



Nyhedsbrev februar 2006

Velkommen til Netarkivets nyhedsbrev februar 2006. Nyhedsbrevet er blevet en smule forsinket pga. almindelig travlhed, og styregruppen for netarkivet har vedtaget, at nyhedsbrevet fremover skal udkomme 6 gange årligt.

Siden november sidste år er der sket en masse, som dette nyhedsbrev meget gerne skulle bidrage til at informere om.

Nyt design til netarkivet.dk

Netarkivet fik sidste efterår lavet et nyt design, som også logoet på dette nyhedsbrev bærer præg af. Det gamle netsted er under portering til det nye design, og inden for ganske kort tid vil det nye design være tilgængeligt på <http://netarkivet.dk>, hvorefter de efterhånden utallige 'gamle' sites for projektet lukker endegyldigt ned.

Indtil da kan interesserede smugkigge på <http://sb-test-net-001.statsbiblioteket.dk> (kun internt).

Første begivenhedshøstning

I november 2005 gennemførte netarkivet den første begivenhedsindsamling omkring det kommunale valg. National- / pligtafleverings-afdelingerne identificerede mere end 2500 relevante netsteder, og især kandidaternes personlige kampagner var stærkt repræsenteret i forhold til tidligere indsamling af kommunalvalg (november 2001) – således havde over 1000 kandidater deres egen hjemmeside.

Indsamlingen forløb efter planen og strakte sig i praksis fra ca. 3 uger før valget til 2 uger efter valget. I alt blev der indsamlet 285 Gbytes repræsenteret af godt 11 millioner objekter.

Første tværsnitshøstning

Under afslutningen af den første tværsnitshøstning løb vi ind i nogle uheldsmæssigheder med unormal høj belastning af nogle få netsteder. På baggrund af disse problemer blev det besluttet at ændre logikken fra at sætte en maksimal grænse på antal objekter pr. domæne til at sætte en maksimal grænse på antallet af bytes, der må hentes. Dette krævede tekniske ændringer i såvel netarkivets administrative system som den høster-software, vi benytter. Det betød i praksis, at tværsnitshøstningen lå stille fra oktober og frem til midten af december 2005.

De nødvendige ændringer er nu implementeret, og det betyder, at netarkivet nu som udgangspunkt høster maksimalt 500 Mbytes pr domæne (4 gange årligt). Kun domæner, der bliver manuelt udvalgt og godkendt til en højere grænse, vil komme forbi denne grænse. Det er vores forventning, at

ca. 4-5000 netsteder potentielt er større end 500Mbytes grænsen, og disse skal altså studeres nærmere af den biblioteksfaglige gruppe, der er tilknyttet netarkivet.

Netarkivets erfaringer med den første tværsnitshøstning er blevet skrevet sammen til en artikel, der udkom i DF-revy i januar 2006. Så snart det nye design til netarkivets hjemmeside er færdigimplementeret, vil artiklen kunne hentes er i såvel en dansk som en engelsk udgave.

Anden tværsnitshøstning

Første skridt – med en grænse på 10 Mbytes pr. domæne - af den anden tværsnitshøstning er veloverstået. Det resulterede i godt 1Tbytes data og knap 49 millioner objekter.

Andet skridt med en grænse på 500 Mbytes er netop sat i gang på de ca. 61000 domæner, der ramte grænsen på 10 Mbytes.

Det betyder, at ca. 600.000 af de domæner, vi kender, blev høstet komplet ved første høstning og altså rummer mindre end 10Mbytes data.

Anden begivenhedshøstning

Netarkivet igangsatte den 8. februar (på opfordring) en begivenhedshøstning, der dækker den nationale krise omkring de berømte Muhammed-tegninger. Der høstes dagligt fra mere end 500 netsteder, der dækker krisen fra vidt forskellige vinkler, lige fra de oplagte medie-sites til private hjemmesider, der siger undskyld på både egne og Danmarks vegne. Desuden høstes udvalgte udenlandske mediers dækning af sagen i det omfang, de vurderes relevante for den danske kulturarv.

Digital bevaring

Såvel Statsbiblioteket som Det Kongelige Bibliotek er med i EU-projektet PLANETS, der skal kigge på både strategier og værktøjer til digital bevaring. Netarkivet forventes at kunne trække meget på resultaterne af dette projekt.

Efter planen igangsættes det inden sommeren 2006 og kører over 3 år.

Udviklingsopgaver

Udviklingsprojektet omkring netarkivet kører videre på naturligt nedsat blus i forhold til det store udviklingsprojekt. Der er stadig mange hjørner af systemet, der trænger til både fejlrettelse, optimering og nyudvikling.

Udviklingsmæssigt har der i den seneste periode været fokuseret dels på den tidligere nævnte bytegrænse pr. domæne, men også på aktiv bitbevaring og forbedringer af interfacets hastighed, hvilket er en stor lettelse i det daglige arbejde med systemet.

Status på bitarkiverne

Den aktive bitbevaring betyder, at man nu fra webinterfacet kan bede bitarkiverne på Statsbiblioteket henholdsvis Det Kongelige Bibliotek om en øjeblikkelig status på både antallet af filer samt

disse filers tjeksum (for at tjekke at filerne stadig er de samme på begge sider af storebælt, som da de blev lagt i arkiverne).

Det er med stor tilfredshed vi kan konstatere, at status i øjeblikket er, at der ikke mangler en eneste fil, og at ikke en af netarkivets filer er korrumpet frem til dags dato. Det samlede bitarkivs øjeblikkelige status er altså, at alt er i skønneste orden.

Bitarkiverne på SB og KB rummer pt. ca. 9.6 Tbytes data.

Virtuelle mødefaciliteter

Både SB og KB har investeret i virtuelle mødefaciliteter i form af videokonferenceudstyr. Det er blevet brugt af såvel udviklingsgrupperne, de biblioteksfaglige som styregruppen. Det er en fantastisk mulighed at kunne holde kortere møder sammen uden at skulle bruge 6 timer i et tog.

Det kan ikke afløse alle fysiske møder, men mange problemer kan klares via dette medie. Især har driftslederen, som i efteråret i en periode var på KB i snit hver anden uge, fået set noget mindre til DSB, siden disse faciliteter blev etableret.

Redaktionsgruppe

Kulturministeriet har nedsat en redaktionsgruppe for netarkivet. Redaktionsgruppen skal vejlede netarkivets fagligt mht. strategier, valg af begivenheder samt konkrete kandidater til særlig indsamling. Redaktionsgruppen har det første møde på Statsbiblioteket den 13. marts.

Redaktionsgruppens medlemmer er:

Niels Ole Finneman, Institut for Informations- og Medievidenskab, Århus Universitet

Ellen Christiansen, Institut for Kommunikation, Ålborg Universitet

Henrik Rewes, Nordjyske Medier på vegne af Foreningen af Danske Internetmedier

Karen Westman Hertz, DR Interaktiv - Foreningen af Danske Internetmedier (på vegne af IT-Brancheforeningen)

Henrik Borberg, Forlaget Multivers Aps på vegne af Den Danske Forlæggerforening

Planer for den kommende periode

1. Flytning af indholdet på det nuværende netarkivet.dk til det nye design
2. Løbende overdragelse af det daglige arbejde med systemet til national-/pligtafleveringsafdelingerne
3. Krav til "Trusted Repositories"
4. IIPC-møde i København i marts
5. Afslutning af den anden tværsnitshøstning