

Predspletne internetne storitve 3

WAIS,
WWW

WAIS

WAIS - Wide Area Information Servers.

- ❖ Iskalniki povezani v distribuiran sistem.
- ❖ Vsak strežnik WAIS je skrbel za iskanje po lokalnih podatkovnih zbirkah.
- ❖ Vsi strežniki so sodelovali pri istem iskanju in posredovali zbrane rezultate iskanja.

WAIS

- ❖ Potek iskanja:
 - ❖ odjemalec prikaže kazalo delujočih strežnikov,
 - ❖ uporabnik izbere poljubno število tistih, za katere misli, da lahko vsebujejo želene informacije,
 - ❖ zastavi iskalno zahtevo,
 - ❖ WAIS posreduje zahtevo izbranim informacijskim strežnikom po celem svetu in
 - ❖ od njih spet zbere odgovore.
- ❖ Iskanje torej teče globalno.

WAIS

Zakaj je bil WAIS pomembna faza pri gradnji omrežnih iskalnikov?

- ❖ Informacijski objekti v sistemu WAIS so bili avtomatsko indeksirani dokumenti.
- ❖ Iskalni vmesnik je omogočal iskalne zahteve v naravnem jeziku.
- ❖ Rezultati iskanja so bili rangirani po izračunani relevantnosti.

WAIS – prikaz delovanja (I. 1998)

```
SWAIS                                     Source Selection                               Sources: 392
#           Server                          Source                                           Cost
001: * [     archie.au] aarnet-resource-guide                          Free
002:   [  munin.ub2.lu.se] academic_email_conf                          Free
003:   [wraith.cs.uow.edu.au] acronyms                                       Free
004:   [  archive.orst.edu] aeronautics                                   Free
005:   [  bloat.media.mit.edu] Aesop-Fables                                  Free
006:   [  ftp.cs.colorado.edu] aftp-cs-colorado-edu                          Free
007: * [nostromo.oes.orst.ed] agricultural-market-news                    Free
008:   [  archive.orst.edu] alt.drugs                                           Free
009:   [  wais.oit.unc.edu] alt.gopher                                           Free
010:   [sun-wais.oit.unc.edu] alt.sys.sun                                           Free
011: * [  wais.oit.unc.edu] alt.wais                                           Free
012:   [alfred.ccs.carleton.] amiga-slip                                           Free
013:   [  munin.ub2.lu.se] amiga_fish_contents                                       Free
014:   [  quake.think.com] Applications-Navigator                               Free
015:   [  132.183.190.21] Arabidopsis-BioSci                                       Free
016:   [weeds.mgh.harvard.ed] Arabidopsis_thaliana_Genome                    Free
017:   [  ftp.tex.ac.uk] archaeological_computing                                   Free
018:   [  archive.orst.edu] archie-orst.edu                                       Free
```

Keywords: information retrieval for full text databases

Enter keywords with spaces between them: <return> to search: ^C to cancel

- ❖ Del seznama strežnikov WAIS in zastavljanje iskalne zahteve **information retrieval for full text databases.**

WAIS – prikaz delovanja (I. 1998)

SWAIS			Search Results	Items: 22
#	Score	Source	Title	Lines
001:	[1000]	(altgopher)	antonr@ohs Re: Express Learning of langu	229
002:	[1000]	(aftp-cs-colorad)	RD.Comparison.ps PostScript /usr/local/f	1074
003:	[940]	(aftp-cs-colorad)	Doc/paper.ps PostScript /usr/local/ftp/p	1318
004:	[920]	(aftp-cs-colorad)	fwmanual.ps PostScript /usr/local/ftp/pu	4504
005:	[850]	(aftp-cs-colorad)	RD.Comparison.ps PostScript /usr/local/f	1074
006:	[850]	(aftp-cs-colorad)	PostScript/RD.Comparison.ps PostScript /	1074
007:	[809]	(aftp-cs-colorad)	Techniques.Wide.Area.ps PostScript /usr/	1538
008:	[809]	(aftp-cs-colorad)	PostScript/Techniques.Wide.Area.ps PostS	1538
009:	[809]	(aftp-cs-colorad)	Techniques.Wide.Area.ps PostScript /usr/	1538
010:	[799]	(aftp-cs-colorad)	Proj.Overview.ps PostScript /usr/local/f	959
011:	[799]	(aftp-cs-colorad)	PostScript/Proj.Overview.ps PostScript /	959
012:	[799]	(aftp-cs-colorad)	Proj.Overview.ps PostScript /usr/local/f	959
013:	[754]	(aftp-cs-colorad)	White.Pages.ps PostScript /usr/local/ftp	1318
014:	[754]	(aftp-cs-colorad)	PostScript/White.Pages.ps PostScript /us	1318
015:	[754]	(aftp-cs-colorad)	White.Pages.ps PostScript /usr/local/ftp	1318
016:	[618]	(aftp-cs-colorad)	Essence.ps PostScript /usr/local/ftp/pub	2075
017:	[618]	(aftp-cs-colorad)	PostScript/Essence.ps PostScript /usr/lo	2075
018:	[618]	(aftp-cs-colorad)	Essence.ps PostScript /usr/local/ftp/pub	2075

<space> selects, arrows move, w for keywords, s for sources, ? for help

- ❖ Del izpisa rezultatov iskanja. Izračun relevantnosti predstavlja stolpec Score.

WAIS

- ❖ WAIS se ni obdržal, čeprav je bil revolucionaren v številnih pogledih.
- ❖ V omrežno okolje je vpeljal iz klasičnega IR najsodobnejša načela gradnje in uporabe zbirk polnih dokumentov.
- ❖ Izvedba in uporabniški vmesnik sta bila pod nivojem, ki ga je že dosegel Gopher.
- ❖ WAISove algoritme za avtomatsko indeksiranje so dolgo uporabljali nekateri sodobnejši spletni iskalniki.

WWW - uvod

- ❖ WWW (World Wide Web), Svetovni splet je:
 - ❖ programska oprema in
 - ❖ mreža dokumentov, dostopnih preko protokola HTTP (Hypertext Transfer Protocol).
- ❖ HTTP je protokol, ki prenese daleč največ podatkov od vseh aplikacij na Internetu.

Predzgodovina: vloga Teda Nelsona

- ❖ Teodor Holm Nelson (*1937), sociolog, filozof in pionir informatike.
- ❖ Bil je eden prvih, ki je opozoril na notranjo strukturo dokumentov in jo predstavil kot informacijski problem.
 - ❖ Dokument je vsebinsko povezana mreža pojmov.
 - ❖ Ta mreža je pogosto ne-zaporedna.
 - ❖ To implicitno strukturo je l. 1963 imenoval hipertekst.
 - ❖ Mrežo, ki bi vključevala tudi ne-besedilne informacije, vključno z zvoki, slikami in videom, je l. 1965 imenoval hipermedia.

Predzgodovina: vloga Teda Nelsona

- ❖ V zgodnjih 70-ih letih začel načrtovati sistem **Xanadu**
 - ❖ sistem za neskvenčno branje in ustvarjanje besedil, ki so razvejana in omogočajo bralcu svobodo izbire pri branju.
 - ❖ Nelson je pisoval Xanadu kot
 - ❖ odprt,
 - ❖ fluiden in
 - ❖ spremenljiv sistem.



Predzgodovina: vloga Teda Nelsona

- ❖ **Xanadu** je vpeljal kazalce kot dvosmerne strukture, ki bi omogočale tudi prehajanje od "citiranega" k "citirajočemu" dokumentu.
- ❖ Končni cilj Teda Nelsona je bil globalni sistem, ki naj bi
 - ❖ povezoval odlomke vseh literarnih del in
 - ❖ omogočal uporabniku njihovo svobodno povezovanje in ustvarjanje novih del.
- ❖ Xanadu je bil daleč pred časom tudi pri strojni opremi. Predvideval je interaktivne zaslone z grafičnim okoljem.

Predzgodovina: vloga Teda Nelsona

- ❖ Xanadu je bil dolga leta "najbolj obetajoča informacijska aplikacija", a ni nikoli prešel eksperimentalne faze.
- ❖ Izreden pomen sistema Xanadu je bil v tem, da je v zavest informacijske srenje vpeljal ideje, ki jih je potem udejanjil Splet.

WWW - uvod

Lastnosti, zaradi katerih je Splet najpopularnejša Internetna aplikacija, so predvsem:

1. Vključevanje hipertekstnih kazalcev na druge dokumente.

S tem je materializirana semantična povezanost vsebinsko sorodnih dokumentov, ki jo sicer mora bralec klasičnih dokumentov izpeljati sam.

WWW - uvod

2. Vključevanje referenc na nebesedilne elemente dokumentov.

Statične in gibljive slike, zvočni posnetki, video posnetki in izvedljivi programi postanejo integralni deli dokumentov.

WWW - uvod

3. "Razumevanje" ostalih najpomembnejših aplikacijskih protokolov na Internetu.

S pregledovalnikom za Splet lahko uporabljamo (ali smo uporabljali) tudi ftp, gopher, telnet, Usenet News in elektronsko pošto.

- ❖ Splet je postal univerzalen vmesnik za informacijske storitve na Internetu.

WWW - uvod

4. Vključevanje programskih vložkov, ki postanejo integralni del spletnih dokumentov, s funkcionalnostjo, ki bistveno presega funkcionalnost običajnih dokumentov.
 6. Sodelovanje skupnosti, ki enakovredno gradi spletne informacijske vire – Splet 2.0.
- ❖ Splet je postavil nove standarde za izrazno moč dokumentov.

WWW - uvod



Tim Berners-Lee, avtor WWW:

- ❖ Obstajajo naloge, ki jih bolje rešujejo računalniki in naloge, ki jih bolje rešujejo ljudje. Med njimi je malo prekrivanja.
- ❖ Računalniki ne znajo postavljati naključnih povezav med različnimi enotami informacije, kar znajo ljudje zelo dobro.
- ❖ Potrebujemo informacijsko okolje, ki bi združilo oba svetova in z računalnikom podprlo človekove asociacijske sposobnosti.

WWW - uvod

Tim Berners-Lee, avtor WWW:

- ❖ “Sanjal sem o skupnem informacijskem prostoru, v katerem bi si ljudje delili informacije.”
- ❖ “Bistvena je bila univerzalnost: hipertekstni kazalec lahko kaže na poljubni informacijski objekt: osebni, lokalni ali globalni.”
- ❖ “Sanjal sem o informacijskem okolju, ki bi odražalo načine, na katere se igramo, delamo in se družimo.”

WWW - zgodovina

❖ 1989

- ❖ Tim Berners-Lee na CERN (Ženeva) predstavi idejo o omrežnem infomacijskem sistemu. Za urejanje dokumentacije CERNa bi sistem izkoriščal prednosti nesekvečnega branja.
- ❖ Prvi projektni predlog je bil neuspešen.

❖ 1990

- ❖ Tim Berners-Lee in Robert Cailliau pripravita nov projektni predlog.

WWW - zgodovina

❖ 1991

- ❖ Januar: prva (zelo enostavna) verzija programske opreme za WWW,
- ❖ Junij: seminar o uporabi WWW za delavce CERN.

❖ 1992

- ❖ Programska oprema pregledovalnika za WWW javno dostopna na Anonymous FTP strežnika na CERN.

WWW - zgodovina

❖ 1993

- ❖ Januar: na spletu 50 spletnih strežnikov.
- ❖ Januar: v NCSA narejen Mosaic, prvi pregledovalnik za WWW, delujoč v grafičnem okolju.
- ❖ September: WWW preseže 1% prometa po hrbtenici NSF.
- ❖ December: 600 – 700 spletnih strežnikov.

WWW - zgodovina

❖ 1994


- ❖ Od 1991 do 1994 se je obremenitev strežnika na CERN vsako leto podeseterila.
- ❖ Maj: 1. mednarodna konferenca o WWW.
- ❖ Junij: na svetu že 1500 - 3000 strežnikov za WWW.
- ❖ Oktober: 2. mednarodna konferenca o WWW.

WWW - zgodovina

❖ 1994

- ❖ September: Tim Berners-Lee ustanovi W3C (World Wide Web Consortium), v katerem sodelujejo MIT (ZDA), INRIA (Francija) in Keio University (Japonska).
- ❖ W3C je neprofitna organizacija, namenjena razvoju Svetovnega spleta - protokolov, aplikacij in novih usmeritev.
- ❖ www.w3.org - odlična startna točka za študij WWW.

WWW - zgodovina

mesec	št. strežnikov za WWW	% iz domene .com
6/1993	130	1,5
12/1993	632	4,6
6/1994	2.738	13,5
12/1994	10.022	18,3
6/1995	23.500	31,3
12/1995	100.000	50,0
6/1996	230.000	 68,0
12/1996	650.000	62,6

Najhitreje razvijajoča se aplikacija v zgodovini

- ❖ Pomemben razlog za eksplozivno uveljavitev Spleta je tudi v tem, da je bil orodje za osebno publiciranje.
- ❖ Tehnologija je bila od vsega začetka, z vidika uporabnika, izjemno enostavna.
- ❖ Izvorni HTML ni zahteval drugega, kot enostaven urejevalnik besedila.
- ❖ Zelo lahko je bilo narediti spletno stran in jo povezati z drugimi spletnimi stranmi.
- ❖ Težave so nastopile s spletišči - spletišče je mnogo več, kot spletna stran ali šop spletnih strani.

Od spletne strani do spletišča

- ❖ Pri gradnji spletišča so pomembne odločitve, povezane
 - ❖ z vsebino - spletišče naj ima neko mero vsebinske skladnosti,
 - ❖ z estetiko - spletišče naj bo vizualno skladno, in predvsem
 - ❖ z arhitekturo spletišča - informacije naj bodo urejene na jasn in enostaven način. To ne velja le za vsebino posameznih strani, ampak še bolj za njihovo povezanost.

Od spletišča do spletne aplikacije

- ❖ Na začetku so si ljudje na spletu delili le statične informacije, shranjene v datotekah.
- ❖ Na spletu je bilo na voljo malo dinamičnih informacij, ki bi jih posredovala informacijska orodja.
- ❖ Izjema so bile informacije, dostopne z Archie v arhivih FTP, na strežnikih storitve Gopher, ali na redkih spletiščih, ki so nudila vremenske napovedi.

Od spletišča do spletne aplikacije

- ❖ Ena prvih spletnih aplikacij je bila iskalnik telefonskih števil na Cernu (avtor TimBL).

coke@cmu

- ❖ Ena najznamenitejših spletnih aplikacij je bila t.i. internetni avtomat za kokakolo.
- ❖ Internetni avtomat za kokakolo je
 - ❖ nastal v zgodnjih 70-ih na Oddelku za računalništvo Carnegie Mellon University (CMU), in
 - ❖ končal kot spletna aplikacija v zgodnjih 90-ih.

coke@cmu

- ❖ Na CMU je bil en sam avtomat za kokakolo in študente je jezilo, da jo je vedno zmanjkalo ravno ko so prišli tja, ali pa je bila topla.
- ❖ V avtomatu so za steklenicami namestili lučke in pred njimi senzorje. Senzorji so šteli lučke, ki jih niso zakrivale steklenice. En senzor je nadzoroval odpiranje vrat. Vrata so se odprla le, kadar so nameščali nove steklenice. Znano je bilo, v koliko urah se steklenica ohladi.
- ❖ Na CMU so predelali internetni program **finger**, sicer namenjen ugotavljanju, če je nek uporabnik prijavljen v omrežje, da je po novem sporočal stanje avtomata za kokakolo.
- ❖ Ukaz **finger coke@cmua** je bil popularen po celem svetu.

Spletni brskalniki: Lynx

- ❖ Lynx je bil prvi resni spletni brskalnik.
- ❖ Narejen je bil 1992 kot brskalnik za Gopher in potem predelan za Splet.
- ❖ Zmogel je komunicirati po številnih internetnih protokolih: Gopher, HTTP, HTTPS, FTP, WAIS, in NNTP.
- ❖ Deloval v znakovnem okolju.
- ❖ Uporabnik je aktiviral kazalce
 - ❖ s premikom utripača (s smernimi tipkami) na poudarjene dele besedila, ali
 - ❖ s tipkanjem ustrezne številke, postavljene v besedilu poleg kazalca.

Lynx, delujoč še 2004

```
xterm
#<<<>>> Lynx (web browser) - Wikipedia, the free encyclopedia (p1 of 5)
#copyright
Your continued donations keep Wikipedia running!
Lynx (web browser)
From Wikipedia, the free encyclopedia
Jump to: navigation, search
CAPTION: Lynx
Wikipedia Main Page displayed in Lynx
Wikipedia Main Page displayed in Lynx
Maintainer: Thomas Dickey
Stable release: 2.8.5 (February 4, 2004) [[+/-]]
Preview release: 2.8.6 (?) [[+/-]]
OS: Cross-platform
Use: web browser
License: GPL
Website: lynx.isc.org
Lynx is a text-only Web browser and Internet Gopher client for use on cursor-addressable, character cell terminals.
Browsing in Lynx consists of highlighting the chosen link using cursor keys, or having all links on a page numbered and entering the chosen link's number. Current versions support SSL and many HTML features. Tables are linearized (scrunched together one cell after another without tabular structure), while frames are identified by name and can be explored as if they were separate pages.
Lynx is a product of the Distributed Computing Group within Academic Computing Services of the University of Kansas, and was initially developed in 1992 by a team of students at the university (Lou Montulli, Michael Grobe and Charles Rezac) as a hypertext browser used solely to distribute campus information as part of a Campus-Wide Information Server. In 1993 Montulli added an Internet interface and released a new version (2.0) of the browser [1] [2] [3].
more- http://en.wikipedia.org/wiki/Image:Lynx_228web_browser229.png
```

WWW

- ❖ Enota informacij na WWW je dokument ali spletna stran.
- ❖ WWW temelji na hipertekstu, zato je nelinearen in nehierarhičen.
- ❖ Ne obstaja privilegirana vstopna točka v celotno omrežje.
- ❖ Vsaka spletna stran, ki ima vsaj en kazalec na drug dokument, je lahko vstopna točka.

WWW

- ❖ Obstajajo privilegirane vstopne točke v organizacijsko omejene dele omrežja:
 - ❖ domača stran (homepage) je načeloma enakovredna katerikoli drugi, vendar je zaradi svoje strukture primerna za vstopno točko v del omrežja, ki opisuje eno osebo, inštitucijo ali storitev,
 - ❖ spletni portali so privilegirane vstopne točke v storitev ali vsebinsko omejeno zbirko kazalcev in storitev.

Spletni gradniki

Osnovni gradniki Spleta so

- ❖ **HTML** - označevalni jezik za "pisanje" spletnih dokumentov,
- ❖ **URI** - enotna shema naslavljanja omrežno dostopnih dokumentov,
- ❖ **HTTP** - aplikacijski protokol, namenjen transportu spletnih sporočil.

URI

- ❖ Naslavljanje spletnih strani je osnova delovanja Spleta.
- ❖ Naslov (identifikator strani) bi moral biti neodvisen od sprememb v konfiguraciji omrežja.
- ❖ Uveljavil se je **URL** - Uniform (izvorno Universal) Resource Locator, ki je odvisen od lokacije.
- ❖ Prvotna zamisel neodvisnega identifikatorja - **URN** - Uniform Resource Name se ni dovolj uveljavila.
- ❖ W3C vpeljal skupno ime **URI** - Uniform Resource Identifier, ki zajema URL in URN.

Splošna shema URL

Splošna shema URL, veljavna tudi za dinamične aplikacije:

**scheme://host[:port#]/path/file_name/
[;url_parameters][?query_string][#anchor]**

- ❖ url_parameters: parametri, ki si jih brskalnik izmenjuje z nekim programom na strežniku (t.i. servlet-om)
- ❖ query_string: parametri, ki jih brskalnik pošilja v neko podatkovno zbirko (rezultati vnosa v spletni obrazec), ali v iskalnik (iskalna zahteva).
- ❖ anchor: oznaka pozicije v zahtevanem dokumentu, ki omogoča skok na določeno mesto v dokumentu.

HTTP - Hypertext Transfer Protocol

- ❖ Protokol (navodila) za izmenjavo spletnih sporočil.
- ❖ Zelo enostaven protokol, kar je njegova moč in pomanjkljivost.
- ❖ Zaradi enostavnosti je zavzel svet, sorodni, zapletenejši protokoli (npr. Xanadu) pa so propadli.
- ❖ Protokol temelji na načelu zahteva/odgovor.
- ❖ Odjemalec in strežnik lahko komunicirata po protokolu HTTPS - tajnopisno kodiranem prenosu po HTTP).

HTTP je "stateless"

HTTP je "stateless".

- ❖ Odjemalec pošlje zahtevo, strežnik pošlje odgovor in s tem je ta komunikacija zaključena.
- ❖ Pri HTTP je življenjska doba komunikacije en sam par zahteva/odgovor.

Če je protokol "stateful",

- ❖ omogoča trajnejšo komunikacijo med odjemalcem in strežnikom.
- ❖ Omrežje se zaveda partnerjev v komunikaciji, omogoča zaporedje akcij v seansi, dokler komunikacija ni izrecno zaključena.
- ❖ Številni Internetni protokoli so "stateful", vključno s FTP in SMTP (e-pošta).

HTTP:

- ❖ Ni mogoče združiti več zahtev in naenkrat prenesti sestavljene spletne strani.
- ❖ Posebej se zahteva in prenese osnovno besedilo in posebej vsaka od slik ali drugih nebesedilnih delov, ki se spet sestavijo na odjemalčevem zaslonu.
- ❖ To je nerodno, a omogoča, da vsak del pride z drugega strežnika poljubne oddaljenosti.

Piškotki (cookies)

- ❖ Kako s "stateless" protokolom izvajati "stateful" aplikacije?
- ❖ Treba je najti način, da bi si partnerja v komunikaciji, poleg informacij, ki so del komunikacije, izmenjevala tudi informacije o "stanju".
- ❖ Tak dogovor se imenuje Cookie.
- ❖ Strežnik pošlje odjemalcu sporočilo **Set-Cookie** in identifikator seanse.
- ❖ Odjemalec pošlje z vsako zahtevo v seansi še sporočilo **Cookie** in identifikator te seanse.

Shema piškotka

```
Set-Cookie: name=value  
[; expires=date]  
[; path=path]  
[; domain=domain_name]  
[; secure]
```