INTERNET QUALITY CIRCLE

OTDK dolgozat

Minőségfogalom az Interneten

XXV. Országos Tudományos Diákköri Konferencia

Minőségfogalom az Interneten
1.0 Verzió

Fejlesztők:
Erdei Szabolcs
Roszik János

Témavezető:
Bátfai Norbert

A technikai megjegyzéseket a következő címre várjuk:
kron@venus.math.klte.hu

Chronos

Choronos weblap
http://kron.sourceforge.net
Choronos összefoglaló
http://sourceforge.net/projects/kron/

Debreceni Egyetem, Természettudományi Kar, Matematikai és Informatikai Intézet
JFC Enterprise - J Group
jfc@venus.math.klte.hu
http://venus.math.klte.hu
Debrecen, 2001 Január 11.

Tartalomjegyzék

1. Bevezetés

Tisztelt Olvasó!

Engedje meg, hogy néhány szóban bemutassuk az általunk kigondolt, és fejlesztés alatt álló CHRONOS rendszert. CHRONOS

A utóbbi években a WORLD WIDE WEB olyannyira bonyolulttá és nagyméretűvé vált, hogy az értékes és minőségi információk megtalálása egyre nehezebb, gyakran szinte lehetetlen. Ezt a jelenséget nevezzük információs túlterhelésnek. A problémát a felismerése óta eltelt időben többen, többféleképpen próbálták megoldani. Sorra jelentek meg a keresők, webdirectory-k és portálok, melyek kezdőlapként segítik ugyan az elindulást, de ezután semilyen segítséget nem tudnak adni a böngészőknek. Időközben a web méretének növekedése tovább gyorsult, amivel a keresők sem tudtak lépést tartani. A frissítés lelassult, a keresések eredménye gyakran áttekinthetetlenné vált. Javulást csak a második generációs keresőmotorok megjelenése hozott, melyek fejlesztésekor már a web jelenlegi struktúráját feltáró kutatások eredményeit is figyelembe vették. *"A régi gépekkel az volt a baj, hogy sohasem azt csinálták, amire utasították őket. A modern gépekkel az a baj, hogy pontosan azt csinálják, amit mondunk, és nem azt, amit mondani akartunk."* - mondta Kemény János a BASIC megalkotója. Örök érvényű megállapítása a keresőkre is megállja a helyét. Lassan már ahhoz, hogy egy keresés eredménye még kezelhető legyen, komoly logikai ismeretekre van szükség, hogy jól meg tudjuk fogalmazni a kérdésünket. Különben találatok ezrei árasztanak el bennünket felhasználókat.

Az általunk készített CHRONOS rendszerrel egy új szolgáltatásként a linkeket lehet validálni és a weboldalakat minősíteni. Ezt úgy kívánjuk elérni, hogy az oldalakhoz tulajdonságokat rendelünk, amit az oldal készítője befolyásol. Ezen tulajdonságok értékei ezután az oldalt látogató felhasználók számától és böngészési szokásaitól függően változnak. Alapvető kérdések: minősítés validálás

- Miért jó a felhasználóknak a CHRONOS használata?
- Mire lehet használni a CHRONOS alaprendszert?
- Miért gondoljuk, hogy világszerte elterjedhet a CHRONOS?

Ha valaki a CHRONOS-on keresztül böngészik, mintha egy prizmán keresztül látná az Internetet, akkor olyan plusz információkhoz jut, amik segítik a linkek kiválasztásakor. Például egy adott oldalon a linkek melletti kis piktogrammal jelezni tudjuk, hogy létezik-e az az oldal, amire a link hivatkozik, és ha létezik, akkor a „minőségére" is utalni tudunk. Mindezt úgy valósítottuk meg, hogy a böngésző felhasználóktól semilyen plusz tevékenységet nem igényel, az eddigi böngészési szokásokon nem kell változtatni.

A CHRONOS számtalan értéknövelő szolgáltatásnak lehet az alapja. Például statisztikai alkalmazások építhetők rá, amiből egy adott webszerveren tárolt oldalakról kaphatunk látogatottsági információkat. A legérdekesebb alkalmazás a CHRONOS-ra alapozott keresés. Egy ilyen keresésnél a mostani keresőknél bevált kifejezéseken kívül megadhatunk plusz megkötéseket is, például az oldal kategóriájára, utolsó módosítására vagy akár a látogatottságára vonatkozóan. A keresés eredményét pedig aszerint rendezhetjük, hogy az adott oldal mennyire „jó" az adott kategóriában. De ez a két példa csak egy kis része a számtalan lehetőségnek.

keresés

Az alap CHRONOS rendszert és forráskódját mindenki számára elérhetővé és letölthetővé tesszük, ezzel biztosítva a rendszer elterjedését. E célból került a nyitott forráskódú fejlesztés fellegvárának számító sourceforge.net-re is a CHRONOS, GNU Lesser General Public License (LGPL) engedéllyel. Elosztott rendszer lévén a CHRONOS annál hatékonyabb, és annál jobban szolgálja a böngészők igényeit, minél több webszerverre van telepítve[1]. A CHRONOS rendszer nyújtotta szolgáltatásokat élvező böngésző felhasználók a Hálón való szörfölés közben észre sem veszik, hogy böngészésükkel a CHRONOS mélyén lapuló matematikai modellt is táplálják. Erre a dinamikai rendszerre épülnek majd kereső szolgáltatásaink.

matematikai modell

Szerentnénk leszögezni, hogy a CHRONOS rendszer a saját ötletünkön alapul, és az alaprendszer fejlesztését teljes egészében magunk végeztük. Álmunk egy sokkal felhasználóbarátabb és jobban használható Internet. Úgy gondoljuk, hogy az ehhez vezető hosszú út egy lépése lehet a CHRONOS rendszer.

[1] A CHRONOS rendszer telepítése nem feltételez privilegizált felhasználót

2. A Chronos rendszer működése

2.1. A Szerver működése

A CHRONOS szerver indulásakor a futtató gépen lefoglalja a CHRONOS portot, és az ide érkező kéréseket továbbítja a webszervernek. A webszerver válaszát feldolgozza, és a feldolgozás eredményét pedig elküldi a böngészőnek.

1903 a CHRONOS kapu

A feldolgozás során a nem `text/html` tartalmú dokumentumokat módosítás nélkül küldi el. A `text/html` típusú tartalmú oldalakon talalható abszolút hivatkozású linkekbe befűzi a CHRONOS szerver portszámát, ha a hivatkozás protokoll típusa HTTP, nincs megadva port szám és az adott gépen fut CHRONOS szerver. Így az egyes linkek kiválasztása esetén az a dokumentum is feldolgozásra kerül amire a link hivatkozik. Relatív linkek esetén a böngésző egészíti ki a címet. Ha az aktuális lapot a CHRONOS porton keresztül kértük le, akkor a relatív linkek is ezzel a portszámmal lesznek kiegészítve kiválasztás esetén.

A feldolgozás gyorsítása érdekében a relatív képhivatkozásokat kiegészíti abszolút hivatkozássá, így azok nem a CHRONOS szerveren keresztül töltődnek le.

A port szám befűzésén kívül a letöltött lappal azonos gépen lévő dokumentumokra hivatkozó linkek elé elhelyez egy kép hivatkozást, attól függően, hogy az adott linket a CHRONOS szerver milyennek minősíti. Ha valamilyen típusú dokumentumra mutató linkeket - például bináris fájlok - nem akarunk minősítéssel ellátni, akkor azt a CHRONOS szerver indítása előtt a konfigurációs fájlban megadhatjuk. Az ilyen link hivatkozásokba a port számot sem fűzi be. Más gépeken lévő dokumentumokra mutató linkek esetén egy a CHRONOS szerver futását jelző kis képre mutató hivatkozást helyez el a link elé.

A linkminősítési szempontok:

- A link által hivatkozott dokumentum utolsó módosításának ideje
- 30 napra visszamenőleg a letöltők számának változása
- A dokumentumhoz tartozó, minőségre utaló értékek

A dokumentumhoz tartozó, minőségre utaló értékek CHRONOSON keresztül nézett HTML lapon CHRONOS által minősített dokumentumra mutató link kiválasztása esetén egy matematikai modell alapján módosulhatnak.

Ezeknek a dokumentumoknak nem szükséges ugyanazon a gépen lenniük, a Chronos szerverek ilyen dokumentumok találkozása esetén kommunikálnak egymással, elküldik egymásnak a szükséges információkat.

A részletesebb megértést segíti a Függelék magyarázó ábrája.

2.2. Statisztikák

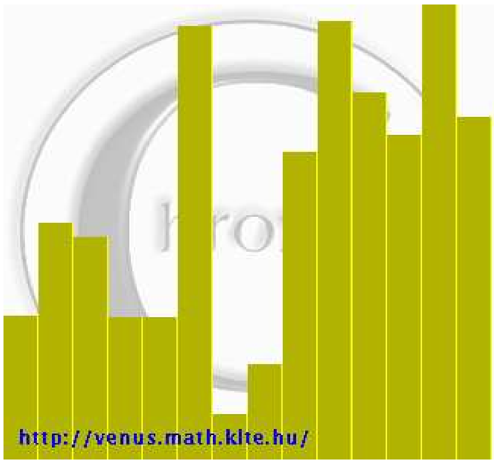
A Chronos alaprendszer által szolgáltatott adatokból a statisztikai modul segítségével változatos kimutatások készíthetők. Az ilyen táblázatok és diagrammok segítségével a weboldalak tulajdonosai tájékozódhatnak az oldalaik népszerűségéről és látogatottságáról. A Chronos így képes a mostani web szerverekhez csatlakoztatható statisztikákat készítő csomagokat teljes egészében kiváltani. De ennél többre is, hiszen a statisztikai modul részére rendelkezésre állnak a látogatottságon kívül a dinamikai modell által szolgáltatott adatok is. Így a táblázatokban szerepelhet az egyes oldalak különböző kategóriákbeli jósága, és a frissítésre utaló linkminősítés is. Lehetőség van arra is, hogy az oldalakat egy összetett tulajdonság alapján rangsoroljuk, mely függ a látogatottságtól, a látogatottság változásától, az oldal saját kategóriáiban elért jóságától és az utolsó frissítés dátumától. Természetesen mindez tetszés szerint konfigurálható és súlyozható. A statisztikák frissítése naponta egyszer, éjszakánként történik, amikor vélhetőleg a legkisebb a szerver terheltsége. A honlapok tulajdonosai ezeket a statisztikákat böngészve sokkal árnyaltabb képet kaphatnak a lapjuk népszerűségéről, mintha csupán a látogatottsági adatokra támaszkodnának. Ez segíthet a helyes következtetések levonásában és a helyes döntések meghozatalában.

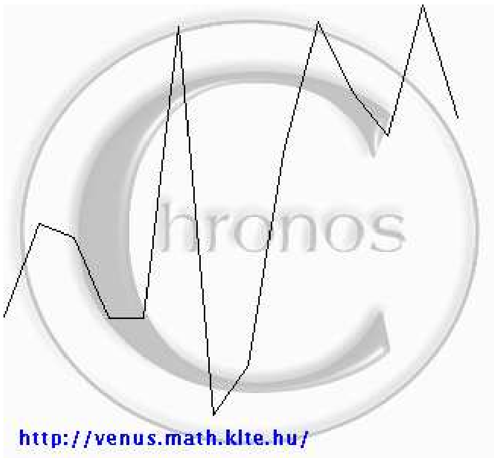
A leglátogatottabb oldalak: 2000 November					
No.	Találatok		URL	Kategóriák	Max.
1	4751	2,18%	/~denevito/capture.mpg		
2	4179	1,92%	/nin.html	java, j csoport	6,4
3	3407	1,56%	/www/		
4	3364	1,54%	/~alice/	MI, chatrobot, AIML	7,4
5	3337	1,53%	/~jg/	java, kutatás, fejlestés	5,2
6	3226	1,48%	/java/summer/	java	3,9
7	3130	1,44%	/wayfarer/Letoltes/konyvtar	download	2,1
8	2682	1,23%	/java/casestudies/	java, oktatás	2,5
9	2127	0,98%	/pub_NOVEL_ON_THE_MOVE/pillanatfelvetelek.html	wap, mobil, játék	7,0
10	2006	0,95%	/novdev.html	wap, mobil, játék	7,0

Az 1. ábrán egy adott hónap 10 leglátogatottabb oldala látható kiegészítve azzal, hogy az oldal milyen kategóriákba tartozik és legjobb kategóriájában mennyire jó. A jóság az $1 - 10$ intervallumra van leképezve.

A legjobb java kategóriájú oldalak				
No.	Pontszám	URL	Találatok	
1	8,2	/	1366	0,63%
2	7,9	/~voice/	2006	0,92%
3	7,4	/javanews/show	1810	0,83%
4	6,4	/nin.html	4179	1,92%
5	5,2	/~jg/	3337	1,53%
6	4,2	/javalinks.html	1794	0,82%
7	3,9	/java/summer/	3226	1,48%
8	2,5	/java/casestudies/	2682	1,23%
9	2,5	/java/summer/Laszlo.Nagy_WG_4/	1826	0,84%

A 2. ábrán a 10 legjobb java kategóriájú oldal látható.





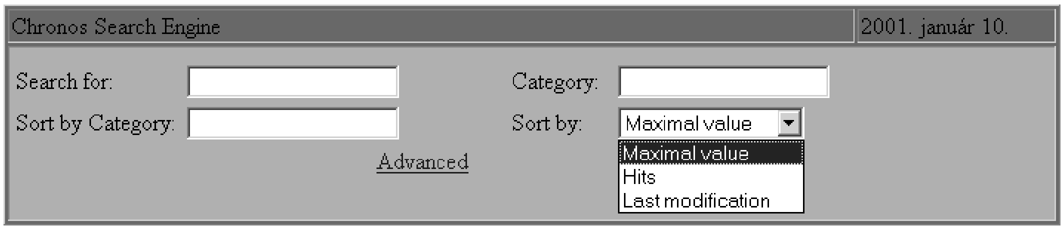
A két ábra a venus.math.klte.hu webszerver fő lapjának java tulajdonságának a változását mutatja az elmúlt 2 hétben. Az ábra viszonyított értékeket tartalmaz, ezért a koordináta tengelyek ábrázolásától most eltekintettünk.

2.3. Kereső szolgáltatások

Minden olyan webhely, amin sok információ található előbb utóbb szembekerül azzal a problémával, hogy a felhasználók egyre kevésbé igazodnak el az oldalak között és sokkal lasabban találják meg a keresett információkat. Egy hirdetésekből élő cégnél ez akár még előny is lehet, de egy magára valamit is adó webmester figyel arra, hogy az internetező közönség igényeit kiszolgálja. Az eligazodást segítheti többszintű menürendszer, a tartalom racionalizálása, weblap-térkép (sitemap) vagy kereső elhelyezése. A keresést egy adott web szerveren képes a Chronost is elvégezni. És egy átlagos keresésnél többre is képes, hiszen itt ismét fel lehet használni az alaprendszer adatait. Például megadható, hogy csak mely kategórián belül akarunk keresni. Ilyenkor a találatokat az oldalak jósága szerinti sorrendben kapjuk meg, de lehetőségünk van más szerint is rendezni a keresés eredményét.

Gráfokat már számos biológiai és mesterséges rendszerben találtak, melyek "small world" hálózatok. Ilyen például az energiahálózat az Egyesült Államok nyugati részén vagy a Caenorhabditis Elegans féreg ideghálózata, ami az egyetlen teljesen feltérképezett ideghálózat. De gondolhatunk nyugodtan az Erdős számra vagy a filmszínészek együttműködési gráfjára is.

Lada A. Adamic a The Small World Web cimű cikkében bemutatja, hogy egy újabb mesterséges, ember által készített hálózat a Word Wide Web is "small world" topológiájú. Ennek a topológiának következményei vannak a szöföző felhasználó útját és az információ összegyűjtésének egyszerűségét tekintve. A linkek struktúrája plusz információt szolgáltat az emberek közötti alapvető kapcsolatokról, érdeklődésükről és a közösségekről. Cikkében Lada A. Adamic bemutatta, hogy kereséseket hogyan lehet jobbá tenni a szorosan összefüggő komponensek (SCC) megkeresésével és vizsgálatával. Részben ezt alkalmazza Google kereső is. Mi is ezt használtuk fel kersőnkben, hogy lehetőleg még jobb eredményeket adhassunk. Persze ezt a plussz funkciót csak akkor lehet kihasználni, ha az adott szerveren sok weboldal található vagy ha a Chronos egy egész tartományra van feltelepítve.



Az ábrán a kerső beviteli formja látható.

2.4. Biztonsági kérdések

A KRONOS rendszer lelke a webes erőforrásokhoz rendelt numerikus értékek egymásrahatása. A szóbanforgó, az adott erőforrást jellemző numerikus értékeket lokálisan az „adott erőforrás mellett” tároljuk. Az INTERNET hajnalán ennyi még elég is lett volna ... De ma mi akadályozhatná meg a legkülönfélébb motivációjú (akár jogosult) felhasználókat, hogy ne szépítsenek a rendszer találkozási algoritmusa determinálta numerikus értékeken?

Melyek a felmerülő problémák?

- Biztosítani, hogy a találkozások alkalmával érvényes adatok hassanak egymásra.
- Biztosítani, hogy az INTERNET közössége számára egy webes erőforrás hiteles adatokat mutasson magáról.

Ezeket mérlegelve kell meghúznunk a határokat a biztonság és a hatékonyság között.

Jelen fejlesztési fázisban ezek a kérdések még nincsenek véglegesen és megnyugtatóan lezárva. Az első problémához elég annyi, hogy a találkozás egy — funkcionalitásában a névszerverekhez hasonló — kitüntetett gépen történik. A KRONOS szájtokon pedig aláírt formában tároljuk a numerikus értékeket becsomagoló objektumokat.

A második probléma kézenfekvő megoldása (attól eltekintve, hogy az adatokat ne lokálisan tároljuk), hogy a KRONOS szájt által önmagáról publikált adatok mellett mindig szerepeltetünk egy hitelesítés gombot. Erre kattintva az előbbi pontban említett kitüntetett gépre eljut az információ, hogy mely adat hitelesítését kérték. A kitüntetett gép letölti az aláírt, numerikus értékeket becsomagoló objektumot. Ellenőrzi az aláírást és visszaküldi az érvényes adatot.

3. A Találkozás

Amikor egy felhasználó egy CHRONOST futtató szerveren levő weboldalt böngészik, akkor megvizsgáljuk, hogy a felhasználó által előzőleg betöltött weboldal CHRONOSOS szerverről származik-e. Ha igen, akkor következik a találkozás (meeting). A találkozás alatt azt az eseménysort értjük, melynek során két weboldal tulajdonságait tartalmazó objektum kölcsönhat egymással. Először a most böngészett oldalhoz tartozó objektumot küldjük az előző oldalt tartalmazó szerverre, hogy hathasson az ottani objektumra, majd azt vissza. Ez a folyamat teljesen automatikus és a háttérben történik, így a web szerverek válaszidejét nem növeli, a felhasználókat pedig nem terheli. A rendszer egyik legfontosabb része a matematikai modell, ami alapján a két oldal tulajdonságai egymásra hatnak. Ez akkor a legegyszerűbb, ha a két oldalnak nincs közös tulajdonsága: ilyenkor nem történik semmi, nem hatnak egymásra.

találkozás

De milyen tulajdonságok tartoznak egy honlaphoz?
A CHRONOS rendszer azokat a tulajdonságokat rendeli egy honlaphoz, amelyeket a készítője kulcsszóként megad. Minden tulajdonsághoz egy valós számérték tartozik. Egy találkozás alkalmával ezek az értékek változhatnak. A weboldalak minőségéről bármit is mondani borzasztóan nehéz, hiszen a jóság az egyik legszubjektívebb fogalmunk. Ezt végiggondolva a minősítés alapjául szolgáló matematikai modellt eleve úgy építettük be a CHRONOSBA, hogy bármikor cserélhető legyen. A CHRONOS első indulásakor egy központi helyről letölti a legfrissebb modellt. Ez a modell magában hordozza azt az információt, hogy mikor kell újra a központi szerverhez fordulni új modell iránt érdeklődve. Így az elosztott rendszer matematikai modellje központilag frissíthető.

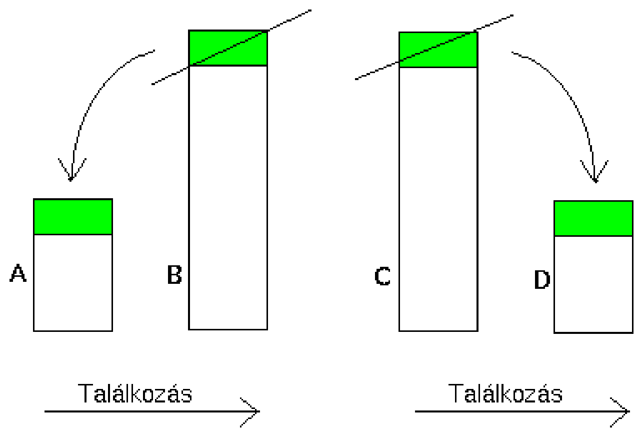
A CHRONOS rendszer fejlesztése során a Jószív referencia modellt használjuk a találkozások során.

referencia modell

4. Modellek

4.1. A Jószív referencia modell

A találkozás a különböző webszájtok lapjai között értelmezett fogalom. Tehát egy CHRONOS rendszerrel felszerelt webszájton történő böngészés során nem kerül sor találkozásokra. A következő ábrán egy CHRONOS rendszerrel felszerelt webszájton lévő lapról ugrunk egy másik – szintén a CHRONOS rendszerrel felszerelt webszájton lévő – lapra. Ekkor sor kerül a találkozásra:



Feltesszük, hogy vannan közös kategóriák; egy ilyet ragadtunk ki. Két lényegi lehetőség van: az a lap, ahonnan jöttünk gyengébb a szóbanforgó kategóriában, mint az, ahova ugrottunk; illetve ennek megfordítása.

A Jószív modellben mindig az erősebb lap "ad át" erejéből a gyengébbnek. Ennek verbális indoklása a következő lehet:

A : jó lapra ugrás javulást eredményez

B : rosszabbak belinkelte lap színvonala esik

C : rosszabb lapra ugrás romlást eredményez

D : jobbak belinkelte lap színvonala emelkedik

4.2. Szimulációk

A jószív modell kiválóan alkalmas a modell szerepének szemléltetéséhez, de — bár verbális indoklása meggyőző — a gyakorlati kipróbálás során nem mutatott megfelelő eredményeket. Mert kellően sok találat után a rendszerben levő oldalak tulajdonságai kiátlagolódnak. Ez könnyen látható, hogy nem jó hiszen a valóságban sem egyformán jók a weboldalak. A CHRONOS mélyén levő dinamikai rendszernek tehát egy olyan modellre van szüksége, ami meg tudja ragadni a böngészés folyamatát és kellőképpen ki tudja fejezni a böngészők értékrendjét. Azt posztuláljuk, hogy a modell üzemének eredménye olyan legyen, hogy: nagyon kevés lap nagyon jó minőségű, a lapok többsége nagyon gyenge minőségű és e kettő között van egy átmenet. Egy ilyen modell megtalálása rendkívül nehéz és hosszantartó munkát igényel, ezért a Debreceni Egyetem matematikusainak a segítségét kértük. A következő nehézség, amivel szembetaláltuk magunkat, hogy a modellek vizsgálatához szükségünk van egy olyan szimulációra amely kellőképpen tudja modellezni a Web-et és a böngészést. Ehhez a Web alapstruktúráját feltáró kutatások eredményeit vettük figyelembe.

A tanulmány nem olyan érdekes csupán az eredményeket figyelve, viszont nagyon is hasznos olyan algoritmusok fejlesztésekor, amelyek a web egy részét akarják feltérképezni, keresést akarnak végezni, illetve valamilyen közösséget akarnak megfigyelni, illetve elemezni fejlődésének jellemzőit. A jelentés a web grafikus struktúrájának lokális és globális jellemzőit hivatott leírni, felhasználva az Altavista két rekurzív keresésének eredményét. Mindkettő 200 milliónál több lapot tartalmazott, illetve több mint 1.5 billió linket. Eme írásból kiderül, hogy a web makroszkopikus struktúrája sokkal bonyolultabb, mint azt a korábbi kutatások sugallták, melyek kevesebb oldallal és linkkel dolgoztak. A web grafikus ábrázolásán azt az irányított gráfot értjük, melynek csúcsai a statikus lapok, és élei a linkek. A gráf irányított, ugyanis a lapok nem kölcsönösen hivatkoznak egymásra. Megvizsgálták eme gráf jellemzőit, ide értve az átmérőjét, fokszámok eloszlását, összefüggő komponenseit és makroszkopikus struktúráját.

A vizsgált részgráf nagysága kb. ötször akkora, mint a legutóbbi legnagyobb kutatás által feldolgozott gráf. A legújabb kutatások [1] azt sugalják, hogy a pontok fokainak (degree) eloszlása az erő törvényt (power law) követi:

Az erő törvény bemenő élekhez tartozó fokokra (in-degree): annak a valószínűsége, hogy egy pont i belső fokszámú, egyenesen arányos $1/i^x$ kifejezéssel, valamely pozitív $x > 1$-re.

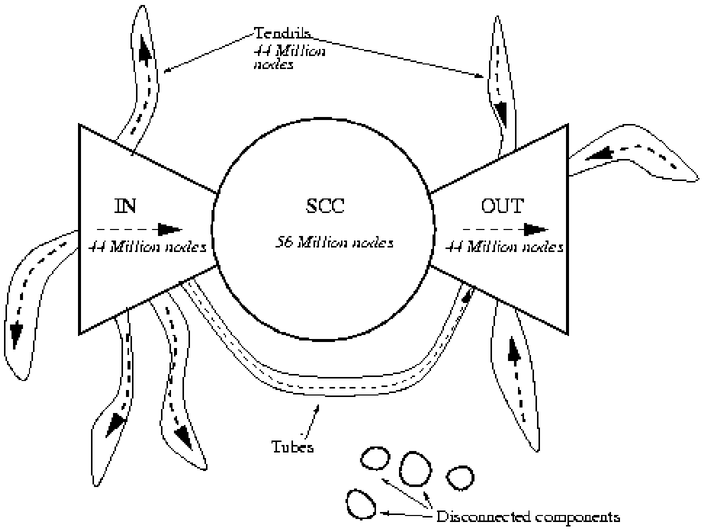
Bebizonyították, hogy az erő törvény helyesnek bizonyult ilyen nagy részgráf esetén is, bizonyítva, hogy ez egy alapvető jellemzője az Internetnek.

Egy korábbi cikk közli azt az érdekes felfedezést, hogy majdnem minden lapot csupán egy maréknyi link választ el egymástól, és ez a szám minden esetben 20 alatt van, és azt állítja, hogy ez a szám a web nagyságának növekedésével logaritmikusan fog nőni. Ezt a megállapítást *small world* jelenségnek is nevezik. Jelen cikk sokkal részletesebb és finomabb képet ad: a web lényeges részei nem érhetőek el más lényeges részekből, és jelentős számú lappár áthidalható, de csak a használt útvonalak haladják meg a többszáz közbeeső lapot.

1999 májusától októberéig három csoportját hajtották végre a begyűjtött oldalak elemzésének. Először legenerálták a be- és kimenő fokszámok eloszlását, bebizonyítva az erő törvényt. Második körben megvizsgálták a web közvetlenül és közvetetten kapcsolódó komponenseit. Megmutatták, hogy az erő törvény megmutatkozik ezen kapcsolódó komponensek méreteinek eloszlásában is. Végül végrehajtottak számos szélességi keresést véletlenszerűen választott kezdő csúcsokból.

Az elemzés egy igen érdekes képet nyújtott a web makroszkopikus szerkezetére vonatkozóan. A legtöbb csúcs (90%-nál több) a hozzávetőlegesen 203 millió csúcs közül egy egyszeresen összefüggő komponenst formáz, ha a hiperlinkeket irányítatlan éleknek tekintjük. Ez az összefüggő komponenes 4 részből áll. Az első darab a központi mag, az összes olyan lap, amelyek közvetlenül elérhetőek egymásból irányított hiperlinken keresztül – ez a "giant strongly connected component" (SCC), azaz "óriás erősen összefüggő komponens", a web szíve. A második és a harmadik darab az IN és OUT nevet kapta. Az IN által tartalmazott lapok elérik az SCC-t, de nem érhetőek el abból – talán az új lapok, amelyeket még nem fedeztek fel, és nem linkeltek be. OUT tartalmazza azokat a lapokat, melyek elérhetőek az SCC-ből, de nem mutatnak vissza. Ilyenek az olyan szervezet lapjai, melyek csak belső linkeket tartalmaznak. Végül a TEDRIL-ek (indák) azokat a lapokat tartalmazzák, melyek nem érhetőek el a SCC-ből, és az SSC sem érhető el róluk. Talán a legmeglepőbb eredmény, hogy az SCC mérete aránylag kicsi – kb. 56 millió lapot tartalmaz. A másik három halmaz egyenként kb. 44 millió lapot tartalmaz, azaz mind a négy komponenes durván ugyanannyi lapból áll.

Megmutatták, hogy az SCC átmérője kevesebb, mint 28, és a teljes gráf átmérője nagyobb, mint 500. Ezenkívül meghatározták, hogy véletlenszerűen választott forrás- és céllapra, annak a valószínűsége, hogy létezik őket összekötő irányított út, 24%. Ha létezik ez az út, akkor annak átlagos hossza kb. 16. Hasonlóképpen, ha irányítatlan út létezik, akkor az átlagos hossza 6. Az Altavista két gyűjtését használva ezek az eredmények megegyeztek.



Az ábra Andrei Broder, Ravi Kumar, Farzin Maghoul, Prabhakar Raghavan, Sridhar Rajagopalan, Raymie Stata, Andrew Tomkins, Janet Wiener[1] cikkéből származik.

Ezeknek az eredményeknek az ismeretében alakítottuk ki saját szimulációs környezetünket. Ez egy C nyelven implementált alkalmazás, amelyben rugalmasan vizsgálható a felmerülő modellek hosszútávú viselkedése.

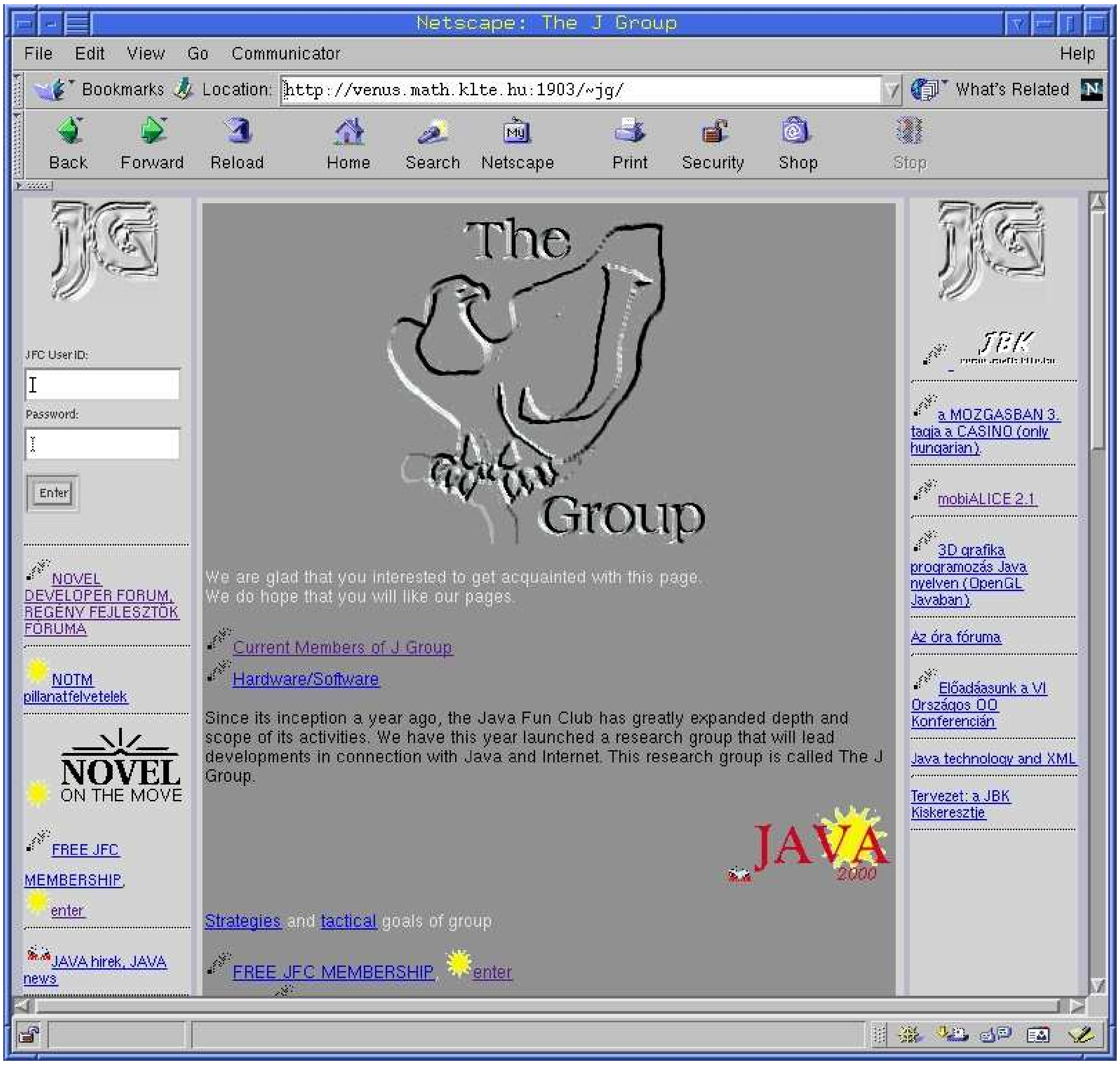
5. A Chronos jövője

A fejlesztést 2001 januártól kezdve a Sourceforge.net-en folytatjuk. Így lehetőség nyílt arra, hogy bárki a világ bármely tájáról bekapcsolódhasson a fejlesztésbe. Ez teljesen öszhangban van a nyílt forráskódú fejlesztéssel. Reményeink szerint ez nagyban hozzá járulhat ahhoz, hogy minél többen próbálják ki és ismerkedjenek meg a CHRONOS-sal.

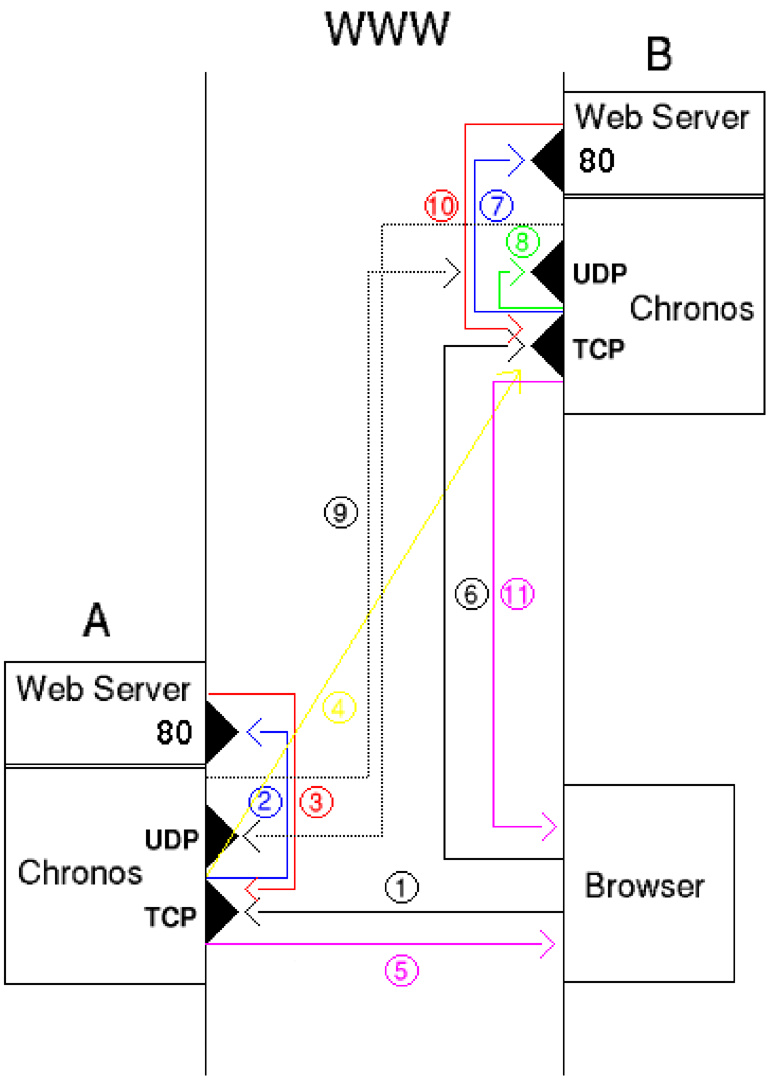
Jelenleg az alkalmazásunk csak a Debreceni egyetem szervrein fut, ezeken a gépeken teszteljük. Mivel a rendszer nagy részének a fejlesztése végső stádiumhoz érkezett ezért szeretnénk még januárban egy bárki számára letölthető verziót kibocsátani. Igy az egyetem falain kívül is megkezdődhetne az alkalmazása.

Az elkövetkezendő időszakra is vannak ötleteink. A következő félévben azt szeretnénk megvizsgálni, hogy lenne-e igény egy olyan globális keresőre, amelyel a teljes Web-en lehetne keresni, és nem csak egy adott szerveren, és a keresés eredményében, ha az adott oldal egy CHRONOS-t futtató szerveren található, akkor a hozzá tartozó plusz információkat is megjelenítené. Ezt próbaképpen - szükös erőforrásainkhoz mérten - a webszerverek csak egy kis csoportján tervezzük csak implementálni és kiprobálni pl.: a klte.hu tartományon belül. Ez, ha elkészülünk vele, reméljük sok embert meg fog győzni a CHRONOS hasznosságáról.

6. Pillanatkép



7. Függelék



1 : A böngésző lekér egy lapot a CHRONOS szervertől.

2 : A CHRONOS szerver továbbítja a kérést a webszervernek.

3 : A webszerver elküldi a CHRONOS szervernek a kért lapot.

4 : A Chronos szerver feldolgozza ezt, megvizsgálja hogy melyik belinkelt gépen fut Chronos szerver.

5 : Elküldi a feldolgozás eredményét a böngészőnek.

6 : A böngésző személy kiválaszt egy linket ezen a lapon, lekéri a lapot a Chronos szervertől.

7 : A Chronos szerver továbbítja a kérést a webszervernek.

8 : A Chronos szerver értesíti az adatfeldolgozó szálat.

9 : Az adatfeldolgozó szál felveszi a kapcsolatot a másik Chronos szerverrel.

10 : A webszerver elküldi a Chronos szervernek a kért lapot.

11 : A Chronos szerver feldolgozza ezt, majd elküldi a feldolgozás eredményét a böngészőnek.

8. Köszönetnyilvánítás

Köszönjük a Debreceni Egyetem Alkalmazott Matematika és Valószínűségszámítás Tanszékének, hogy informatikai infrastruktúráját a rendelkezésünkre bocsátotta. Továbbá, hogy munkánkat a modellt illető tanácsaival folyamatosan támogatja.

Köszönjük Horváth Balázs hallgatónak a logókkal kapcsolatos munkáját. Továbbá Rab Zsigmondnak a linkminősítés új ikonjainak elkészítését.

Hivatkozások

[1] **Andrei Broder, Ravi Kumar, Farzin Maghoul, Prabhakar Raghavan, Sridhar Rajagopalan, Raymie Stata, Andrew Tomkins, Janet Wiener.** Graph structure in the web, 2000. *http://www.almaden.ibm.com/cs/k53/www9.final/*

[2] **Lada A. Adamic** The Small World Web, 1999. *http://www.parc.xerox.com/iea*

[3] Google. *http://www.google.com*

[4] Sourceforge.net. *http://www.sourceforge.net*

....	....	....
Bátfai Norbert témavezető	Erdei Szabolcs hallgató	Roszik János hallgató