

Web - világ

Magyar tenger - a web hazai világa

Több mint 5 millió, de kevesebb, mint 40 millió magyar site van. Átlagosan 9 kattintással eljuthatunk az egyik hazai oldalról bármelyik másikra. Ragozni egyik kereső sem tud. Apró cseppet tesz ki csupán a magyar oldalak tengere a világháló óceánjában. De mekkora is ez a csepp? Akárcsak a teljes világháló méretével kapcsolatos kérdésekre, erre sem lehet pontos választ adni.

Annyi bizonyos, hogy több mint 5 millió, de kevesebb mint 40 millió lapról van szó: a Sztaki és az ELTE számítógép-tudományi tanszékének kutatói egy nemrégiben készített tanulmányukban 10-15 millióra becsülték a .hu domain alá tartozó lapok számát.

"A becslésnél komoly problémát okoznak a duplikátumok" - vélekedik Halácsy Péter, a Budapesti Műszaki és Gazdaságtudományi Egyetem Média Oktató és Kutató Központjának munkatársa. A domainliberalizáció hatásaként ugyanis számos oldal több címen is elérhető, tartalmuk viszont azonos. Persze ezenkívül - például .com végződésű címeken - is vannak magyar nyelvű dokumentumok a neten, közel 300 ezer oldal.

A hálózat szorossága

A keresés nehézségei

A magyar nyelvű keresések során problémát okoz a magyar nyelv két tulajdonsága: a szélesebb karakterkészlet, valamint a ragozás. Az utóbbit jelenleg egyik internetes kereső sem kezeli. Az ékezetek kezelése változó: míg a Google figyelembe veszi az ékezetes betűket, addig a Vizsla nem tesz különbséget például egy "a" és egy "á" betű között. "Ez utóbbi problémákhoz vezethet, ha például a láda szó után kutatunk, így ugyanis könnyen a Lada gépkocsival foglalkozó oldalakra bukkanhatunk" - mutat rá a problémára Benczúr András, a Sztaki kutatója.

A világháló azonban nem különálló oldalakból áll, legfőbb tulajdonsága, hogy valamely site-ról egy összekötő kapcsolat (link) segítségével eljuthatunk egy kapcsolódó lapra. Az oldalak kapcsolatát így pontokkal és az azokat összekötő szakaszokkal (matematikai kifejezéssel: gráfokkal) lehet modellezni. Az ilyen hálózatok vizsgálata jelenleg fénykorát éli. A téma népszerűségéhez nagyban hozzájárult Barabási Albert László Linked (Behálózva) című műve, amely a tavalyi év tudományos sikerkönyve volt (Figyelő, 2003/43. szám).

A hálózat szorosságának fokmérője lehet a linkek számossága is. A magyar web vizsgált részénél a linkek száma az oldalak számának 8,42-szerese volt, tehát egy átlagos .hu lap esetében hozzávetőlegesen 8,5 a lapra mutató, illetve a lapon szereplő linkek száma. A site-ok közötti kapcsolatháló tulajdonságai sokban emlékeztetnek más típusú hálózatokra, például emberek közötti ismeretségekre. Megfigyelhető az úgynevezett kisvilág-jelenség is. A weblapok között szoros a kapcsolat: az idézett tanulmány szerint átlagosan 9 kattintással eljuthatunk az egyik hazai oldalról bármelyik másikra.

A legértékesebbek

A hazai kutatók kísérletet tettek arra is, hogy megkeressék a legértékesebb oldalakat. Ezt a Google kereső által is használt PageRank algoritmus segítségével végezték: ennek alapján az az oldal számít jobb minőségűnek, amelyre több link mutat. Ennek a alapján a tavalyi év végén a legmagasabb pontszámokat a www.freeweb.hu, a www.777sms.hu, illetve a www.westel.hu/talalkozasok site-ok kapták.

A kutatók vizsgálták továbbá a weblapok minősége és a frissítések gyakorisága közti kapcsolatot. Itt arra jutottak, hogy a jó minőségű lapok általában gyakran frissülnek, ám ez fordítva nem igaz: a kevésbé értékes lapok között is szép számmal akadnak olyanok, amelyeket gyakran frissítenek.

Ha érdeklí sajtó internetes megjelenése látogatottságának javítása, keresse meg ezzel foglalkozó partneroldalunkat: www.mahirnet.hu

A teljes cikk a Figyelő 2003/47. számában olvasható.

2004. 02. 26.