hand in their exam "papers" on diskettes and for the teacher to have the machine suggest a preliminary evaluation.

The old-fashioned language laboratories are another problem: a student may well spend a lesson drilling the wrong pronunciation, because the teacher still only has minutes to spend on each student. Software is available that will check and correct the pronunciation of the student in comparison with the voice of the teacher or a native speaker, and there are also systems that will give students who are hard of hearing or deaf a visible indication of whether vowels are pronounced correctly.

One person felt that a "good NCCH" would be a type of interface where competence in computing met with competence in the humanities. Today, NCCH is too heavy on computing and too light on competence in the humanities.

Several people wanted to define the role of NCCH as that of the "locomotive" which would develop expert competence within a field and then pass the competence on to others, such as the computing sections at the arts faculties of the universities, and move on to other fields.

In conclusion, director Hauge said it might well be necessary for NCCH to reduce the number of current tasks, in order to strengthen the best parts of the activity.

## *SEMINAR OM LENKING AV NOMINATIVE DATA*

---

### *SEMINAR ON HISTORICAL RECORD LINKAGE*

---

This seminar was arranged at the National Archives in Oslo 29–31 October 1990, for 16 invited participants.

Nominative, historical data are source data, such as censuses, church registers and emigration protocols where individuals may be identified by name. Linking means to combine information from different sources in order to reconstruct the life of an individual or to find out what happened to a group of people.

The seminar took its starting point in work done in Norway, but the main purpose was to give the participants new impulses and information about the work that is done abroad. Special guests were Hans Christian Johansen, Dept. of History, Odense University and Gérard Bouchard, director of SOREP, Centre Universitaire de Recherches sur les Populations, Université du Québec à Chicoutimi. SOREP has developed advanced techniques for automatic linkage.

One of the problems involved in linking information from various sources, such as a church register and a census, is the unexpectedly high degree of error in these records: The name of one person may be given differently in the various records, or in different entries in the same record. One of the seminar participants gave examples from her own work: The name "Hansdatter" was entered as "Monsdatter", a girl's name, "Johanne", had been entered instead of "Johannes", the boy's actual name; the wrong person was registered as buried; ages were given wrongly; individuals or even whole families were entered twice in the same census.

Linkage may be carried out manually, and historians have long been doing that. Linking by the use of computers is a more recent option, with two possible strategies: interactive linkage in which the computer analyses the material and suggests links, while the operator makes the decisions about which links are to be established. Interactive linkage is therefore more akin to manual linkage, sharing most of its strengths and weaknesses. Automatic linkage means that the computer makes the links on the basis of a set of criteria.

Errors are likely to occur both in automatic and manual linkage. In general there are two types of errors. Overlinking means to establish more links than the material contains: information about several people may be put together as relating to one and the same person, creating inauthentic lives. Underlinking means that links that should have been established are left out, and the representational value of the material is not fully appreciated.

Linkage as a method has become more interesting in recent years, because historians now take more interest in the individual. The alternative to linkage is to analyze the source material at a higher level of abstraction, which is still a common research method. But linkage is also a source of information in other fields of study: epidemiologists and geneticists are also concerned with the lives of individuals and families.

In automatic linkage one establishes a set of criteria for the linkage process and lets the computer carry out the work: a number of fields may be selected, as name, age and sex, and the program will compare and evaluate the similarity between record entries. The result is given as a score – a high score most likely indicating that the entries in fact refer to the same individual. It is important to establish a common cutoff-point, a standard for what is deemed a legitimate link between record entries.

In order not to have to compare all possible connections between two sources, the material may be organized in advance, say by grouping name variations into standardized forms. The article gives examples of such standardization: names as different as Ambret, Engebret, Ingebret and Ingebrikt may be grouped as one name, because studies have shown that the actual variations in entries of the same name are at least this extreme.

Bouchard recommended material of a certain size, in order to optimize linkage projects. It was best to work with entire populations, counties or regions. Smaller areas, like parishes, can provide a suitable basis, but people who move will create problems for such limited investigations, so that he suggested that one at least worked on several parishes in order to improve the value of the linkage project.

The advantage of automatic over manual linkage is the fact that the criteria will not change in the course of the project. With manual linking it is not always possible to work with uniform consistency over a long period of time – the criteria for similarity may change.

For a check on the results of the automatic linkage, it may be advisable to reserve certain fields of information, such as age at marriage and death, for use against the findings. Bringing in too many criteria in the original linking will probably result in underlinking, i.e. reducing the representative value of the material.

The seminar showed that it is possible to develop computer systems that will make good use of historical material, but building a good, practicable system, like the Canadian one at SOREP is expensive. The best way is to bring in other fields of research as early as possible – epidemiology and genetics, in order to secure funds.

## ***ONLINE 90, LONDON, 11-13 DECEMBER 1990***

---

The **14th International Online Information Meeting** was arranged by **Learned Information** in London, 11–13 December 1990. The conference drew some 2,400 participants, and the exhibition was seen by almost 8,000 people. The conference itself was divided into parallel sessions for English, French, German, Italian, Dutch and Spanish language users.

One of the main points of the conference was the relationship of online data bases to optical storage media: D. Raitt opened the conference by stating that online is no longer what it once was – because so many of the online data bases now are available on CD-ROM. One speaker discussed the practical features of the ideal CD-ROM work station of the 1990's.

Another part of the online data base industry's relationship to the market is a new demand for quality. There is now more criticism and an increasing demand for a dialogue between users and producers, as an indication also of the importance of the online industry.

One speaker criticized the user interface, and suggested that a few thought-out changes might make the data bases far more accessible to the users. His conclusion was that the online tool has not been good enough.

## ***INTERNASJONALT SEMINAR OM FORSKNINGSDOKUMENTASJON***

---

### ***INTERNATIONAL SEMINAR ON RESEARCH DOCUMENTATION***

---

This seminar in Bergen 7–9 February 1991 was arranged by the **Norwegian Computing Centre for the Humanities and the Committee for Reseach Documentation** under the Joint Board of the Norwegian Research Councils.

Research documentation is becoming increasingly more important, also as a field of international cooperation. At the seminar in Bergen speakers from 11 European countries presented the various national approaches to the problem of reseach documentation – which varied from absolute minimum solutions to some very advanced and sophisticated systems.

The conference speakers discussed a wide range of topics: the purpose of research documentation; the function of such a service for the various user groups; targeting and defining the user groups; standards for documentation information; copyrights; user support.

One of the most interesting novelties presented at the conference was



the CERIF manual developed by the EEC as a suggestion for a common European format for research documentation. The work leading up to the production of the manual was presented by I.R. Perry from the Directorate General for Science, Research and Development under the Commission of the European Communities.

The point of such a manual is to standardize data bases so that it will be possible to make all European research documentation available throughout Europe. There are problems that must be solved before such international use of the data bases is possible, but many people are interested in establishing such international connections. For that reason many of the participants spoke warmly for a follow-up seminar in the not too distant future.

### ***SENTRAL FOR INFORMASJON OM FORSKNINGSPROSJEKTER***

---

### ***CENTRE FOR INFORMATION ON RESEARCH PROJECTS***

---

In 1990 the Committee for Research Documentation under the Joint Board of the Norwegian Research Councils decided to start a pilot project to transfer information on all research projects to one common data base. The entry of the data was completed in the autumn of 1990.

The administration of the data base is located at the Norwegian Computing Centre for the Humanities in Bergen, and the data base makes use of an IBM 4381 computer belonging to the University of Bergen. The base employs a search system for free text and is available 24 hours a day via DATEL, DATAPAK and UNINETT. Limited assistance is available between 8 a.m. and 10 p.m., and full assistance including search assistance is available between 8 a.m. and 4 p.m. In addition there is a detailed manual, so the availability is very good. All services are provided free of charge.

Data are transferred from all the Norwegian research councils, and even though there are differences in quality and the amount of information in the documents, there is still a common core of information.

By 1 January 1991 the data base contained information on a total of 25,278 projects. These projects cover research on virginity rituals in the South Pacific, choral singing, astral physics and protection of the environment. It is therefore possible to carry out interdisciplinary searches, and find areas of common interest to many fields of research. E.g., a search for *environment*, *algae* and *radiation* showed that these topics were represented in a number of research projects. This search, which took a couple of minutes, came up with 1,380 documents on environmental research. It would probably have been impossible to find

all this information in any other way than through an electronic data base.

### ***KOMITÉINNSTILLING OM NY TEKNOLOGI PÅ HF-FAKULTETET VED UiB***

---

#### ***COMMITTEE REPORT: NEW TECHNOLOGY TO BE IMPLEMENTED AT THE FACULTY OF ARTS AT THE UNIVERSITY OF BERGEN***

---

A committee established by the Council for The Faculty of Arts at the University of Bergen presented its report in December 1990. The committee was to propose a plan for the future use of computers in the Arts Faculty and the organisation and function of the Faculty's Computing Section.

The committee's main proposals are,

1. Computing in the Arts Faculty should not be limited to a separate section.
2. The faculty needs computer expertise separated from the individual departments.
3. The Computing Section should remain as a service and training section and should at all times relate its activities directly to the demands of the users.
4. The Computing Section of the Arts Faculty should not develop competence available outside of the institution.

The committee also suggests that the Computing Section should continue to give courses in computing in the humanities, and that it should take over responsibility for the faculty's language laboratories.

There is also a discussion of the responsibilities of the Computing Section in relation to the Norwegian Computing Centre for the Humanities. The committee sees possibilities for a division of labour between the two institutions. The Computing Section should seek to cooperate with NCCH in fields where NCCH has expertise, such as in multimedia.

### ***DISKUSJONSSEMINAR INTERAKTIV VIDEO I SPRÅKOPPLÆRING***

---

#### ***DISCUSSION OF INTERACTIVE VIDEO IN LANGUAGE TEACHING***

---

On Friday 14 December 1990 Researcher/NAVF Signe Marie Sanne held a discussion seminar on the use of interactive video in language teaching



at Høyteknologisenteret in Bergen. About twenty college and upper secondary school teachers participated.

The starting point for the discussion was a presentation of Sanne's project in interactive video for Italian language teaching. This project has created an integrated work station where the student is primarily intended to work on his own. A fellow pedagogue, *Ingrid Hatlestad*, was specially invited to present her computer program which enables the teacher to check what use the student makes of the system.

The discussion was concerned with the pedagogical uses of this teaching system, whether the teacher should concentrate on the practical use of the language or move on to extralinguistic factors, cultural elements, etc.

Another question was concerned with the possibility for working out exercises to be used in connection with the material on the video record, multiple choice exercises, etc. Basically it is only the teacher's imagination that limits the number of possible computer assisted learning tricks that may be appended to this material.

Then there was the question of whether the program should record the student's voice for comparison with a model pronunciation. Technically this presents no problem at all.

The participants in the seminar saw the advantages of giving the students easy access to a great deal of material. They felt it was more likely that a student would look up a word in the dictionary on the screen than on the desk.

## ***LESERUNDERSØKELSE OM HUMANISTISKE DATA***

---

### ***READER SURVEY ON HUMANISTISKE DATA***

---

Last autumn we asked our 474 personal subscribers how they felt about Humanistiske Data. Unfortunately only 114 completed forms were returned, and that percentage is a little too low to make the answers fully representative. However, there are also other reasons why it is problematic to draw any substantial conclusions from the answers. The answers indicate that the readers represent a wide variety of subjects and interests, and that they would like to see more material in Humanistiske Data from *their* field.

One of the questions was:

*Which topics and types of material are you most interested in?*

| Choices:   | No of answers | Percentage of answers |
|--|---------------|-----------------------|
| Computing in the Humanities in Scandinavia             | 69            | 63%                   |
| Theoretical articles about computing in the humanities | 65            | 60%                   |
| Presentations of projects                              | 62            | 57%                   |
| Reviews of software                                    | 59            | 54%                   |
| Computing in the Humanities in the rest of the world   | 50            | 46%                   |
| Computing in education                                 | 46            | 42%                   |
| New products   | 45            | 41%                   |
| Computing in cultural and research dissemination       | 38            | 35%                   |
| Book reviews   | 35            | 32%                   |
| Optical storage media                                  | 33            | 30%                   |
| Multimedia   | 33            | 30%                   |
| Work at NCCH   | 30            | 28%                   |
| Technical articles                                     | 21            | 19%                   |
| Computing in library work                              | 11            | 10%                   |

The five most interesting categories are unspecific – all the specific categories receive gradually diminishing scores. A good many of those who returned the completed form wanted more material on specific subjects. This seems to indicate that there is now more need for specific solutions than for general ones in computing in the humanities.

We interpret the answers thus that we should continue to report on a wide variety of subjects. We will continue to present reviews of machines and software, and reports of projects and arrangements. We will try to bring material on the subjects on the "wanted" list.

BUT THEN THERE WAS THE FORM in which we should report.

Obviously, the readers' evaluation of Humanistiske Data is mixed. Some say the periodical is good, some say its boring, with old-fashioned lay-out and the wrong typography. Many ask for debates, and want the editorial board to evaluate programs and systems instead of just presenting the novelties uncritically. Some people write that technical descriptions are too detailed and inaccessible beyond the initiates.

However, many of our readers characterize Humanistiske Data as an outstanding periodical – in Scandinavia. We have few readers outside of Norway, but a high percentage of these have taken the opportunity to express the view that Humanistiske Data provides information which is not available through any other channel.

Obviously, we cannot satisfy everyone. However, we've heard it said that the biggest room in the world is the room for improvement, and we will start by changing the layout of the magazine. We also aim to publish four issues of Humanistiske Data per year. As far as debates go, it's very much up to our readers to use the periodical for debating purposes.

We wish to thank all of those who took the time to answer the questionnaire.



## SENTERETS RAPPORTSERIE

### RAPPORTER UTGITT F.O.M. 1987

RAPPORT NR. 41. *Tove Fjeldvig og Anne Golden: Automatisk gjenkjenning og splitting av sammensatte ord.* Desember 1987. ISBN 82-7283-049-3 Pris kr. 55.

RAPPORT NR. 43. *NKKMs Datafaglige Råd: Rapport fra studiereise til USA, Canada og Japan. Mars 1988.* NAVFs edb-senter for humanistisk forskning. Oktober 1988. ISBN 82-7283-051-5 Pris kr. 55.

RAPPORT NR. 44. *Claus Huitfeldt and Viggo Rossvær: The Norwegian Wittgenstein Project Report 1988.* October 1989. ISBN 82-7283-052-3 Price NOK 150.

RAPPORT NR. 46. *Claus Huitfeldt: Ny teknologi til bevaring og formidling av primærkilder.* Oktober 1989. ISBN 82-7283-055-8 Pris kr. 150.

RAPPORT NR. 47. *Årsmelding 1989.* NAVFs edb-senter for humanistisk forskning. ISBN 82-7283-056-6 Gratis.

RAPPORT NR. 48. *Konferanserapport "Edb & kunstfag", Bergen.* September 1990. ISBN 82-7283-057-4 Pris kr. 100.

RAPPORT NR. 49. *OCR - teknologi for innlesing av humanistisk kildemateriale. Status for optisk tegngjenkjenning i dag.* September 1990. ISBN 82-7283-058-2 Pris kr. 60.

RAPPORT NR. 50. *Signe Marie Sanne: AskSam - et databaseprogram med mye attåt.* September 1990. ISBN 82-7283-059-0 Pris kr. 60.

RAPPORT NR. 51. *Elektroniske Bøker - Multimedia oppslagsverk, Bergen 21. - 23. november 1990.* NAVFs edb-senter for humanistisk forskning. Mars 1991. ISBN 82-7283-4 Pris kr. 175.

# C

ISSN 0800-6792

## I DETTE NUMMER:

- *Multimedia*
- *Bildebehandling*
- *Senterets tjenester*



NAVF'S EDB-SENTER  
FOR HUMANISTISK FORSKNING  
NORWEGIAN COMPUTING CENTRE  
FOR THE HUMANITIES

**RETURADRESSE:**

NAVF's edb-senter for humanistisk forskning, Boks 53 – Universitetet,  
N-5027 Bergen