



Nyhedsbrev marts 2007

Velkommen til Netarkivets nyhedsbrev for marts 2007.

Netarkivet hjælper Esbjerg Kommunes Biblioteker

I januar fik netarkivet en henvendelse fra Esbjerg Kommunes Biblioteker. De havde i forbindelse med kommunesammenlægningen omlagt bibliotekets netsted, og i den proces var der forsvundet nogle sider. De ville høre om netarkivet havde arkiveret deres gamle netsted, og om de eventuelt kunne få en kopi af det mistede tilbage.

Sidste tværsnitshøstning af www.esbbib.dk var fra medio oktober. De relevante data blev kopieret fra arkivet og leveret på cd-rom tilbage til en meget glad webmaster.

Netarkivet får tilknyttet medieforsker

Statsbiblioteket har januar 2007 ansat en medieforsker (Ditte Laursen). Hun er blevet tilknyttet Statsbibliotekets faglige arbejde med netarkivet og især opgaven med at udvælge og definere de 80 netsteder til den såkaldte selektive høstning. Det er dejligt med noget kvalificeret input og sparring. Velkommen til Ditte.

De første forskere får adgang til netarkivets data

Netarkivet har i starten af 2007 fået foden indenfor i to århusianske forskningsprojekter.

Et forskningsprojekt med base i Århus skal kigge på DR og TV2s opfyldelse af deres public service forpligtelser. I den forbindelse har de udvalgt uge fire og uge 12 som analyseuger. Netarkivet indsamler i de to uger lidt mere end normalt fra disse netsteder.

Forskerne vil naturligvis gerne have adgang til arkivet. Da netarkivet endnu ikke har fået implementeret et slutbruger adgangssystem, bliver det en meget basal og simpel adgang, hvor forskerne kan vælge en bestemt høstning og så browse rundt i den – der er derfor f.eks. ingen søgemuligheder.

Et andet forskningsprojekt, ledet af vores gamle kending Niels Brügger, skal se på, hvorledes materiale, der måtte være arkiveret andre steder (fx. privat eller af producenten selv), kan indlemmes i netarkivets samling. Det er et forprojekt til et større projekt fra efteråret 2007, hvor Niels Brügger skal skrive dr.dk's historie fra 1996-2006. Ud over at se på, hvordan det eksterne materiale kan indlemmes i netarkivets samling, skal kvaliteten af forskellige mindre samlinger af materiale fra dr.dk sammenlignes med kvaliteten af netarkivets høstninger af samme. Det bliver spændende at få resultatet af den undersøgelse.

Netarkivets system i open source

Udviklerne på Statsbiblioteket og Det Kongelige Bibliotek arbejder på højtryk med at klargøre netarkivets system til udgivelse i Open Source. Det er planen, at det bliver sendt ud 1. juli i år under den såkaldte LGPL software-licens.

Der har været vist stor interesse fra flere af de nordiske lande samt England, Frankrig, Tyskland, Tjekkiet og Østrig, så forhåbentlig vil der med tiden blive opbygget et community omkring systemet, der sikrer, at den videre udvikling og vedligeholdelse bliver på et højere niveau, end hvis vi sad som eneste brugere af systemet.

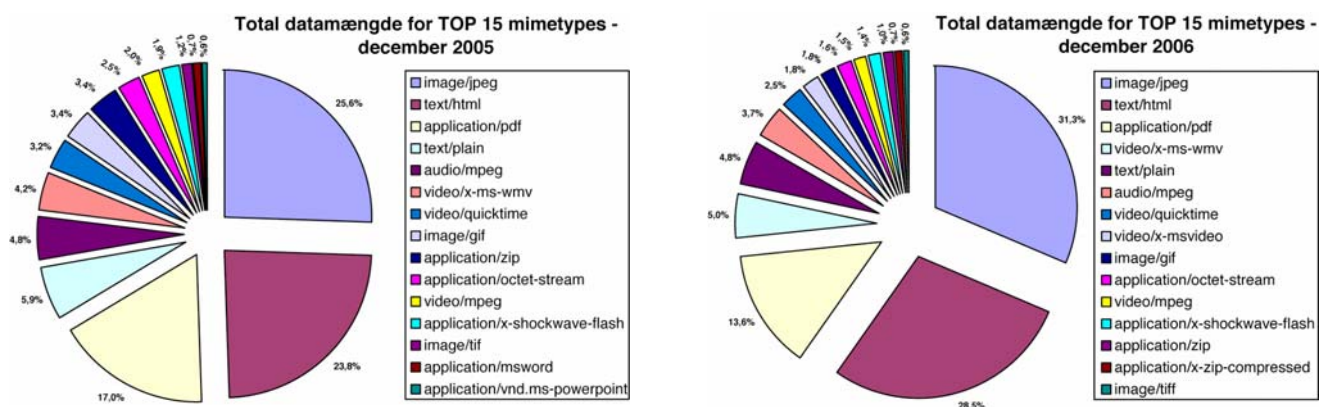
IIPC møde i Paris

18-20. april afholdes det første stormøde i den nye IIPC organisation, hvor såvel gamle som nye medlemslande/-organisationer vil være tilstede.

Bjarne Andersen er blevet indstillet som medlem til TC (Technical Committee), der skal fungere som faglig styregruppe for hele konsortiet. IIPC er uden tvivl et meget spændende og indflydelsesrigt sted at være med i, så vi afventer med spænding valget.

Udviklingen af .dk domænet

Efter netarkivets tredje tværsnitshøstning har vi genereret en del statistik. Blandt de mere interessante er statistikken over, hvordan de forskellige filtyper fordeler sig på antal, og den udvikling der er sket fra 2005 til 2006.



Mest markant er det, at jpeg-billeder er gået signifikant frem i den totale datamængde fra godt 25% i 2005 til over 31% i 2006. Desuden tegner der sig et klart billede af, at video er på vej i samme retning. Video er nu både på en fjerdeplads på TOP 15 samt repræsenteret med fire filtyper på TOP 15 i 2006 mod kun tre i 2005.

Styregruppemøde 29. januar 2007

Styregruppen for netarkivet havde styregruppemøde på Det Kongelige Bibliotek den 29. januar. Her blev det bl.a. besluttet, at netarkivet medio 2007 skal lave en redegørelse til Kulturministeriet omkring indsamlingsstrategi kontra lagerpladsforbrug og budget. Diskpriserne ser ikke ud til at falde så meget som forventet, så det betyder i praksis, at vi ikke har råd til fire årlige tværsnitshøstninger inden for det nuværende budget.

Så enten skal der flere penge til, eller også skal vi til at kigge på at lægge strategien mere permanent om, således at vi ikke laver fire årlige tværsnitshøstninger men måske begynder at differentiere, således at nogle netsteder måske kun høstes én gang årligt, mens andre høstes tre eller fire gange årligt, baseret på opdateringsfrekvensen.

Danske domæner uden for .dk

I perioden op til den netop igangsatte tværsnitshøstning arbejdede pligt på Det Kongelige Bibliotek med identifikation af danske domæner uden for .dk (som beskrevet i sidste nyhedsbrev). Det resulterede i, at der blev lagt knap 38.000 nye domæner i systemet i januar 2007. Disse bliver i øjeblikket høstet for første gang.

Styregruppen besluttede på styregruppemødet 29. januar, at domæner ikke inkluderes i netarkivets indsamling, hvis de pågældende domæner indsamles andetsteds af en lignende national organisation. Det betyder indtil videre, at vi har udeladt .se (Sverige) og .no (Norge), idet begge lande selv indsamler det komplette domæne.

Fjerde tværsnitshøstning

Den fjerde tværsnitshøstning blev igangsat 1. marts. Den inkluderer nu ca. 803.000 domæner, hvoraf det forventes, at ca. 80% er aktive – dvs. netarkivet indsamler ca. 640.000 danske hjemmesider.

I denne tværsnitshøstning er der nu 2548 domæner, der har fået godkendt en øvre grænse på 1 GByte, og 173 domæner der har en øvre grænse på 2 GBytes.

Netarkivet dokumenterede strukturreformen

Netarkivet lavede umiddelbart før jul en ekstra indsamling af de netsteder, der var i fare for at forsvinde eller undergå kraftige forandringer som følge af strukturreformen. Det blev til en indsamling af i alt 383 domæner – primært de kommunale og amtskommunale samt en række sammenlægningsnetsteder.

Status på bitarkiverne

Bitarkiverne rummer nu følgende datamængder:

Tværsnitshøstning: 26.141 GBytes i 269.543 filer

Selektiv høstning: 3.466 GBytes i 40.985 filer

Begivenhedshøstning: 3.325 GBytes i 35.028 filer

De tre bitarkivmaskiner, der blev omtalt i sidste nyhedsbrev, har været en tur i Århus for at blive genopfyldt med de manglende data. De er ankommet til København og er ved at blive sluttet til systemet igen. Derefter vil en komplet tjeksumskontrol af begge bitarkiver forhåbentlig vise, at netarkivet igen har 100% konsistens mellem de to online-kopier på henholdsvis Statsbiblioteket og Det Kongelige Bibliotek.

Planer for den kommende periode

1. Redaktionsgruppemøde 22. marts
2. IIPC-møde i april
3. Fortsat omlægning af de selektive høstninger
4. Implementering af DS-484 i forhold til netarkivets data