



Nyhedsbrev august 2009

Velkommen til Netarkivets nyhedsbrev.

Status på bitarkiverne

De nye bitarkiv-maskiner på Det Kongelige Bibliotek begynder at være på plads. Der er nu 7 nye, kraftige maskiner oppe at køre med overførte data. De kører fint - som forventet.

De resterende 5 nye maskiner er planlagt til at blive sat op løbende - senest klare til oktober. I takt med at data bliver overført til de nye maskiner bliver de gamle maskiner slukket og skrottet.

Der vil med de nye maskiner være ledig kapacitet år 2009 ud. I øjeblikket fylder Netarkivet knap 120 TB og med de nye maskiner optimalt udnyttet burde vi have ca. 180 TB totalt.

Bitarkivet på Statsbiblioteket er fortsat to kraftige maskiner, der løbende bliver tilført nye diske.

Ottende tværsnitshøstning

Denne seneste tværsnitshøstning er påbegyndt d. 9. juli 2009.

Efter planen er der forventninger om følgende tidsforbrug:

- 2x10 dage til indeksering
- step-1 skal køre på ca. 8 dage
- step-2 kan køre på ca. 2 måneder

Det er ret præcist 3 måneder tilsammen, og vi kan starte næste tværsnitshøstning op umiddelbart (1-3 dage) efter en afsluttet tværsnitshøstning. Men der er ikke plads til problemer i denne tidsplan.

Og problemer har vi dog stadig indimellem. Det var ikke kun hardwareproblemer der kunne forsinke planen, men også software, firewall, organisatoriske forhold som prioritering, ferie, sygdom osv.

I denne tværsnitshøstning startede vi med at have problemer med at generere indekser. Derfor var det nødvendigt at køre indeksgenereringen om. Nødløsninger blev sat ind og bugs blev rapporteret og fejlhåndteret, men forsinkelsen var uundgåelig. Vores tiltag sikrer dog, at vi i hvert fald ikke får samme fejl ved næste tværsnitshøstning.

Da tværsnitshøstninger erfaringsmæssigt vokser ca. 20 % for hver høstning, forventer vi denne tværsnitshøstning på ca. 26-28 TB. Dertil kommer diskforbrug for de løbende høstninger af selektive og begivenhedshøstninger.

Alt i alt betyder dette, at vi realistisk ikke når op på 4, men kun måske 3 tværsnitshøstninger pr år.

Vi har dog – i princippet - mulighed for at skruer op for høstningskapaciteten.

Der er stadig ledige høster kapacitet på de 2 nye høstermaskiner på KB samt en maskine, der slet ikke er taget i brug endnu, så der kan maskinemæssigt startes i alt 7 nye høsterapplikationer - ud over de nuværende 13 tværsnitshøstere. Dette vil forøge kapaciteten med ca. 50 %, men kræver at KB

godkender merforbrug af netværk. Det kræver igen at KB lægger om så de kan styre Netarkivets forbrug mere individuelt. Mulighederne for at øge høstningskapaciteten væsentligt undersøges nu.

Statistiske oplysninger er nu tilgængelige på <http://netarkivet.dk/tvaersnitsHostninger.html>.

Samlingsafdelingen på Det Kongelige Bibliotek står for arbejdet med tværsnitshøstninger.

Selektive høstninger

De selektive høstninger arkiverer - fra seks gange dagligt til månedligt - de godt 80 udvalgte netsteder. Listen af netsteder ændrer sig løbende.

Formålet med den selektive høstning er at indsamle sider, der ændrer sig hurtigere end tværsnitshøstningen kan nå at fange. Ifølge bemærkningerne til loven om pligtaflevering skal Netarkivet indsamle:

1. Netsteder, der fungerer som nyhedsmedie for den danske offentlighed. Hertil hører især de landsdækkende avisers og tv-stationers netsteder, de større regionale avisers netsteder, Folketingets netsted samt et mindre udvalg af nyhedsmedier
2. Netsteder, der dækker et repræsentativt udsnit af typiske netsteder, dvs. meget besøgte netsteder (fx portaler af typen netdoktor), og/eller netsteder, der repræsenterer typiske anvendelser (fx virksomheders netsteder, byportaler og kommunale portaler, netsteder, der anvendes som lokalt medie)
3. Netsteder, der er særligt karakteristiske, eksperimentelle eller unikke (fx åbne debatmiljøer, netfællesskaber, personlige sider, netkunststeder) i forhold til indhold og udtryksformer.

Samlingsafdelingen på Statsbiblioteket står for arbejdet med de selektive høstninger.

Begivenhedshøstninger

Begivenhedshøstninger planlagt i 2009:

- EU-parlamentsvalg i juni – er høstet og afsluttet
- Kommunalvalg 2009 i november
- Klimakonference i København i december. Høstningen er allerede gjort klart og høstninger er begyndt - dog med lav frekvens. Alle relevante danske og engelsksprogede sider fra .dk ønskes høstet, fx COP15.org og unfcc.int. Blog kan høstes sent, da de er kumulative.

Ikke-planlagte begivenheder er:

- Influenza-pandemien
- Begivenheden omkring irakere i Vor Frue kirke og Brorsons Kirke

Det er relative små datamængder der høstes.

Samlingsafdelingerne på Det Kongelige Bibliotek og Statsbiblioteket koordinerer sammen arbejdet med begivenhedshøstningerne.

Internationalt samarbejde og konferencer

IIPC (International Internet Preservation Consortium) bliver afholdt i San Francisco, USA d. 7. okt. 2009.

ECDL-konference / IAWAW konference afholdes på Korfu, Grækenland fra 27.sept. til 2. okt. 2009.

Netarkivet deltager begge steder.

NetarchiveSuite

Dagene 27-29. maj 2009 var afsat til en workshop på Det Kongelige Bibliotek om NetarchiveSuite- og høster-erfaringer. Her deltog 5 udviklere/kuratorer fra ONB (Østrigs nationalbibliotek) og BnF (Frankrigs nationalbibliotek) sammen med udviklere og samlingsfolk fra Netarkivet.

Målet er et mere internationalt samarbejde omkring udvikling af NetarchiveSuite, således som det var ønsket, da det blev gjort til et Open Source projekt.

Efterfølgende er samarbejdet i fuld gang, således at de udenlandske folk nu bidrager med blandt andet kode. Samarbejdet ser lovende ud.

Netarkivets NetarchiveSuite blev d. 25.6.2009 opgraderet i produktionsmiljøet til version 3.8.0 (stable release). Det viste sig at være en nyttig ”øvelse”, som gav nogle uventede driftsforstyrrelser. Disse er efterfølgende blevet udbedret og installationsfiler samt dokumentation er opdateret.

Redaktionsgruppemøde

Netarkivets redaktion skal holde møde igen til efteråret og Kulturministeriet skal beskikke nye medlemmer eller genudnævnelser til redaktionsgruppen.

Det tager noget ekstra tid at finde de nye ansigter, da Foreningen af Danske Interaktive Medier, FDIM først kan tage stilling til det ved bestyrelsesmøde medio september.

Claus Lomborg,
Netarkivet