

**Debreceni Egyetem
Informatikai Kar**

XML adatfeldolgozás

Készítette:
Megyesi Emil
Programtervező Informatikus (BS.c.) hallgató

Témavezető:
Dr. Juhász István
egyetemi adjunktus

**Debrecen
2007**

Tartalomjegyzék

| | |
|--|----|
| Bevezetés..... | 3 |
| A szabványosítás szükségessége..... | 5 |
| Hivatalos, de-jure szabványok | 5 |
| Ipari, de-facto szabványok | 6 |
| Egyéni, de azért is szabványok | 6 |
| W3C World Wide Web Consortium..... | 9 |
| Az eXtensible Markup Language (XML)..... | 11 |
| Története | 11 |
| Előnyei és hátrányai | 12 |
| XML verziók | 14 |
| Google..... | 17 |
| A goto.com története..... | 21 |
| AdSense, AdWords..... | 22 |
| A keresőmarketingről..... | 28 |
| Apache..... | 32 |
| PHP..... | 34 |
| MySQL..... | 37 |
| KeyAdMan - a hirdetéskezelő..... | 38 |
| A problémakör nehézségei..... | 39 |
| Összefoglalás..... | 41 |
| Irodalomjegyzék..... | 43 |
| Köszönetnyilvánítás..... | 44 |

Ábrajegyzék

| | |
|--|----|
| 1. ábra: Google hirdetések egy weboldalon..... | 21 |
| 2. ábra: Más stílusú Google hirdetések egy másik weboldalon..... | 21 |
| 3. ábra: A bejegyzett oldalak számának alakulása..... | 23 |
| 4. ábra: Az egyes webszerverek használtsága..... | 28 |

Bevezetés

Napjaink egyik legjobban és leggyorsabban fejlődő informatikai ágazata az internet (world wide web). Szinte naponta kerül a figyelem középpontjába valamilyen új fejlesztés vagy egy új szabvány (vagy csak annak jelölt).

Jelen dolgozatomban egy teljesen aktuális webes témán, az online marketingen keresztül szeretném megmutatni a szabványok betartásának fontosságát és nélkülözhetetlenségét, valamint egyes szabványokat, amelyek nélkül ma nem lehetne könnyebben megoldani egyes problémákat.

Ezt egy jelenleg az iparban működő, és napi használatban lévő szoftveren keresztül szeretném megmutatni, amiben az XML és az Unicode előnyeit éles körülmények között tapasztalhattam meg.

Kezdetben áttekintem, hogyan lett az internetes marketing a mai egyik legfontosabb reklámozási felületté. Ezt a legnagyobb internetes marketing szereplő – és nem mellékesen az egyik legfontosabb elindítója – a Google történetének és szolgáltatásainak áttekintésével illusztrálom.

Az internetes szabványok elterjesztésében a Google vezető szerepe kétségtelen, ám hosszú és rögös út vezetett eddig. A szabványok szükségességét is bemutatom a kezdeti lépésektől a mai kaotikus állapotokig. Rámutatok, hogy miért fontosak a szabványok, miért

kellene a szabványokat kötelezővé tenni.

Bemutatom az egyik legnagyobb, az interneten elterjedt szabványt, az XMLt, előnyeivel és hátrányaival egyetemben.

Dolgozatom végén a példaalkalmazás problémakörét tekintem át, majd rámutatok a továbbfejlesztési lehetőségekre. Ez a web dinamikus fejlődése következtében nem csak az asztali szoftvereknél megszokott “mivel lehet még jobba tenni” kérdés, hanem a jövőtől függő változtatások és fejlesztések sorozata.

A szabványosítás szükségessége

A "szabvány" szónak több különböző értelmezése létezik, különösen az informatikában. A szóhoz olyan képzetek kapcsolódnak, mint időtállóság, minőség, tekintély, együttműködés és konszenzus. Általában egy iparági megállapodást értünk alatta, melynek keretében egy termék előállítása vagy egy szolgáltatás üzemeltetése előre specifikált (szabványosított) módon történik.

Ahogy a világ minden területén, úgy az informatikában is szükségessé váltak a szabványok. A legkorábbi időszakától kezdve felismerték az informatikusok, hogy megfelelő szabványok nélkülözhetetlenek az informatikában.

Háromféle szabványt különböztetünk meg az informatikában:

Hivatalos, *de-jure* szabványok

Ide tartoznak azok a szabványok, melyeket a különböző államok által törvényi szinten elismert, szabványok megalkotására létrejött szervezetek adnak ki. A szabványügyi testületek három szinten helyezkednek el, és az általuk kiadott szabványok is általában ezekre szintekre érvényesek:

- globális, nemzetközi szint (pld. ISO, IEC, ITU-T)
- regionális szint (pld. az Európai Unióban: CEN, CENELEC, ETSI)
- nemzeti szint (pld. Magyar Szabványügyi Testület)

Ipari, *de-facto* szabványok

A legtöbb ilyen szabvány egy adott iparág, konzorciumba tömörült érdekelt feleinek együttműködési törekvése kapcsán jön létre. Habár ezek nem de-jure szabványok, hitelességüket mégis a nyitott és átlátható együttműködési törekvések biztosítják. Példa erre a W3C (World Wide Web Consortium), mely több mint 350 informatikában érdekelt cég és non-profit szervezet együttműködése. A W3C is ajánlásokat (recommendations) bocsát ki szabványok helyett, hiszen neki sincs jogi engedélye de-jure szabványok kiadására.

Egyéni, *de azért is* szabványok

Néhány specifikáció, habár egyik szabványügyi szervezet sem adta ki vagy hagyta jóvá, annyira elterjedt, hogy lényegében szabvánnyá vált. Ilyen például a Microsoft® Windows® operációs rendszer. Ez az informatikában kialakult “minél előbb legyünk a piacon” felfogásnak lehet egyrészt köszönni. Mindegyik nagy gyártó/szolgáltató minél előbb piacon akar lenni a termékével, hogy a felhasználók minél előbb szokjanak rá a saját termékükre. Hiányzó informatikai érdekvédelmi szervezet híján ezt megtehetik és meg is teszik.

Sajnos az informatikában a szabványosítás területén hatalmas ellentmondások és hibák találhatóak - lásd az utolsó szabványcsoportot. A cégek igyekeznek minél több helyen a saját megoldásaikat használtatni, így - szabványosítás előtt - olyan előnyre és pozícióra szert tenni, hogy később vagy szabvány lesz a megoldásuk, vagy elfogadják és mindenki használja - hivatalos

szabványosítás nélkül.

Ezt hívják **vendor lock-in**nek.

A *vendor lock-in* angol kifejezést arra a helyzetre alkalmazzuk a számítástechnikában és a köré csoportosuló iparban, mikor egy ügyfél egy gyártó (vendor) bizonyos termékét használva csapdába, kényszerhelyzetbe kerül.

Példa: egy ügyfél elkezdi használni X gyártó irodai programcsomagját. Ez a programcsomag olyan fileformátumokban tárolja az információt, amely semelyik másik konkurens gyártó hasonló termékével sem kompatibilis. Emellett a formátum nem nyílt, így a konkurensnek esélye sincs kompatibilis programcsomagot készíteni. Az ügyfél hosszú évekig használja a programcsomagot, és ennek következtében nagymennyiségű fontos dokumentuma tárolódik már ezekben a formátumokban. Közben úgy döntene, hogy Y gyártó irodai programcsomagjára váltana, mert az jobban megfelelne igényeinek. Sajnos nem tud váltani, mert az évek óta felhalmozódott dokumentumokkal nem tud mit kezdeni. Az ügyfél kénytelen kényszerűségből X gyártó termékvonálánál maradni és ha naprakész szeretne maradni, akkor kénytelen a verzióváltásokkal együtt haladni.

W3C World Wide Web Consortium



"A webtársadalomhoz tartoznak mindazok, akik böngésznek vagy információkat helyeznek el a Weben, akiknek a Web jelenti a munkahelyüket, akik új termékeket és szolgáltatásokat fejlesztenek a Weben, ezenkívül az internetszolgáltatók, valamint azok, akik elképzelésekkel rendelkeznek arról, hogyan fejlesszék tovább a Webet... A Konzorcium az a semleges közeg, melyben ezek a csoportok megegyezésre juthatnak a szabványokat illetően. Ezeknek az erőknél a találkozási révén gazdagodik a W3C."

Tim Berners-Lee, a W3C igazgatója

A World Wide Web Consortiumot (W3C) 1994-ben alapította Tim Berners-Lee, a Web szülőatyja. A cél elsősorban az volt, hogy a webtechnológiák fejlesztésével foglalkozó vállalatok és kutatóintézetek a jövőben ne forgácsolják szét feleslegesen tudásukat, elért eredményeiket, hanem egymással vállvetve, egymást segítve tereljék ugyanazon irányba az informatika újításait. Mára már kiderült, hogy a Web nyújtotta lehetőségek tárháza rendkívüli tempóban bővül napról napra a kereskedelem, a kommunikáció, az információs fórumok és a kollektív tudás terén egyaránt. A W3C éppen ezeknek a folyamatosan bővülő lehetőségeknek a kiaknázására szerveződött.

Az MTA SZTAKI 1995 óta tagja a Konzorciumnak. A W3C itt nyitotta meg első kelet-közép-európai irodáját, a W3C Magyar Irodát.

A W3C olyan egymással együttműködő technológiák (specifikációk, irányelvek, szoftverek) fejlesztésével foglalkozik, melyek segítségével kiaknázzhatjuk a Web összes lehetőségét az információs fórumok, a kereskedelem, a kommunikáció és a kollektív tudás terén egyaránt.

Egy W3C-ajánlás (*W3C Recommendation*) azt jelzi, hogy egy specifikáció stabil, elősegíti a Web interoperabilitását, a W3C tagjai átnézték, és helyeslik annak ipari alkalmazását.

A W3C karolta fel az XML specifikációt is.

Az eXtensible Markup Language (XML)



Az **Extensible Markup Language** (XML, kiterjeszhető leíró nyelv) a W3C által ajánlott általános célú leíró nyelv speciális célú leíró nyelvek létrehozására. Az SGML egyszerűsített részhalmaza, mely különböző adattípusok leírására képes. Az elsődleges célja strukturált szöveg és információ megosztása az Interneten keresztül. Az XML-en alapuló nyelvek (például RDF, RSS, MathML, XSIL, SVG) formális módon vannak leírva, így lehetővé téve a programok számára a dokumentumok módosítását és validálását a formátum előzetes ismerete nélkül.

Története

Az XML-t Tim Bray fejlesztette ki, miután az IBM, az Oxford University Press és a University of Waterloo egy által támogatott internetes szótáron dolgozott. Mivel hatalmas mennyiségű adatot kellett tárolni és feldolgozni, kereskedelmi szoftvermérnököket vontak be a projektbe, hogy megoldást találjanak az adatok indexelésére és tárolására. Az Association for Computing Machinery (ACM) számára adott interjú során Bray azt nyilatkozta, hogy bevonták a projektbe és megmutatták neki a szótár számára kifejlesztett belső struktúrát: "kis beágyazott címke határozta meg, hogy mi bejegyzés illetve szó, és aztán kiejtés, etimológia, rövid idézet, és aztán adat, forrás, szöveg és így tovább" (ACM Queue, 2005). Ez vált az XML elődévé. Miután kifejlesztették a technológiát a szótár projekthez, Bray megalapította az Open Text Corporation-t,

kifejlesztett egy kereső motort, valamint meghívták a W3C-be, hogy legyen az XML specifikációjuk szerkesztője.

Előnyei és hátrányai

Az XML azon tulajdonságai, melyek alkalmassá teszik adattovábbításra:

- mind ember, mind gép számára olvasható formátum
- támogatja a Unicode-ot, ami lehetővé teszi bármely információ bármely emberi nyelven történő közlését
- képes a legtöbb általános számítástudományi adatstruktúra ábrázolására (rekord, lista, fa...)
- öndokumentáló formátum, amely struktúra- és mezőneveket ír le speciális értékekkel együtt
- szigorú szintaktikus és elemzési követelményeket támaszt, ami biztosítja, hogy a szükséges elemzési algoritmus egyszerű, hatékony és ellentmondásmentes maradjon

Az XML-t gyakran használják dokumentumtárolási és feldolgozási formátumként, mind online mind offline módon, és több előnnyel is jár:

- internetes szabványokon alapuló erőteljes, logikailag ellenőrizhető formátum
- a hierarchikus struktúrája megfelel a legtöbb (de nem mindegyik) dokumentum típusnak

- egyszerű szöveg formátumban valósul meg, licencektől és korlátozásoktól mentesen
- platform-független, így viszonylag immúnis a technológiai változásokkal szemben
- az XML-t és elődjét, az SGML-t már több mint tíz éve használják, így széles tapasztalat és eszközkészlet áll rendelkezésre

Bizonyos alkalmazások szempontjából a következő hátrányokkal rendelkezik:

- A szintaxisa elég bőbeszédű és részben redundáns. Ez nehezítheti az emberi olvashatóságot és az alkalmazások hatékonyságát, valamint nagyobb tárolási költséggel jár. Nehézzé teszi az XML alkalmazását korlátozott sávszélesség esetén, bár bizonyos esetekben a tömörítés csökkentheti a problémát. Ez részben igaz a telefonokon és PDA-kon futó multimédiás alkalmazásokra, melyek XML-t szeretnének használni képek és videók leírására.
- A szintaxis számos homályos, felesleges tulajdonsággal bír, ami az SGML hagyatéka.
- Az alapvető elemzési követelmények nem támogatják az adattípusok túl széles körét, így néha a kívánt adat kinyerése a dokumentumból plusz munkával jár az elemző részéről. Például nincs lehetőség XML-ben a "3,14159" lebegőpontos számként való megjelölésére hét karakterből álló sztring helyett.
- Nincs lehetőség a dokumentum egyes részeinek közvetlen

elérésére és frissítésére.

- Egymást részben átfedő (nem hierarchikus) adatstruktúrák modellezése külön erőfeszítést igényel.
- Az XML relációs és objektum orientált paradigmához kötése néha fáradságos.

XML verziók

Jelenleg két XML verzió létezik. Az *XML 1.0*-t 1998-ban definiálták. Többször átnézték már azóta anélkül, hogy új verziószámot kapott volna. Jelenleg a 2004. február 4-én publikált harmadik kiadás aktuális. Széles körben elterjedt, és még mindig ajánlják általános felhasználásra. Az *XML 1.1*-et egy napon adták ki az XML 1.0 harmadik kiadásával. Sok olyan – néhány vitás – tulajdonsággal bír, aminek célja az XML használatának egyszerűbbé tétele több felhasználói csoport (főleg mainframe programozók) számára. Az XML 1.1 nem túl elterjedt, és csak azoknak ajánlják, akik ki tudják használni az egyedi képességeit.

Az XML 1.0 és az XML 1.1 az elemek, attribútumok stb. nevében megengedett karakterekben különbözik. Az XML 1.0 csak a Unicode 2.0 szabványban megengedett karaktereket engedélyezi, ami magában foglalja a legtöbb világnyelvet, de nem tartalmazza azokat a nyelveket, melyeket a későbbi Unicode szabvány verzióban vezettek be. Ilyen nyelv a mongol, a kambodzsai, a burmai stb. Az XML 1.1 csak bizonyos vezérlő karakterek használatát *tiltja*, ami azt jelenti,

hogy minden más karakter használható, még ha nem is definiálták az aktuális Unicode verzióban.

Meg kell jegyezni, hogy az XML 1.0 megkötései csak az elemek illetve attribútumok neveire vonatkoznak: mind az XML 1.0, mind az XML 1.1 megengedi a teljes Unicode készlet használatát a tartalomban. Így az XML 1.1 csak akkor szükséges, ha a Unicode 2.0 után publikált szöveghez szeretnénk elem és attribútum neveket fűzni.

Másik kisebb különbség az XML 1.0 és az XML 1.1 között, hogy a vezérlő karakterek nem szerepelhetnek, hacsak nem "escape" karakter formátumúak, és két speciális Unicode karaktert tartalmaz, amit üres helyként kezelnek.

Az XML 1.0 dokumentumok helyesen formázott XML 1.1 dokumentumok egy kivétellel: a nem "escape" formátumú C1 karakterek hibásak: az XML 1.1 megköveteli a C1 vezérlő karakterek számmal történő hivatkozását.

Megbeszélés folyik az XML 2.0 szabványról, bár nem látható előre, hogy egyáltalán létre fog-e jönni. Az XML-SW (SW: skunk works, titkos projekt) – amit az XML egyik eredeti szerzője írt – tartalmaz néhány utalást az XML 2.0 formátumára: a DTD kiküszöbölése a szintaxisból, névterek, XML Base és XML Information Set (infoset) integrációja az alap szabványba.

A W3C rendelkezik egy XML Binary Characterization Working Group-pal, ami bevezető kutatást végez a binárisan kódolt XML infoset-ek felhasználásának terén. A munkacsoportnak nem célja bármilyen szabvány létrehozása. Mivel az XML definíció szerint szöveg alapú, a Sun Microsystems a *gyors infoset* nevet adta a saját bináris infoset

implementációjának, hogy elkerülje az összetévesztés lehetőségét, és az ISO-n keresztüli szabványosításán dolgozik.

Az XML rövid idő alatt az internet kedvelt nyelvává vált, a fentebb említett hasznos tulajdonságai alapján. Az eredeti elképzelése, a platformfüggetlenség, egyik legjobban kihasználható tulajdonsága az internetes környezetben, így a világháló tényleg a teljes szabadságot kínálja – feltéve, hogy a honlapok készítői figyelnek a szabványokra és betartják azokat.

Látni fogjuk, hogy egy olyan országokon-kultúrákon átívelő területen, mint az online marketing, az XML és vele az Unicode a leghasznosabb segítőtárs, a legkönnyebb és legegyszerűbb megoldásokhoz nélkülözhetetlenek.

Most pedig lássuk az internetes hirdetések bölcsőjét, hogyan és honnan indult ez a napjainkat egyik legjobban befolyásoló hirdetési forma.

Google



A nagy portálokkal egyidőben pedig sokan elkövették azt a hibát, hogy leírták a keresőgépeket - sokan úgy hitték, hogy igazából nincs szükség keresőmotorokra... A “Google-éra” gyermekeinek, meglehetősen érthetetlennek tűnik ez a gondolat. Larry Page és Sergey Brin, a Google két alapítója sem így gondolták - és megszületett napjaink leginkább használt keresőmotorja.

Talán nehéz elhinni, de amikor már megtette első sikeres lépéseit a Google, mint üzleti vállalkozás - nos, még akkor sem igazán volt túl sok elképzelésük az alapítóknak arról, hogy miképp is termeljenek pénzt a keresőmotor felhasználóiból. A kezdeti időkben a Google abszolút nem akart semmiféle reklámot megjeleníteni - a különféle bannereket hihetetlenül bosszantónak tartották. Igen, a felhasználókban már akkor is igen gyorsan kialakult a reklám-undor, nem akartak reklámokat nézegetni - Page és Brin meg nem akart reklámokat megjeleníteni a Google főoldalán. Mégis, az AdSense¹ révén ők váltak az online reklámpiac egyik legmeghatározóbb figuráivá.

Sokan úgy hiszik, hogy ezt a szöveg-alapú reklámot, ami az AdSense legfőbb jellemzője, Larry Page és Sergey Brin találták ki, holott ez így ebben a formában nem igaz. Egy Bill Gross nevű úriember 1997-ben alapította meg a GoTo.com-ot², ami már azokkal az

¹ lásd a következő fejezetet

² lásd a fejezet végén

alaptulajdonságokkal kísérletezett, amik később az AdSense-t tették sikeressé.

Havi fél millió dollár kiadás - 20 millió dollár a bankszámlán - bevétel meg gyakorlatilag nuku. Míg egyre többen és többen kattantak rá a Google használatára, a céget üzletileg nagyjából ez jellemezte '99 végén.

Ezen a nem túl rózsás felálláson változtatott a Google AdWords bevezetése. Lássuk hát a történet második felvonását.

Larry Page és Sergey Brin nem akarták, hogy a Google úgy nézzen ki, mint a legtöbb portál: nem akarták mindenféle csiricsaré, villogó bannerekkel telepakolni az oldalt, holott valószínűleg több millió dollárt kaszáltak volna, ha erre az útra lépnek.

Nem ezt tették, legalábbis nagyon-nagyon ódzkodtak tőle. Amikor elkezdett formálódni az üzleti elképzelésük, akkor azért úgy gondolták: ha minden kötél szakad, akkor mégiscsak eladnak bannerhelyet a Google.com nyitólapján, de ha lehet, megpróbálkoznak valami mással.

Két dologhoz ragaszkodtak minden áron: egyrészt, a reklámok csak és kizárólag szövegesek lehettek - csiricsaré képek kizárva. Másrészt, a reklámok akkor bukkantak fel, ha valaki olyan kulcsszóra keresett, amihez passzolt a reklám - ha autóra kerestél, akkor jelent meg pl. a Ford hirdetése.

A reklámozó itt a megjelenések számának arányában fizetett. Továbbá a hirdetéseket reklámszervező hozta, nem lehetett “csak úgy” kampányt indítani a rendszerben.

Ez a rendszer jó pénzt hozott ugyan a konyhára, de nem eleget - nagyon úgy nézett ki, hogy a Google mégiscsak meg fog állapodni a DoubleClick nevű reklámszolgáltatóval, aki a hagyományos bannerértékesítésben utazott. Csakhogy egy szép napon a DoubleClick korábban 150 dollárt érő részvényei lezuhantak 15 dollárra.³

A Google előtt nem állt más lehetőség, mintsem hogy előremeneküljön - fogták tehát, és gyakorlatilag lemásolták a GoTo.com módszerét (kis ablakban a goto módszere), egy fontos különbséggel: a Google továbbra is élesen elkülönítette a keresőprogram természetes (organikus) találatait a fizetett hirdetésektől. És kezdetben továbbra is a megjelenések száma alapján kellett a hirdetőknak fizetniük.

És volt még egy jelentős különbség - az átlagember sokkalta könnyebben vásárolhatott magának reklámot, és talán ez az a fő szempont, ami miatt sikeressé válhatott az AdWords szolgáltatás.

³ A sors különös fintora, hogy a Google pár héttel ezelőtt végül mégis felvásárolta a DoubleClicket – a történelem eddigi legnagyobb készpénzes kifizetésével.

A goto.com története

Egy Bill Gross nevű úriember 1997-ben alapította meg a GoTo.com-ot, ami már azokkal az alaptulajdonságokkal kísérletezett, amik később az AdSense-t tették sikeressé.

Gross egyszerre három nagy újítást vezetett be: egyrészt, a hirdetők nem a reklám megjelenéséért fizettek, hanem azért, ha egy felhasználó a reklámra rákattintott. Másrészt, a reklámok nem össze-vissza jelentek meg, hanem akkor, ha egy felhasználó bizonyos kulcsszavak után kutakodott: ha autókereskedő vagy, akkor érdemes az autó kulcsszót megvenned - és akkor a reklámod akkor jelenik meg, ha valaki autók után kutakodik. Harmadrészt, hihetetlen alacsony árakkal indult a GoTo.com: kezdetben kattintásonként mindössze 1 centet kellett fizetnie a hirdetőnek.

És volt még egy nagy újítása a GoTo.comnak: a hirdetőknél licitálniuk kellett a kulcsszavakra. Azaz, ha több autókereskedő akarta megvenni az "autó" kulcsszót, akkor a sorrendet az döntötte el, hogy ki volt hajlandó többet fizetni egy kattintásért.

AdSense, AdWords



Tekintsünk egy picikét vissza, az AdSense előtti időkbe, hogy jobban megérthessük, mi vezetett ahhoz, hogy a világ legsikeresebb marketing-eszközévé válhatott ez a reklám-típus.

A történet a nagy portálokkal, a keresőgépek megjelenésével, meg az első online reklámokkal kezdődik: az interneten megjelentek az olyan óriási méretű portálok, mint a Yahoo (itthon pár évvel később az Index-előd Internettó, illetve az Origo), amelyek üzleti modellje alapvetően egyszerű volt: azt az irgalmatlan mennyiségű forgalmat “adták el”, amit sikerült elérniük: megjelentek a reklámbannerek minden oldalon, a reklámozók pedig a reklámmegjelenés száma alapján fizettek.

Az AdWords lényege a mai napig az, hogy bárki, bármilyen vállalkozást is vezessen, könnyen és egyszerűen tudja magát regisztrálni a rendszerben, és olcsón jusson minőségi találatokhoz. 2003-ban még annyi lehetősége sem volt egy kicsi vagy közepes honlap üzemeltetőjének arra, hogy reklámhelyeket adjon el, mint manapság: ma azért már van rá mód, hogy pl. az itthon Adnetwork néven futó szolgáltatáshoz csatlakozzon bárki, és a látogató / letöltésszámát bedobja a “nagy közösbe”: az Adnetwork jellegű cégek bizonyos tematikák mentén összegyűjtene egy csomó, minimális szintet megütő honlapot, és a reklámozók, jellemzően marketingügynökségek felé egy nagy egészként adják el a teljes hálózatot.

Az AdWords lényege, hogy a felhasználót abban a pillanatban találják meg a reklámmal, amikor a legfogékonyabb annak üzenetére: a Google keresőmotorjának segítségével sebészi pontossággal megtalálható ez a pillanat: ha én mondjuk videókamerák után kutakodok a Google segítségével, akkor abszolút nyitott vagyok egy olyan reklámra, amely valamilyen videózással kapcsolatos terméket, szolgáltatást, üzletet ajánl.

Nos, az AdSense ezt a nyitottságot terjeszti ki a blogokra, és egyéb csatlakozó honlapokra is, mint amilyen az SFblogs.net rendszere.

Az AdSense nem sokkal a Blogger megvásárlása után, 2003 márciusában indult el. Az AdSense azt tette lehetővé a világon akkoriban egyedülálló módon, hogy bárki, akinek honlapja van, akármilyen kicsi a látogatottsága és a letöltésszáma, szelhessen magának egy szeletet a globális reklámtortából.

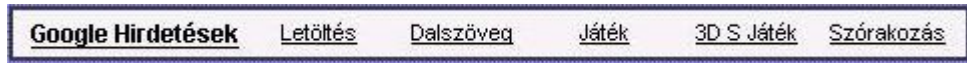
Az AdSense azonban továbbra is a világ legnagyobb tartalmi és hirdetői hálózata: épp olyan egyszerű módot kínál azóta is a csatlakozásra, mint az AdWords a hirdetők részére: regisztrálsz a honlapodat, blogodat a rendszerben, kapsz egy javascript kódot, amit be kell illesztened - és voilá, már meg is jelennek az első reklámok a weblapodon.

Az AdSense rendszer technikailag más, filozófia szempontjából azonban nagyon is hasonló elven működik, mint eredetileg működött az AdWords: ha én a fenti példával élve, videó után keresek a Google.hu-n, akkor olyan reklámokat jelenít meg, amik kapcsolódnak a videózáshoz. Az AdSense ugyanezt csinálja: az AdSense robot elemzi a honlapodat, megpróbálja "kitalálni", miről is szól az, és ezen tartalomnak megfelelő reklámokat helyez el rajta.



1. ábra: Google hirdetések egy weboldalon

Az AdSense egyik nagy előnye, hogy majdnem tetszőlegesen alakítható, többféle méretben és részletezettségben lehetséges a reklámokat a saját weboldalra elhelyezni, és a mai napig mindegyik szöveges. (igaz, már vannak kísérletek már a Google-nél is képes sőt újabban mozgóképes hirdetésekre is) formátumú.



2. ábra: Más stílusú Google hirdetések egy másik weboldalon

Az AdSense nagyon gyorsan nagyon nagy sikereket könyvelhetett el magának: ezrével csatlakoztak a honlaptulajdonosok a rendszerhez - főleg azok, akikkel korábban a hirdető, reklámügynökségek nem törődtek: túl picik voltak ahhoz, hogy bármi eredményt el tudjanak érni hirdetésszervezés tekintetében... pláne, hogy azok a honlapok, amik a web túlnyomó részét adják, nem tudják, hogyan kell hirdetéseket szervezni, és pénzük sincs ezzel foglalkozó szakember felvételére.

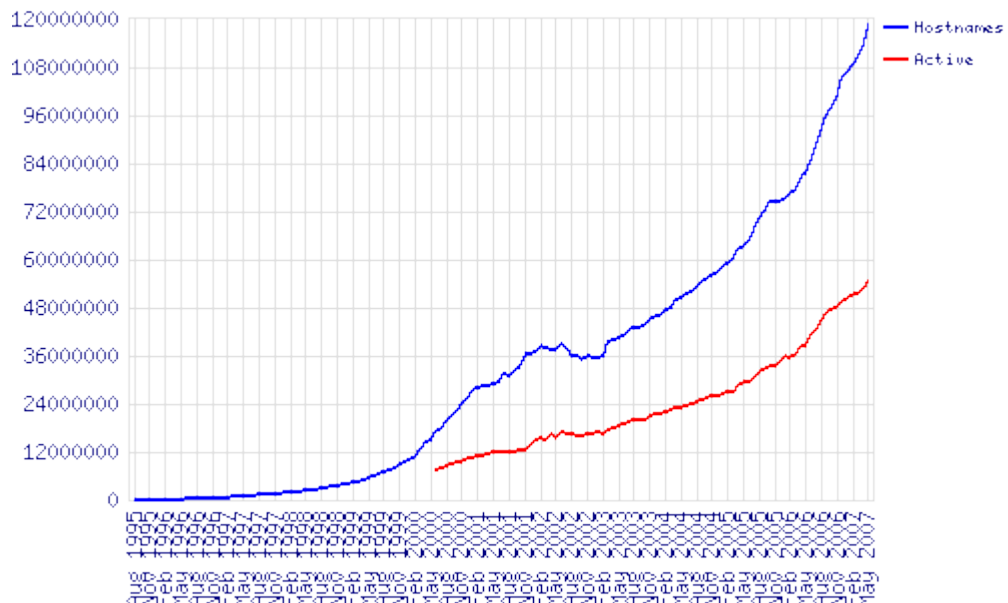
Értelemszerűen nagy eséllyel olyan reklámok jelenítődnek meg az adott oldalon, ami az oldallal kapcsolatos – például az autózás történetéről szóló oldalakon autókereskedők hirdetései fognak megjelenni – hiszen így sokkal nagyobb az átkattintás⁴ esélye, mint ha például drogériák reklámait látná a látogató.

A Google hajlandó volt kiterjeszteni a rendszerét ezekre a weboldalakra is - felismerték, hogy az AdWords esetében fényesen bevált “sok kicsi sokra megy” elve kiterjeszthető: “még több kicsi még többre megy”. A Google a mára már több milliárd dolláros vagyonát gyakorlatilag centenként kereste meg - 2005. elejére minden 15. cent az AdSense rendszeréből származott.

A Blogger megvásárlása így utólag teljesen egyértelmű: a masszív felhasználói réteg, amit ezzel megszerzett, egy világméretű hirdetési hálózat alapkövét jelentette számára.

A Google korán felismerte azt, hogy minél több ember használja a szolgáltatásait, annál több helyen látják a reklámait, annál többen kattintanak rá, ami számára pénzt jelent, ráadásul nem kevés pénzt. Másrészt azt is felismerte, hogy minél könnyebben lehet használni a szolgáltatásait, annál többen és gyakrabban fogják használni.

4 a hirdetésre való rákattintás



3. ábra: A bejegyzett oldalak számának alakulása

A grafikon a bejegyzett oldalak számának növekedését mutatja. A bejegyzett oldalakkal arányosan emelkedik az internetet aktívan használók száma, és így a potenciális célközönség is növekedik.

Minél több ember használja az internetet, annál több ember válik – főleg az interneten keresztül történő vásárlási lehetőségek megjelenésével – a keresőmarketing célpontjává. A következő fejezetben azt tekintem át, hogy mi ez a fiatal, ám láthatóan dinamikus növekedő ágazat.

A keresőmarketingről

Láthattuk eddig a Google cég rövid történetét, megismerhettük, hogyan hirdethetünk a keresőóriás oldalain, illetve hogyan helyezhetünk el célzott hirdetéseket az internetes oldalak millióin. Néhány szót ejtenék arról, hogy ez miért fontos.

Az internet felhasználói - amennyiben egy, számukra eddig ismeretlen weboldalt szeretnének megtalálni - leggyakrabban a keresőket használják. A keresés során a látogató egy adott kifejezésre keres rá, és várja a site-okat, amelyeken további információkat talál. A potenciális vevők nagy része is a keresőkön keresztül jut el a honlapokra. Lényeges tehát, hogy egy adott keresés eredményeként megjelenő találatok között milyen helyet foglal el az adott cég honlapja. Iparágtól, illetve a begépett kulcsszavaktól függően akár több ezer, sőt több millió találat is lehetséges a keresőprogramokban.

A keresőbeli viselkedésre vonatkozó statisztikák adatai szerint az érdeklődők nagy része leggyakrabban csak az első találati oldalt (azaz az első tíz találatot) nézi meg. A második-harmadik oldalon szereplő találatok nagyságrenddel kisebb arányban generálnak látogatást, a további találatok átkattintási aránya pedig elhanyagolhatóan kis mértékű. Ebből is kitűnik tehát, hogy egy keresés során a találati listában az első oldalnak van a legnagyobb szerepe.

A keresőmarketing feladata, hogy az internetes keresőket használó emberek közül minél többet irányítson megrendelő weboldalára.

Ennek egyik fejtája a keresőkben való hirdetés.

A **keresőkben való hirdetés** során a találatok előtt vagy mellett elhelyezett hirdetési felületeken jelennek meg a hirdetések, amelyek a cégek tevékenységére utaló meghatározások (terméknevek, szolgáltatások) köréből kiválasztott kulcsszavakra jelenik meg, a hirdető cég webhelyére vezető hivatkozással. Ezekért a rövid szöveges hirdetésekért általában átkattintásonként kell fizetni, vagyis akkor, ha ténylegesen rákattintanak a hirdetésre. A hatékonysága abban mutatkozik meg, hogy a keresőkből érkező látogatók már befogadóképes állapotban érkeznek a cég oldalára, mivel ők maguk keresnek az adott témára, emiatt könnyebben megszólíthatók, jobb a konverziós arány, mint az általános látogatóké.

Ez a típusú hirdetés jellégéből adódóan rengeteg adatot termel, amit később ki kell elemezni, hogy a kampányt pontosítani, finomítani lehessen. Erre a rengeteg adatra egyben a megrendelőnek is szüksége van - melyik kulcsszó mennyire hatásos, illetve mennyibe kerül a kampány. Természetesen minden megrendelőnek, más szempont alapján van szüksége az adatokra, ezért lényeges egy saját rendszer ezen adatok tárolására és a jelentések készítésére.

A KeyAdMan, a hirdetéskezelő rendszer, ezt a feladatot látja el, a keresőktől (egyelőre csak a Google-től, de úgy lett megtervezve a rendszer, hogy a későbbiek folyamán más hirdetési adatokat is tudjon kezelni) érkező adatokat dolgozza fel, a Report Maker (jelentés készítő) része pedig a megrendelők igényeinek (a cég vezetőinek) megfelelő formában szolgáltatja a kampányaikról az adatokat. Így elősegítve a döntéshozókat a későbbi lépések megtételéhez.

A továbbiakban bemutatom azt a környezetet, amiben dologzva elkészítettem ezt a jelenleg az iparban működő, online marketinggel foglalkozó cég által használt terméket.

Apache



The **Apache Software Foundation**

<http://www.apache.org/>

Az Apache projekt célja egy olyan Web-szerver program létrehozása, karbantartása, és fejlesztése, amely megfelel a gyorsan változó Internet követelményeinek, biztonságos, üzleti, vállalati felhasználásra is megfelelő és szabadon használható. Az Apache a régi NCSA httpd szerverre épül, az Apache Szoftver Licenc feltételei alatt terjesztik.

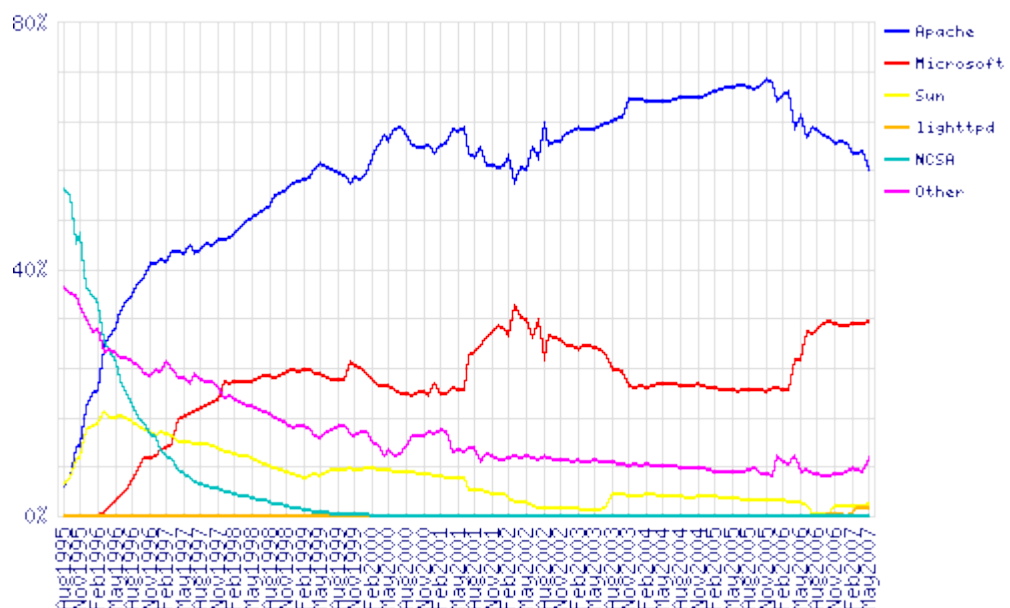
Az Apache robosztus, erőteljes, flexibilis webszerver, amely HTTP/1.1 (RFC2616) protokoll kompatibilis. Az Apache projekt koordinálását az Apache Software Foundation végzi. Néhány vezető és több száz fejlesztő van e mögött a projekt mögött.

Az Apache webszervernek semmi köze az Apache indiánokhoz, akik az Egyesült Államok dél-nyugati részén élő őslakók. A név egy szójáték: "a patchy server", ami arra utal, hogy a szervert eleinte a webmesterek által készített patchekkel frissítették. Az Apache az NCSA Web-szerver toldozgatásából fejlődött ki. Az NCSA egyike volt az első Web-szervereknek, amelyet az Illinois-i egyetemen található NCSA fejlesztett ki.

Az Apache projektet webmesterek, rendszergazdák kezdték el, amikor Rob McCool, az NCSA (National Center for Supercomputing Applications) webszerverének írója 1994. közepén kilépett az NCSA-tól, és a szoftver nem fejlődött tovább. A webmesterek levelezéssel tartották egymással a kapcsolatot, azon keresztül adták át egymásnak a patcheket, amelyek így körbe jártak. Ezt látva Brian Behlendorf és Cliff Skolnick a Hotwired segítségével elindított egy levlistát, ahova

a core fejlesztők feliratkozhattak. Így indult el az Apache projekt. 1995. februárjában 8 core fejlesztővel megindult a koordinált fejlesztés.

Jelenleg a legnépszerűbb webservert, a serverek 60%-án fut Apache a Netcraft adatai alapján.



4. ábra: Az egyes webserverek használtsága

Jelenleg két verzió (1-es és 2-es) fejlesztése fut párhuzamosan egymás mellett, 1.3.37 és 2.2.4-es verziószámoknál tartva.

PHP



A dolgozathoz használt programozási nyelv legfőbb előnyei: A PHP, bővebben "PHP: Hypertext Preprocessor" egy széles körben használt, nyílt forráskódú, általános célú programozási nyelv, különösen jó web-fejlesztés támogatással, és HTML-be ágyazási képességekkel. A szintakszisa a C, Java és Perl nyelvekre épül, könnyen megtanulható. A nyelv fő célja lehetőséget teremteni dinamikusan generált weboldalak gyors készítésére, ám a PHP ennél sokkal többre is képes.

A PHP használható a legfontosabb operációs rendszereken, beleértve a Linuxot, sok Unix változatot (beleértve a HP-UX, Solaris és OpenBSD rendszereket), a Microsoft Windows-t, a Mac OS X rendszert, a RISC OS-t, és másokat.

A PHP a legtöbb webszervert is támogatja, beleértve az Apachet, Microsoft Internet Information Servert, és másokat.

A PHP rendelkezik egy DBX adatbázis absztrakciós kiterjesztéssel is, amellyel egyöntetűen és áttetsző módon lehet kezelni bármilyen adatbázist, amit ez a kiterjesztés támogat. Ezen kívül a PHP támogatja az ODBC-t, ezért bármilyen más, ezt a szabványt támogató adatbázishoz is lehet kapcsolódni.

Az XML dokumentumok feldolgozásához és eléréséhez PHP4-ben a SAX és DOM szabványok is használhatóak, az XSLT kiterjesztés XML dokumentumok átalakítására használható. Mivel a nyelv jelenleg az 5-ös verziónál tart, ezért a teljességhez meg kell említeni, hogy a PHP 5 az XML-el kapcsolatos feladatokat egységesen, a libxml2 függvénykönyvtárra támaszkodva látja el. Ezen biztos alapokon a PHP 5 bevezeti a SimpleXML és az XMLReader támogatást is.

Jelenleg az 5.2.1-es verziónál tart a fejlesztése, de párhuzamosan fejlesztik a 4-es verziót is, ez a 4.4.6-os verziónál jár.



MySQL

A MySQL egy gyors, több szálú, több felhasználós robusztus SQL adatbázis-szerver. A MySQL Unix, OS/2 platformok alatt általános esetben ingyenesen használható, míg Microsoft platformok alatt a 30 napos próba használat után meg kell vásárolni (a licenst) a MySQL honlapján, a <http://www.mysql.com/> helyen. Jelenleg az 5.1.17-es verzióval tart a MySQL, de párhuzamosan fejlesztik tovább a 4-es verziót, ez jelenleg a 4.1.21-nél tart.

A fejlesztések kb. 1996-ban kezdődtek el, amikor a szerzőknek TeX-nél egy nagy adatbázisokat biztonságosan és gyorsan kezelő SQL szerverre volt szükségük, ugyanis ekkor más adatbázis-rendszer fejlesztő cég nem rendelkezett ilyen környezethez igazán hatékony SQL szerverrel - a szerzők szerint.

KeyAdMan - a hirdetéskezelő

Most pedig röviden tekintsünk át egy példa alkalmazást, ami az eddig ismertetett szabványokat és eljárásokat használja fel és az iparban jelenleg is működik.

A KeyAdMan program egy online marketinggel foglalkozó cég, amely Google-ben (lábléc: és később más hirdetési rendszerekben) gondozott kampányai adatait dolgozza fel és tárolja le, hogy a cég döntéshozóinak és ügyfeleinek könnyen és gyorsan lehessen kimutatásokat előállítani.

A Google AdWordsbeli kampányok hirdetési adatai XML formátumban érkeznek minden nap hajnali 5 órakor. Az XML állomány mindig az előző nap adatait tartalmazza, kulcsszó szinten megmutatva az előző napi megjelenési adatokat és átkattintási rátákat. A KeyAdMan a Google-től érkező adatok feldolgozását végzi. Működése során a megérkező adatokat könnyen felhasználható formába konvertálja át és letárolja a helyi adatbázisban.

A program másik fele ebből az adatbázisból a kiválasztott formában készít jelentést, melynek ismertetése nem témája a szakdolgozatomnak, ezért csak említés szinten foglalkozom vele.

Lássunk egy példát kitalált adatokkal illusztrálva a rendszert:

```
<row date="2007-03-21" acctname="autókereskedő"
timezone="(GMT+01:00) Budapest" campaign="használt autó"
adgroup="márkák" keyword="Ford" kwType="Broad"
kwStatus="Active" keywordMinCpc="100000" adwordsType="Search
Only" budget="1000000000" maxCpc="800000" agStatus="Enabled"
campStatus="Active" currCode="USD" imps="334" clicks="4"
ctr="0.011976047904191617" cpc="365000" cost="1460000"
pos="1.3982035928"></row>
```

1. kód: A Google-től érkező XML egy sora (kitalált adatokkal)

```
<row date="2007-03-21" acctname="autókereskedő"
campaign="használt autó" adgroup="márkák" keyword="Ford"
kwType="S" budget="1000" currCode="USD" imps="334"
clicks="4" ctr="1.20" cpc="0.37" cost="1.46"
pos="1.4"></row>
```

2. kód: A feldolgozás után, az adatbázisban való tárolás előtt

A problémakör nehézségei

- A Googletől kapott XML szerkezete nem egységes, többször előfordult, hogy változott az adatok formája (pl. a dátum).
- Ha változott egy kulcsszó átkattintási értéke, akkor ugyanazon a nap több sorban érkeznek az adott kulcsszó adatai.

- Az egyes értékek valamikor lebegőpontos valamikor pedig tizedes tört alakban érkeznek. Ez egy XMLen belül is változó, és naponta más kulcsszavaknál fordul elő.
- Különösen körültekintően kellett bánni a rendelkezésre álló erőforrásokkal, mert a feldolgozandó adatmennyiség rengeteg.
- Az XML teljes Unicode támogatása miatt könnyen megoldható a nem latin karakteres írások feldolgozása. Az internet széles körben való elterjedése miatt ez nagyon lényeges elem, mivel nem kell Európán kívül menni, hogy a latin abc-től különböző karakterekkel találkozzunk. A cég ügyfelei között is van, akinek a kulcsszavai tartalmazzak nem latin karaktereket, tehát éles, ipari körülmények között tapasztalhattam meg az Unicode előnyeit.

Összefoglalás

Dolgozatom célja volt az egyik legnépszerűbb, interneten is széles körben használt szabvány, az XML által nyújtott lehetőségek áttekintése.

Láthattuk, hogy éles körülmények között milyen előnyei származnak egy cégnek ha szabványos megoldásokat használ: egyszerűbben, könnyebben és gyorsabban lesz képes fejleszteni, hamarabb használhatja más cégek szintén szabványon alapuló megoldásait és könnyebb a későbbiekben integrálnia új, szintén a szabványon alapuló, szolgáltatásokat. Valamint későbbi fejlesztések hozzájárulása a programhoz szintén egyszerűbben történik.

Későbbi fejlesztésekre is biztosan szükség lesz, mert – ahogyan az informatikában minden – ez is egy olyan szegmens, ami még messze nem érte el a határait. Konkrét példánál maradván a dolgozatomban is említett DoubleClick felvásárlás is olyan kérdéseket és megoldandó problémákat fog felvetni, amivel minden bizonnyal foglalkozni kell a rövid időn belül. Ám ezek olyan távlati problémák (még ha csak hetekben legrosszabb esetben is csak hónapokban mérhető is), amivel jelenleg nem tudtam foglalkozni.

A program továbbfejlesztése is várható, mert az AdWords és az AdSense is hatalmas fejlődésen ment keresztül az elmúlt hónapokban, és várhatóan ez folytatódni is fog.

A felhasznált forrásanyagon is látszik, hogy – szintén ahogy említettem az elején – a “Google-éra” gyermekeként nagy mértékben támaszkodok az internet adta lehetőségekre – főleg a blogokra. A magyar keresőmarketing szakemberei olyan korrekt blogokat írnak, amikből a témához nem is értő emberek is nyugodtan meríthetnek, ha erre a témakörre kíváncsiak.

Irodalomjegyzék

<http://hu.wikipedia.org/wiki/XML>

<http://www.w3.org/XML/>

<http://www.w3c.hu>

<http://keresomarketing.wordpress.com/>

<http://onlinemarketing.blog.hu/>

<http://reklam.sfblogs.net/>

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

<http://www.php.net>

<http://www.mysql.com>

<http://www.apache.org>

<http://googleblog.blogspot.com>

Köszönetnyilvánítás

Hálás köszönetemet szeretném kifejezni a következő személyeknek, akik nélkül nem jutottam volna el idáig:

- **Édesanyámnak**, aki mindig hitt bennem és támogatott
- Témavezetőmnek, **Dr. Juhász Istvánnak**, az elmúlt 3 év támogatásáért
- **Szabó Csillának**, aki nélkül biztosan elvesztem volna valahol, valamikor
- Az **ONIK Öntevékeny Csoport** minden tagjának – nélkületek, bizony, szintén elvesztem volna
- **Barátaimnak**, akik mögöttem állnak és támogatnak mindig, minden körülmények között.