



Nyhedsbrev maj 2009

Velkommen til Netarkivets nyhedsbrev.

De sidste måneder har stået arkitekturens tegn – hardware arkitektur vel at mærke. Det har nemlig været en lang og sej proces at få valgt og implementeret en ny løsning til noget professionelt isenkram til bitarkivet på Det kongelige Bibliotek. Men nu er de første nye maskiner heldigvis kørende og viser gode takter.

Forhåbentligt når vi op på 3 eller 4 tværsnitshøstninger om året med de nye bitarkivmaskiner på Det Kongelige Bibliotek.

Der er lagt op til nogle interessante begivenhedshøstninger og der skal sikkert komme flere til som vi endnu ikke kender.

Syvende tværsnitshøstning

Denne tværsnitshøstning blev påbegyndt d. 16. februar 2009. DR var inviteret til at være med og lave et TV-indslag om Netarkivets start af en succesfyldt tværsnitshøstning. TV indslaget blev bragt i DR1 nyhederne kl. 21.21 d. 16.4.2009 og varede ca. 2½ minut. Et af budskaberne var, at arkivet først bliver åbnet om mange år, men at der trods det, er personer der synes at det er spændende alligevel. I indslaget fulgte man en forældre der havde lagt billeder af sine børn på nettet.

Samlingsafdelingen på De kongelige Bibliotek sikrer kvaliteten af tværsnitshøstningerne.

Selektive høstninger

De selektive høstninger arkiverer løbende (fra seks gange dagligt til månedligt) de ca. 80 udvalgte netsteder. Pt. er antallet lidt højere men det ændrer sig løbende.

Samlingsafdelingen på Statsbiblioteket arbejder løbende på at kvalitetssikre det indsamlede og ajourføre puljen så nye, og mere interessante aktive netsteder kan blive høstet i stedet for.

Begivenhedshøstninger

Den særskilte høstning omkring den finansielle/økonomiske krise viste kun få opdateringer. Derfor er den nu blevet stoppet.

Den selektive høstning af det nye parti, borgerligtCentrum.dk, blev også startet op da der var udsigt til ændringer i det politiske billede. Heller ikke dette blev til nogen stor høstning.

I øjeblikket er det følgende begivenhedshøstninger der er planlagt i 2009 – og delvist startet op allerede:

- EU-parlamentsvalg i juni
- Kommunalvalg2009. Her kan tages udgangspunkt i hvordan vi gjorde omkring kommunalvalget 2005.
- Klimakonference i København i december. Høstningen er allerede gjort klart og den forventes at blive relativ stor. Alle relevante danske og engelsk sprogede sider fra .dk ønskes høstet, fx COP15.org og unfcc.int. Blog kan høstes sent, da de er kumulative.

Allersenest har begivenhederne omkring svineinfluenzaen givet overvejelser, men der har vist sig ikke at være særligt meget at høste samtidigt med at influenzaen viste sig ikke at være helt så farligt som først forventet.

Begivenheden omkring irakere i Vor Frue kirke bliver ligeledes overvåget.

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

Status på bitarkiverne

Bitarkiverne rummer nu (tal fra 15/5/2009) følgende datamængder:

- Tværsnitshøstning: 78498 Gb
- Selektiv høstning: 11311 Gb
- Begivenhedshøstning: 6415 Gb

Totalt har Netarkivet altså arkiveret 96 Tb data.

Den nye bitarkiv platform på Det kongelige Bibliotek er baseret på professionelle maskiner og diske – den såkaldte DL380 serie. Der er indkøbt i alt 12 maskiner, der erstatter de nuværende 58 maskiner. Diskbehovet i 2009 skulle være dækket ind da den samlede bitarkiv plads med de nye maskiner bliver 180TB.

De første to maskiner er sat i produktion og fungerer helt tilfredsstillende. Løbende vil nye maskiner blive koblet på og sideløbende med de kørende høstninger vil indholdet af de gamle maskiner blive overført til de nye maskiner. Til sidst kontrolleres at arkiverne på Det Kongelige Bibliotek og på Statsbiblioteket er identiske og derefter kan de 58 gamle PC'ere slukkes for evigt.

Internationalt samarbejde og konferencer

IIPC (International Internet Preservation Consortium) blev afholdt i Ottawa, Canada.

Fra Det Kongelige Bibliotek deltog Birgit Nordmark Henriksen og fra Statsbiblioteket deltog Bjarne Andersen. "Netarkivet afholdt en åben session om Netarchive Suite - det open source software Netarkivet har udviklet og bruger til at foretage alle indsamlinger med. Der var stor interesse omkring mødet og mange nye potentielle samarbejdspartnere og brugere af systemet."

Næste ECDL-konference afholdes på Korfu, Grækenland fra 27.sept. til 2. okt. 2009. Information på <http://www.ecdl2009.eu/>

I IIPC-regi deltager Netarkivet også i "2009 European Election Web Harvesting Project". Vores bidrag er væsentligst erfaringer, da vi alligevel høster de berørte emner.

NetarchiveSuite

Dagene 27-29. maj 2009 er afsat til en workshop om NetarchiveSuite- og høster-erfaringer.

Her kommer 5 deltagere fra ONB (Østrigs nationalbibliotek) og BnF (Frankrigs nationalbibliotek) til Det Kongelige Bibliotek og sammen med udviklere og samlingsfolk fra Netarkivet vil vi bidrage til at et mere internationalt samarbejde omkring udvikling af NetarchiveSuite, således som det var planlagt ved at gøre det til et Open Source projekt.

Redaktionsgruppemøde

Netarkivets redaktion holdt møde 27.april 2009 på Det Kongelige Bibliotek.

Antallet af domæner uden for .dk som vi høster er nu på 46000. Det er blevet meget sværere at finde nye interessante kandidater. Danmark hoster mange udenlandske sites, måske fordi danske host-virksomheder er billige og gode til det.

Vores høstere bliver blokeret af 363 domæner fordelt på 33 forskellige IP numre, viser den seneste måling. Det viser sig yderligere at ét IP nummer dækker over 281 url'er. Kun et enkelt domæne var redaktionsgruppen kede af ikke blev høstet, og man vil derfor rette henvendelse til dem for at bede dem om at åbne op igen.

Redaktionen bemærkede, at Facebook profiler bliver meget vigtige i valgkampen til kommunevalg 2009, ligesom Twitter. Derimod i mindre grad YouTube, MySpace og Flickr.

Oversigter over selektive høstninger og begivenhedshøstninger kan ses på <http://netarkivet.dk/indsamlingsDoku.html>

Kulturministeriet genbemander redaktionsgruppen og vi kommer til at se nye ansigter på næste møde i redaktionsgruppen.

Bemanding af Netarkivets udviklingsfunktion

På Statsbiblioteket er Kåre Fiedler Christiansen er tilbage fra orlov.

Colin Rosendahl er i gang med at implementere den vellykkede testinstallation af Wayback på produktionsmiljøet.

Desuden er den nyuddannede Henrik Kirk blevet ansat, og vi ser frem til et godt samarbejde.

Det kongelige Bibliotek har pt. tilknyttet 4 udviklingsmedarbejdere plus en projektleder til Netarkivet.

Planer for den kommende periode

1. Workshop for udviklere
2. Fortsat etablering af ny hardwarearkitektur på Det kongelige Bibliotek og overførsel af arkivdata til nye maskiner.
3. Implementering af Wayback access-modul i produktionsmiljø
4. Begivenhedshøstninger: Eu-valg, kommunevalg og klimatopmøde

Mange hilsner
Claus Lomborg, Netarkivet