



Runsaudensarvi vai musta aukko: palvelevatko verkkoarkistot tulevien tutkijasukupolvien tarpeita?

Historiantutkimuksen päivät, 25.10.2012

Jyrki Ilva (jyrki.ilva@helsinki.fi)

”Reaalimaailma” siirtyy verkkoon

- Verkoaineistojen ja verkossa tapahtuvan aktiivisuuden määrä kasvaa räjähdysmäisen nopeasti
- Yhä suurempi osa myös ihmisten arkipäiväisestä elämästä tapahtuu ainakin osittain verkossa
 - Monet aiemmin verkon ulkopuolella tapahtuneet asiat siirtyneet nopeasti verkkoon
 - Samalla syntynyt kokonaan uusia ilmiöitä joita ei ole aiemmin ollut olemassa
 - Uutiset ja mielipiteet leviävät nopeasti verkossa, samoin poliittinen propaganda
- Miten kaikki tämä verkkomaailmassa tapahtuva tai ainakin sinne dokumentoituva aktiivisuus on tulevien tutkijasukupolvien tavoitettavissa?



Verkkoarkistoinnin idea

- Verkkoaineistoja voidaan kerätä joko luovutuksina tai haravoimalla
 - Haravointi toimii samoilla periaatteilla kuin Googlen kaltaisten hakupalveluiden tekemä verkkosisältöjen keruu ja indeksointi
 - Yksinkertaisimmillaan toimii siten että automaattinen hakurobotti poimii talteen verkkosivuja ja muita verkosta löytyviä tiedostoja seuraamalla sivuilla olevia linkkejä
- Verkkoaineistojen pitkäaikaissäilytys ja pysyvä saatavuus
 - Hakukoneista poiketen verkkoaineistoon kerätyt aineistot kuitenkin tallennetaan pysyvästi ja ne pyritään pitämään saatavilla ja käyttökelpoisina myös tulevaisuudessa
 - Verkkoarkiston käyttöliittymässä voidaan selata (lähes) samanikäisiä sivuja ja liikkua ikään kuin paikalleen pysäytetyssä kuvassa jostain tietystä hetkestä

Ketkä arkistoivat?

- Verkkoaineistoja keräävät ja tallentavat eri tarkoituksiin useat eri tahot:
 - Internet Archive
 - Eri maiden kansalliskirjastot ja muut julkishallinnolliset organisaatiot (usein lakisäätöistä toimintaa)
 - Kaupalliset toimijat (mm. Google)
 - Tutkimusryhmät, yksityiset kansalaiset ym.
- Kaikki kerätty aineisto ei välttämättä päädy julkisiin arkistoihin...
- Lisäksi verkkosivustoja, jotka arkistoivat tehdyt muutokset (esim. Wikipedia ja wikit)
 - Myös visioita aikaulottuvuuden integroimisesta entistä syvemmälle verkkopalvelinten tarjoamiin toiminnallisuuksiin (ks. <http://mementoweb.org/>)

Wayback Machine (Internet Archive)

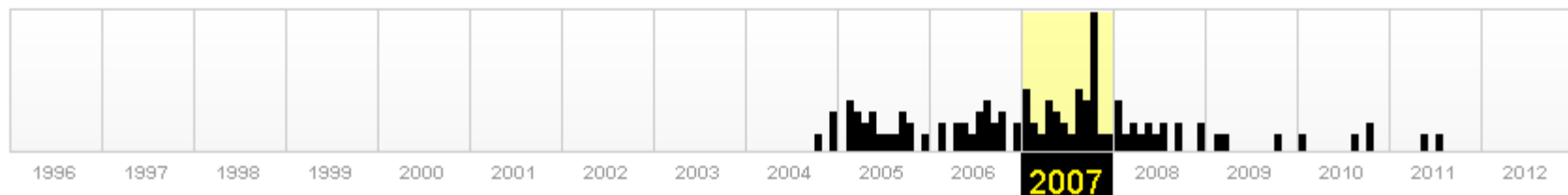
- Tunnetuinta globaalia verkkoarkistoa ylläpitää **Internet Archive** (<http://archive.org>)
 - Amerikkalaisen IT-miljonäärin Brewster Kahlen perustama
 - Internet Archivella verkkoarkiston lisäksi runsaasti muitakin vapaasti käytettäviä aineistoja, mm. digitoituja kirjoja ja musiikkia
- Internet Archiven **Wayback Machine** - käyttöliittymä sisältää vuodesta 1996 lähtien haravoituja verkkoaineistoja, joita voi hakea URL-osoitteen perusteella

The screenshot shows the Internet Archive website interface. At the top, there's a navigation bar with links for Web, Video, Texts, Audio, Projects, About, Account, TVNews, and Open Library. A search bar is prominently displayed. Below the search bar, there are several featured content boxes. The 'Web' box highlights '150 billion pages' and includes a 'Take Me Back' button. The 'Video' box features 'MACHUNAS: A Performance' with a 4.5-star rating. The 'Live Music' box highlights 'Grateful Dead Live at Nashville Municipal' with a 4.5-star rating. The 'Audio' box features 'Melody Scenes' with a 4.5-star rating. The 'Texts' box features 'Sainte Jeanne-Françoise Frémoyot de Chantal, sa' with a 4.5-star rating. At the bottom, there's a forum section titled 'Most recent posts' with a table of recent discussions.

Subject	Poster	Forum	Replies	Date
Re: Question (Dead-Related)	corfelle	GratefulDead	0	14 minutes ago
Brewster Kahle's Internet Archive	Laura Mihy	news	0	18 minutes ago
Re: Question (Dead-Related)	craven714	GratefulDead	1	22 minutes ago
Re: Franklins Tower + Lindley Meadows SF 09/20/75	johnnyonthespot	GratefulDead	0	41 minutes ago
TDH - Phil Your Day With A Philo Stomp	rdennis87	GratefulDead	0	1 hour ago

<http://agricola.utu.fi> has been crawled **113 times** going all the way back to [October 11, 2004](#).

A crawl can be a duplicate of the last one. It happens about 25% of the time across 420,000,000 websites. [FAQ](#)



JAN

1	2	3	4	5	6	
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30	31			

FEB

				1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28			

MAR

			1	2	3	
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	31

APR

			1	2	3	4	5	6	7
8	9	10	11	12	13	14			
15	16	17	18	19	20	21			
22	23	24	25	26	27	28			
29	30								

MAY

1	2	3	4	5		
6	7	8	9	10	11	12
13	14	15	16	17	18	19

JUN

				1	2		
3	4	5	6	7	8	9	
10	11	12	13	14	15	16	

JUL

				1	2	3	4	5	6	7
8	9	10	11	12	13	14				
15	16	17	18	19	20	21				

AUG

								1	2	3	4
5	6	7	8	9	10	11					
12	13	14	15	16	17	18					

Internet Archive: rajoituksia ja ongelmia

- Internet Archiven toiminnalla ei selkeää julkista mandaattia tai rahoitusta
 - Internet Archiven toiminta perustuu USA:n tekijänoikeuslainsäädännön fair use -periaatteeseen, ei olisi mahdollista esim. Suomessa
 - Tehnyt kuitenkin läheistä yhteistyötä eri maiden kansalliskirjastojen kanssa
- Joutuu ottamaan aineiston julkaisijoiden näkemyksiä huomioon
 - Noudattaa esim. palvelinten robots.txt-tiedostoissa annettuja määräyksiä myös takautuvasti: jo kerättyä aineistoa poistuu saatavilta (ei todennäköisesti poisteta kokonaan!) erilaisten teknisten muutosten myötä
- Tukee toisaalta myös käyttäjälähtöistä verkkoarkistointia ja haravoitujen teemakokonaisuuksien luomista (<http://www.archive-it.org/>)

Kansalliskirjasto ja verkkoaineistojen arkistointi

- Laki kulttuuriaineistojen säilyttämisestä ja tallettamisesta (voimaan 2008 alkaen)
 - Suomessa julkaistujen tai suomalaiselle yleisölle suunnattujen verkkoaineistojen arkistointi Kansalliskirjaston tehtäväksi
- Kirjaston näkökulmasta ”uusi vapaakappalelaki”, jatkoa aiemmalle painotuotteiden ja audiovisuaalisten aineistojen keräämiselle
 - Näkyy ajoittain toiminnan painotuksissa: saatetaan keskittyä perinteisiä kirjoja tai muita aiempia vapaakappaleaineistoja muistuttavien aineistojen (esim. julkishallinnon julkaisut) keräämiseen verkolle tyypillisten aineistotyyppien sijaan
 - Lakia valmisteltaessa esim. keskustelupalstat nousivat keskusteluissa esiin aineistona jonka keräämistä ei haluttu priorisoida: toisaalta keräämiselle ei varsinaista estettä
 - Kansalliskirjastolla varsin suuret vapaudet keräystavoitteiden (”edustava otos”) määrittelyssä

Aineiston keruu

- Aineistoja kerätään sekä kustantajilta saatavina luovutuksina että verkkoharavoinnin avulla
- Luovutuksina kerätään esim. verkossa julkaistua musiikkia tai e-kirjoja
- Suurin osa aineistoista kerätään kuitenkin haravoimalla
 - Yleiset ”Suomi-keräykset” ja erilliset teemakeräykset (esim. vaalit, olympialaiset, blogit, seksuaalivähemmistöt, ilmastonmuutos, jne.)

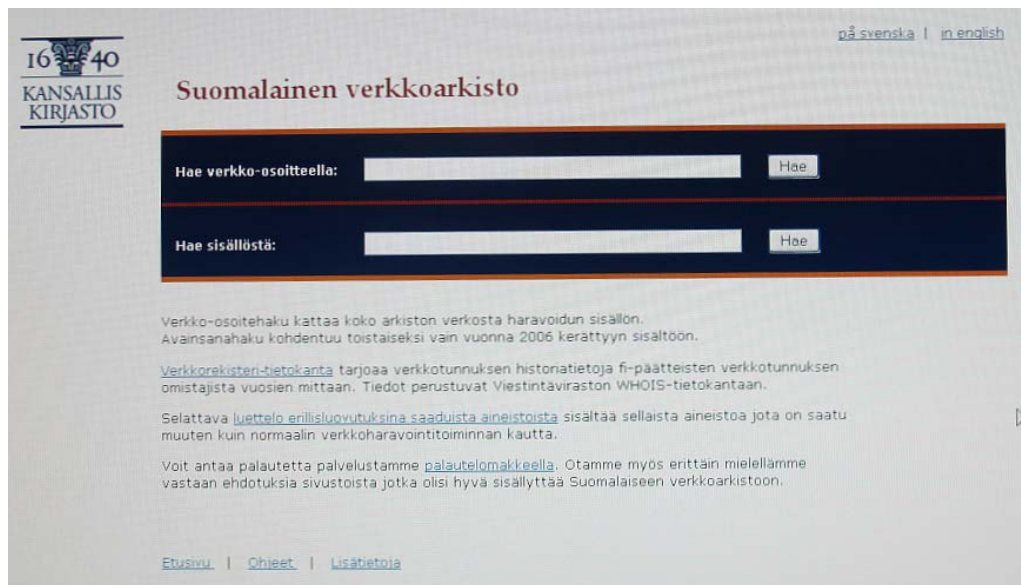
The screenshot shows the VARiA website interface. At the top, there are navigation links for 'Suomeksi', 'På svenska', and 'In English'. Below that, there's a search bar with the text 'Hae 2' and a dropdown menu with options 'Hae tästä kokoelmasta' and 'Hae Variasta'. The main content area is titled 'Verkköäänitteet' and contains a list of items. Each item has a small thumbnail image, a title, and the author 'Holik' and date '2012-08-15'. The items listed are:

- Nimeke: *Bemused/Death Is In Fashion*
- Nimeke: *The Likes You And The Likes Of Me*
- Nimeke: *This Shit Just Got Darker*
- Nimeke: *Like A Landslide*

On the right side of the page, there are two sections: 'Selaa yhteisöä' with a list of links (Nimekkeet, Tekijät, Julkaisijat, Asiasanat, Uusimmat, Sivukartta) and 'Omat tiedot' with links (Kirjautu sisään, Rekisteröidy).

Suomalainen verkkoarkisto

- Haravoitu aineisto kootaan Suomalaiseen verkkoarkistoon
- Hyödyntää samanlaisia teknisiä ratkaisuja kuin Internet Archiven Wayback Machine
- Hakuja mahdollista tehdä sekä URL:in perusteella että indeksoidusta sisällöstä



The screenshot shows the homepage of the Finnish National Library's website. At the top left is the logo with the text '1640 KANSALLIS KIRJASTO'. In the top right corner, there are links for 'på svenska' and 'in english'. The main heading is 'Suomalainen verkkoarkisto'. Below this, there are two search bars: 'Hae verkko-osoitteella:' and 'Hae sisällöstä:'. Each search bar has a 'Hae' button to its right. Below the search bars, there is a paragraph of text: 'Verkko-osoitehaku kattaa koko arkiston verkosta haravoidun sisällön. Avainsanahaku kohdentuu toistaiseksi vain vuonna 2006 kerättyyn sisältöön.' This is followed by a link to 'Verkkorekisteri-tietokanta' and a paragraph explaining that it provides historical data for domain owners. Another link 'Luettelo enillisuovutuksina saaduista aineistoista' is mentioned, along with a note that it contains material obtained through the normal network archiving process. At the bottom, there is a link to 'palautelomakkeella' and a note that the library is also looking for suggestions of websites to include in the archive. At the very bottom, there are links for 'Etusivu', 'Ohjeet', and 'Lisätietoja'.

Käyttörajoituksia

- Lainsäädännössä määritellyistä käyttörajoituksista tuli osin suurten mediayhtiöiden painostuksen tuloksena tiukemmat kuin etukäteen ennakoitiin
- Aineistot käytettävissä vain vapaakappalekirjastoissa ja KAVA:ssa erillisiltä työasemilta
- Verkon kautta käytettävissä vain verkkoarkiston hakemisto, <http://verkkoarkisto.kansalliskirjasto.fi/>



Käyttörajoitusten vaikutus tutkimuskäyttöön

- Tiukat käyttörajoitukset haittaavat oleellisesti aineistojen tutkimuskäyttöä
 - Aineistoja ei saa vapaakappaletyöasemillakaan kopioida digitaalisessa muodossa (tulostaminen ja valokuvaaminen ruudulta sallittua)
 - Esim. tietokantamuotoisten aineistojen ominaispiirteitä ei mahdollista hyödyntää
 - Tiedon louhinta laajoista datamassoista (vapaakappalekirjastojen ulkopuolella?) edellyttäisi sopimuksia oikeudenhaltijoiden kanssa
- Verkkoarkiston sisältämä aineisto alkaa vuodesta 2006
 - Mukana toki jo aiemmin julkaistuja aineistoja, mutta nämä haravoitu vasta myöhemmin
 - Suomalaisen Internetin varhaishistoria jää osin hämärän peittoon?

Verkkoarkisto ja sananvapauden rajat

- Verkkoarkistoon kerätään pääsääntöisesti kaikki aineistot lapsi- ja väkivaltapornoa tai verkkoon vahingossa päätyneitä henkilösuojan kannalta sensitiivisiä aineistoja lukuun ottamatta
- Toisaalta äskettäin saatiin toisenlainenkin esimerkki aineistojen poistamisesta

The screenshot shows the website of the National Library of Finland (Kansalliskirjasto). The page is in Finnish and features a navigation menu on the left with categories like 'Etusivu', 'Yleistiedot', 'Kokoelmat ja palvelut', 'Kulttuuritoiminta', 'Kirjastoille', and 'Julkaisijoille'. The main content area displays a news article titled 'Kansalliskirjasto poistaa Halla-ahon loukkaaviksi katsotut verkko kirjoitukset yleisökäytöstä' (National Library removes Halla-aho's defamatory online writings from public use). The article, dated 31.7.2012, discusses a court decision by the Supreme Court (KKO) regarding the removal of defamatory content from the digital archive. A search bar is visible in the top right corner, and there are several small promotional banners on the right side of the page.

Verkkoarkistoinnin haasteita (1): Mistä lähdetään liikkeelle, miten syvälle mennään ja kuinka usein?

- Joudutaan tekemään valintoja sen suhteen miltä sivuilta lähdetään liikkeelle, miten paljon sivuja kerralla haravoidaan talteen ja miten usein haravointi toistetaan
 - Koko verkon kattava haravoiminen talteen käytännössä mahdotonta, ainakin ilman Google-luokan resursseja
- Oikean syvyyden määrittäminen haastavaa: vaihtelee palvelimittain
 - Jos ei mennä tarpeeksi syvälle, oleelliset dokumentit eivät tule mukaan
 - Toisaalta arkistoon voi kertyä paljon roskaa

Esimerkki: Hommafoorumi (1)

- Homma näyttäisi ensi silmäyksellä olevan melko hyvin hallussa...

Notef Forms, search boxes or links to external sites may not function properly. Site: http://hommaforum.org/index.php?board=22.0.html Time: 6/16/17 Jun 02, 2011 [Web Archive] [Info]

Etusivu Forum Ohjeet Haku Kirjautu Rekisteröidy

Tervetuloa, **Vieras**. Ole hyvä ja kirjautu tai rekisteröidy.
Jaikö aktivoitsähköposti saamatta?

Viimeisimmät, FAQ, Säännöt

Kirjautu käyttäjätunnuksen, salasanan ja istunnonpituuden mukaan: **692 841** viestiä **21 545** aiheetta- kirjoittanut **6 155** jäsentä

HOMMAN POLITTIINEN SIIPPI Puolueet

Alakategoriat

	1266 viestiä 28 aiheetta	Uusin viesti kirjoittanut Kaapo Vi: Kamesta Katsiaan... Tänään kello 05:11:27
	14403 viestiä 303 aiheetta	Uusin viesti kirjoittanut Veb Vi: Suora demokratia esi... 01.06.2011, 23:01:52
	6900 viestiä 159 aiheetta	Uusin viesti kirjoittanut Juh Eemila Vi: Kymn piiri bedotta... 28.05.2011, 23:28:34

Sivuja: [1] Siirry alas

Alue	Aloittaja	Vastauksia	Lukukerrat	Uusin viesti
	Hommaforum Admin	0	1080	23.11.2009, 14:44:49 kirjoittanut Hommaforum Admin

Sivuja: [1] Siirry ylös

HOMMAN POLITTIINEN SIIPPI Puolueet

Notef Forms, search boxes or links to external sites may not function properly. Site: http://hommaforum.org/index.php?board=23.0.html Time: 7/04/17 Jun 02, 2011 [Web Archive] [Info]

Etusivu Forum Ohjeet Haku Kirjautu Rekisteröidy

Tervetuloa, **Vieras**. Ole hyvä ja kirjautu tai rekisteröidy.
Jaikö aktivoitsähköposti saamatta?

Viimeisimmät, FAQ, Säännöt

Kirjautu käyttäjätunnuksen, salasanan ja istunnonpituuden mukaan: **692 810** viestiä **21 543** aiheetta- kirjoittanut **6 154** jäsentä

HOMMAN POLITTIINEN SIIPPI Puolueet Perussuomalaiset

Sivuja: [1] 2 3 ... 6 Siirry alas

Alue	Aloittaja	Vastauksia	Lukukerrat	Uusin viesti
	IDA	1	223	28.05.2011, 23:28:34 kirjoittanut Juh Eerola
	Reich	54	2850	27.05.2011, 15:28:23 kirjoittanut Lalli IzoTalo
	Miniluv	0	145	17.05.2011, 12:23:43 kirjoittanut Miniluv
	Ant.	0	387	11.05.2011, 16:18:32 kirjoittanut Ant.
	James Hirvisaari	55	5775	05.05.2011, 22:49:23 kirjoittanut Janne
	Teemu Lahtinen	7	917	05.05.2011, 10:42:45 kirjoittanut Tuujalainen
	Olli Immonen	78	7706	02.05.2011, 20:39:30 kirjoittanut Perja
	Riikka Slunga-Poutsalo	11	1962	02.05.2011, 20:05:54 kirjoittanut Riikka Slunga-Poutsalo
	Teemu Lahtinen	338	25922	02.05.2011, 07:54:40 kirjoittanut Simo Salmela

Esimerkki: Hommafoorumi (2)

- Mutta lähempi tarkastelu paljastaa, että arkistointi ei ole yltänyt foorumin viesteihin asti
- Internet Archiven puolelta löytyisi tässä tapauksessa enemmän luettavaa...

The screenshot shows a web browser window with the URL `arkisto.kansalliskirjasto.fi:8080/20110602045433/http://hommaforum.org/index.php/topic,49565.0.html`. The page is titled "Suomalainen verkkoarkisto" and has language options for "på svenska" and "in english". There are two search boxes: "Hae verkko-osoitteella:" with the text "http://" and "Hae sisällöstä:". Below the search boxes, the text reads "Puuttuva sivusto" and "Hakemaasi sivua ei ole tallennettu. Miksei sivua ole arkistossa? Ks. [lisätietoja](#). Voit myös [ehdottaa sivun tallentamista](#)." At the bottom, there are links for "Etusivu", "Ohjeet", and "Lisätietoja".

Verkkoarkistoinnin haasteita (2): Verkon globaali luonne

- Yhä suurempi osa suomalaisten tuottamista ja suomalaisille suunnatuista aineistoista muunmaalaisilla palvelimilla
- Suomalaisten aineistojen löytäminen ja poimiminen talteen vaikeaa
- Koti- ja ulkomaiset aineistot linkittyvät toisiinsa moniin eri tavoin: ”suomalaisten” aineistojen suodattaminen erikseen rikkoo verkon perusluonteen



The image shows a YouTube video player interface. At the top, the YouTube logo is on the left, and a search bar contains the text "matti klinge". Below the search bar, the video title is "Professor emeritus Matti Klinge: Eurooppa 1812 (osa I, del I)". Under the title, there are buttons for "addetovideo" (with a plus sign), "Tilaa", and "32 videoita" (with a dropdown arrow). The video player itself shows a man in a dark suit and glasses, identified as Professor Emeritus Matti Klinge, speaking. The video progress bar at the bottom indicates the video is at 00:15 / 13:21. Below the video player, there are buttons for "Tykkään" (with a thumbs up icon), "Lisää soittolistaan" (with a plus icon), "Jaa" (with a share icon), and "6 näytökertaa" (with a play icon). At the bottom of the page, there is a description in Finnish: "Julkaistu: 2.10.2012. Julkaisija: addetovideo. Turku 1812 - tieteellinen seminaari 23.8.2012 Turun pääkirjastossa. Professor emeritus Matti Klingen esitelmä Eurooppa 1812. Professor emeritus Matti Klinges föredrag på Åbo 1812-seminariet torsdagen den 23 augusti 2012." On the right side of the description, it says "0 tykkää, 0 ei tykkää".

Verkkoarkistoinnin haasteita (3): Tekninen monimutkaistuminen

- Nykyiset verkkopalvelut yhä monimutkaisempia teknisiä kokonaisuuksia
 - Samalla sivulla erilaisia koodinpätkiä, upotettuja videoita, riippuvuuksia toisilla palvelimilla sijaitsevista aineistoista ja ohjelmakoodeista
 - Koskee mm. monia suurten verkkomedioiden ja yritysten sivustoja
- Hankalaa kerätä talteen ja säilyttää toimivana kokonaisuutena
 - Joskus sivut toimivat vain osittain, osa ei toimi käytännössä ollenkaan
 - Pitäisikö sivuista ottaa tallentamisen yhteydessä saman tien myös kuvankaappaus alkuperäisen ilmiasun säilyttämiseksi?

Esimerkki: Helsingin Sanomat (1)

- Vuonna 2006 lehden verkkosivut olivat vielä melko yksinkertaiset

Notel Forms, search boxes or links to external sites may not function properly. Site: http://www.hs.fi/ Time: 13:42:49 Jun 22, 2006 [Web Archive] [hide]

HELSINGIN SANOMAT Nyt Radio Helsinki Älypää Tilaa Hesari Mediatiedot Oikotie Sanoma Osakeyhtiö

HS.fi
Torstaina 22.6.2006

ASIAKASPALVELU | Tilaa Hesari
Osoitteenmuutos | Jakeluhäiriöt
Jätä ilmoitus

Hae HS.fi:stä: HAE

In English | Tekstiversio | Uutiskirje | Kirjautu sisään | RSS

KOTISIVU

UUTISET
Kotimaa
Kaupunki
Ulkomaat
Talous
Urheilu
Kulttuuri
Sää

MM-FUTIS '06

TEEMAT
Asuminen
Autot
Henk. Koht.
Matkailu

OSALLISTU
Blogit
Uutisvinkki
Keskustelut
Palaute

HUPI
E-kortti
Sudoku
Tv-opas
Uutisguru
Viivi & Wagner
Älypää

KUUKAUSILITE

NYT
Elokuva
Musikki
Revintola
Menot

UUTISET
EK: Sairauspoissaolot lisääntyivät työpaikoilla
(11.19)
Elinkeinoelämän keskusliitto EK haluaa pysäyttää sairauspoissaolojen kasvun. EK:n jäsenyrityksissä sairaus- ja tapaturmapoissaolot nousivat viime vuonna 4,5 prosenttiin teoreettisesta säännöllisestä työpästä.


MUUT PÄÄUUTISET
► Juhannusruuhkat pahimmillaan tänään iltapäivällä (Päivitetty 0.21) (21.6. 19.12)
► Uimavedet ajankohtaan nähden erittäin lämpimiä (12.00)
► Tukholman ruuhkamaksukokeilu ylitti odotukset (7.29)
► Keräilyaseita valuu rikollisten käyttöön (21.6. 22:00)
► Kuljettaja loukkaantui vakavasti Hämeentiellä Helsingissä (10.56)

Lisää uutisia »

ILMOITUS

E-KORTIT
Rauhallista juhannusta!
Lähetä ystäväillesi sähköinen juhannustervehdys.

URHEILU
Formula 1
Kanadan osakilpailu.
HS.fi tarjoaa reaaliaikaisen tulospalvelun viikonlopun kisasta Montrealin radalla ja paljon muuta.

VAHINGOSSA VERKKOON

Osa 3: Sari Kalkaja Stadionin tornissa
Sarjassa haastatellaan sattumanvaraisesti valittuja ihmisiä eri puolilla Helsingiä.

KESKUSTELUT
• Helsingin rajansiirrot saivat hyväksynnän isoilta hallituspuolueilta
• Tukholman ruuhkamaksukokeilu ylitti odotukset
• Paljonko loma saa maksaa?
• TV-maksu nousee vuodenvaihteessa
• Mikä Turussa on parasta?

Connecting to www.google-analytics.com

Esimerkki: Helsingin Sanomat (2)

- Vuoden 2008 lehdessä ongelmia on näkyvissä jo enemmän
- Uusimmissa lehdissä ei sitten enää olekaan kovin paljon näytettävää...

53319 asuntota 36124 autoa 1529 työpaikkaa
 Maanantaina 14.04.2008
 Nimiävät Taito · Onnittele E-kortilla
 HELSINKI SAA lähimpinä
 MA 14:00 paikkakunnittain >
 Tilaa Hesari | Osoitteenmuutos | Jätä ilmoitus

Palautte | Uutisvinkki | Apua Hae HS:stä HAE

Etusivu UUTiset TUOREIMMAT UUTiset >

Uutiset
 Kotimaa
 Poliittika
 Kaupunki
 Ulkomaat
 Talous
 Saa
 Pääkirjoitus
 Urheilu
 Kulttuuri
 Kirjat
 Elokuvat
 Musiikki
 Televisio
 Menokone

Viihde
 E-kortti
 Uutisguru
 Sudoku
 Alypää
 Viini & Wagner
 Fingerponi
 Ravintolakone
 Musiikkilataamo

Oma elämä
 Testit

Oma kaupunki

Autot
 Vikasarja

UUTiset
Inflaatio kiihtyi jo lähes neljään prosenttiin
 (9:33)
 inflaatio kiihtyi jo lähes neljään prosenttiin
 Suomen inflaatio kiihtyi maaliskuussa jo 3,9 prosenttiin. Helmikuussa inflaatioluku oli 3,7 prosenttia. Inflaatio tarkoittaa kuluttajahintojen vuosimuutosta.

POLITIKKA
 Kanerva ei ole ilmoittanut sairauslomaa jatkosta Ilkka Kanerva
Kanerva ei ole ilmoittanut sairauslomaa jatkosta
 (12:55)
 Kansanedustaja Ilkka Kanerva (kok) palanee tiistaina eduskuntatyöhön sairauslomaltaan. Vielä maanantaina puolittapäivän eduskunta ei ollut saanut ilmoitusta entisen ulkoministerin sairauslomaa jatkosta.

TALOUS
Kauppakamari vaatii selvitystä tasaverosta
 (Päivitetty 11:42) (11:38)
 Keskuskauppakamari vaatii tasaveron selvittämistä valtion tuloverotuksessa. Jos uudistus toteutuisi, valtaosalla suomalaisista olisi sama veroprosentti.

ULKOMAAT
Poliisi sai vihjeitä surmaajasta jo ennen Englan surmaa
 (Päivitetty 12:53) (13:4, 20:37)
 Poliisi sai vihjeitä surmaajasta jo ennen Englan surmaa 10-vuotiaan Engla Juncosa-Höglundin kuolleen Pernilla Helgrenin surmat sunnuntaina tunnustanut mies oli tuttu. Poliisi sai miehestä viime marraskuussa vihjeen

Viini ja Wagnerin parisuliteen hoitokone
 Testaa ja kerää rahanaivoisia pisaroita meren suojeluun.

PAIVÄN KYSYMYKSIÄ
 Soditaanko Irakissa vielä viiden vuoden kuluttua?
 Kyllä Ei

KESKUSTELUT
 • Opiskelijat kiistelevät aseenkannosta Yhdysvalloissa
 • Ansaitsevatko johtajat korkeat palkkansa?
 • Uskotko ammatillaisen luomaan imagoon?
 • Kauppakamari vaatii selvitystä tasaverosta
 • Vanhasen väitetyt viestit Ruusulle päätyivät nettiin

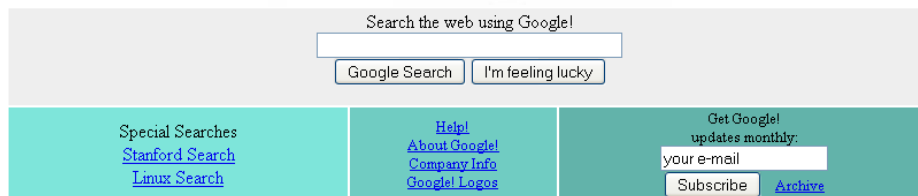
VIDEOT
 Kymmenettuhannelle haluavat kantaa aseita kampuskuolella (13.4.2008)
 Lisää videoita >

BLOGIT
 Innostunut Johnny
 • Rajatapauksia
 Mitä ajattelin tänään
 • Massacre
 Sami Sykon tyylillä
 • Jotain tutua
 Sami Dumanovassa

Connecting to adserver.sanomainsv.fi

Verkkoarkistoinnin haasteita (4): Käyttökokemus

- Verkossa julkaistujen dokumenttien vs. käyttökokemuksen säilyttäminen?
 - Vielä 1990-luvulla verkkosivut melko kiinteitä kokonaisuuksia, jotka näyttäytyivät kaikille käyttäjille samanlaisina
- Verkkopalvelut yhä enemmän personoituja, erilaisia eri käyttäjille
 - Verkkopalvelut keräävät käyttäjistä ja näiden valinnoista dataa, jonka pohjalta pyritään tarjoamaan käyttäjille kiinnostavaa sisältöä (etenkin mainoksia)
 - Jopa Google-haun antamat tulokset vaihtelevat nykyään radikaalistikin useiden eri tekijöiden vaikutuksesta



Verkkoarkistoinnin haasteita (5): Aidatut puutarhat

- Vapaan verkon rinnalle kehittynyt erilaisia kaupallisia palveluita, joihin syntynyt/siirtynyt paljon sisältöä
 - Perustuvat usein sosiaalisten verkostojen rakentamiseen, monien sisältöjen käyttöoikeus usein rajattu tietyille käyttäjäjoukolle
 - Käyttöoikeudeltaan rajatun aineiston haravoiminen verkkoarkistoihin käytännössä vaikeaa tai jopa mahdotonta
 - Käyttäjillä vaihtelevat mahdollisuudet oman datansa tallentamiseen
 - Sama koskee myös mobiilimaailmassa yleistyneitä appseja, jotka osin syrjäyttämässä osittain vapaassa verkossa toimivia palveluita
- Toisaalta Kongressin kirjasto (USA) sopinut Twitterin kanssa twiittien kattavasta arkistoinnista, eli arkistointi ei ole mahdotonta, jos vain poliittista tahtoa löytyy

Esimerkki: Facebook (1)

- Internet Archive ei arkistoi Facebookissa sijaitsevia verkkosivuja lainkaan
- Facebook kieltänyt palvelussaan sijaitsevien julkistenkin sivujen arkistoinnin robots.txt-tiedostossaan kesästä 2010 lähtien
- Vaikuttaa takautuvasti myös aiemmin kerättyihin sivuihin



<http://www.facebook.com/>

Latest

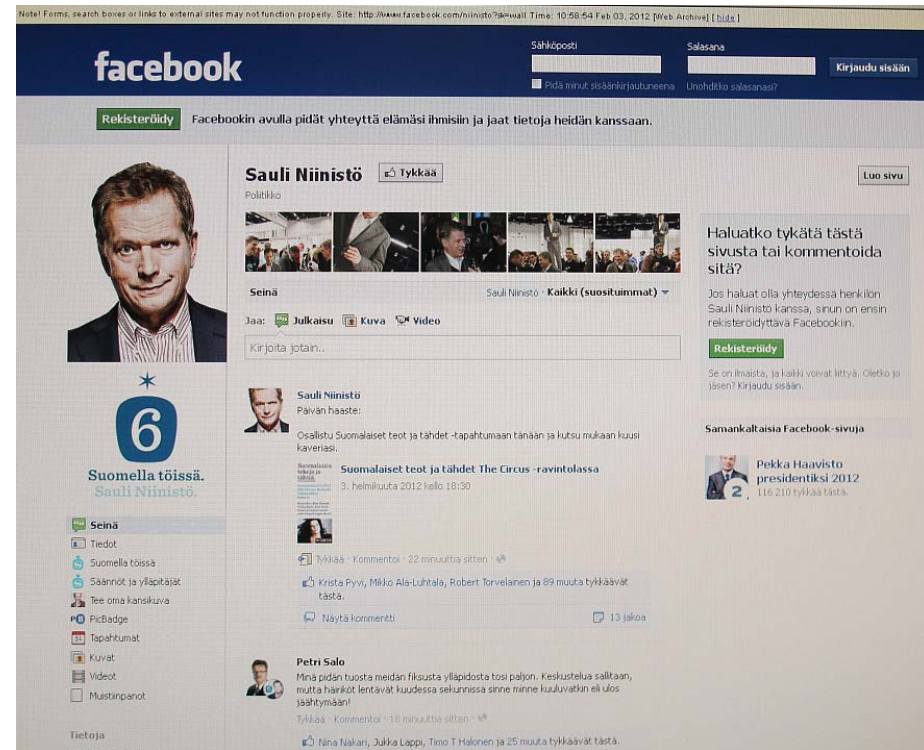
Show All

Page cannot be crawled or displayed due to robots.txt.

See [www.facebook.com robots.txt](http://www.facebook.com/robots.txt) page. [Learn more](#) about robots.txt.

Esimerkki: Facebook (2)

- Kansalliskirjaston verkkoarkisto ei sen sijaan aina noudata palvelinten robots.txt-tiedostojen määräyksiä
- Arkistoon kerätty myös julkisia Facebook-sivuja, tosin varsin satunnaisesti
- Oheinen sivu lienee päätynyt talteen vaalikeräyksen yhteydessä...



Verkkoarkeologiaa?

- Käytännössä verkkoarkistot sisältävät aina vain otoksia verkossa saatavissa olleista dokumenteista ja näkymistä
 - Otokset väistämättä monin tavoin epätyytyttäviä ja satunnaisia
 - Oleellisia dokumentteja puuttuu, monet dokumentit rikki
 - Konteksteja ja dokumenttien välisiä yhteyksiä voi olla vaikea hahmottaa
- Aineistojen tulkinta edellyttää lähdekritiikkiä; aikanaan olemassa olleen verkkotodellisuuden piirteitä joudutaan rekonstruoimaan eri lähdekokonaisuuksissa säilyneiden jäänteiden perusteella
 - Aineistoja käyttäviltä tutkijoilta vaaditaan vähintään verkkotekniikoiden ja verkkoarkistoinnin perusteiden tuntemista

Verkkoaineistojen arkistointi: vaikeaa, mutta erittäin tärkeää?

- Kuten edellä on käynyt ilmi, verkkoaineistojen arkistoinnissa mahdotonta päästä lähelle minkäänlaista täydellisyyttä
 - Silti tärkeää aikakautemme kulttuuriperinnön säilymisen kannalta
 - Suurta osaa nyt keräämättä jäävistä lähteistä ei saada talteen enää jälkikäteen
- Toistaiseksi suomalaiset tutkijat olleet hyvin passiivisia verkkoarkistoinnin suhteen
 - Tavallaan ymmärrettävää: Internet-aikakautta tutkitaan vielä vähän etenkin historiantutkimuksen piirissä, ja aineistojen käyttö siten lähes olematonta
 - Silti tutkijayhteisön panosta ja palautetta kaivattaisiin keskustelussa siitä, millaisia prioriteetteja verkkoaineistojen keruussa pitäisi olla ja missä kohdin toimintaa pitäisi kehittää
 - Tai no, kun nyt ensin edes olisi se keskustelu...



16 40

KANSALLIS
KIRJASTO