

Osztályozási rendszerek a világhálón

1. Bevezető

A World Wide Web (WWW, röviden Web) és vele az általánosan hozzáférhető és felhasználóbarát felületek megjelenését követően viharos gyorsasággal születtek meg és fejlődtek hatalmas rendszerekké a különféle *keresőszolgáltatások*, mint például az Alta Vista, HotBot, Infosec, Magellan, excite, Yahoo*, Magyarországon pedig az Alta Vizsla, Heuréka, Hudir. Ezekre a rendszerekre szemantikai szempontból nemcsak a hagyományos online adatbázisokhoz képest könnyebb kezelhetőség, hanem az asszociatív gondolkodással jobban összhangban álló keresési struktúrák is jellemzők.

A Web keresőszolgáltatásaiban alkalmazott információkereső nyelvek/osztályozási rendszerek a felhasználó szempontjából fontos jellemzői a következők:

- kizárólag természetes nyelven alapulnak;
- rendkívül rugalmasan használhatók;
- rendkívül változatos típusú információk kereshetők egyazon információkereső nyelvvel/osztályozási rendszerrel (a dokumentumleírásoktól kezdve a teszteti információkon, menetrendeken, telefonszámokon, statisztikai adatokon, utazási információkon át a magánjellegű tényadatokig).

Könyvtáros szempontból figyelemre méltó és nem valami köztudott, hogy a hálón belül ugyanúgy megjelent az a két alaptípus, amely a könyvtári világban hagyományos: az egyedi szavakat alkalmazó „nyelv”, ill. a hierarchikus struktúrákba szervezett osztályozási rendszer.

Az előbbieket túlnyomórészt automatikus indexeléssel feltárt indexszavas (szabad szövegszavas) információkereső nyelvek. Az ilyen keresőnyelvet használó szolgáltatások (Hot Bot, Infosec, Heuréka stb.) leszedői általában kötött szótár nélkül válogatják ki a szövegszavakat. Ritkán előfordulnak közöttük már olyanok is, melyekben kötött szótárat, néhányukban (pl. Alta Vizsla, Kolibri) tezauruszt is használnak. Ez utóbbi szolgáltatások információtételeinek relevanciája általában lényegesen nagyobb.

Az utóbbiak hierarchikus osztályozási rendszerek (melyeket a Yahoo, Hudir stb. keresőszolgáltatásokban alkalmaznak). Az osztályozási rendszert és a benne többnyire intellektuálisan feltárt információtételek (html-rekordok, tényadatok) együttesét szokták nevezni Internet katalógusnak. Bennük a hierarchikusan strukturált osztályok szerint rendeződnek a tételek.

Ugyanaz a keresőszolgáltatás sokszor mindkét rendező rendszerrel szolgált (pl. Alta Vista, Lycos, Magellan, Hudir).

* A pontos márkanév a felkiáltójelet is tartalmazza (Yahoo!), melyet a jobb olvashatóság kedvéért elhagytunk.

Az alábbiakban a Web keresőszolgáltatásaiban alkalmazott osztályozási rendszerek néhány jellemzőjét fogjuk áttekinteni.

2. Az Internet-katalógusok osztályozási rendszerei

A Web keresőszolgáltatásaiban alkalmazott osztályozási rendszereket a Yahoo példáján mutatjuk be, mely a jelenleg ismert katalógusok között egyike a leggazdagabb választékot kínáló szolgáltatásoknak. Belépőlapja (portálja) az 1. ábrán látható.

Yahoo! Auctions - 1000's of items to bid on - <u>Pokemon</u> , <u>Beanie Babies</u> , <u>video games</u> , <u>Furbys</u> ...		
<u>Shopping</u> - <u>Yellow Pages</u> - <u>People Search</u> - <u>Maps</u> - <u>Travel Agent</u> - <u>Classifieds</u> - <u>Personals</u> - <u>Games</u> - <u>Chat</u> <u>Email</u> - <u>Calendar</u> - <u>Pager</u> - <u>My Yahoo!</u> - <u>Today's News</u> - <u>Sports</u> - <u>Weather</u> - <u>TV</u> - <u>Stock Quotes</u> - <u>more...</u>		
Arts & Humanities <u>Literature</u> , <u>Photography</u> ...	News & Media <u>Full Coverage</u> , <u>Newspapers</u> , <u>TV</u> ...	In the News · <u>NATO - Serbia war</u> · <u>New solar system discovered</u> · <u>Rodman released by Lakers</u> <u>more...</u>
Business & Economy <u>Companies</u> , <u>Finance</u> , <u>Jobs</u> ...	Recreation & Sports <u>Sports</u> , <u>Travel</u> , <u>Autos</u> , <u>Outdoors</u> ...	Marketplace · <u>Charity Auctions</u> - for the Kosovo relief effort · <u>Get ready for Mother's Day</u> <u>more...</u>
Computers & Internet <u>Internet</u> , <u>WWW</u> , <u>Software</u> , <u>Games</u> ...	Reference <u>Libraries</u> , <u>Dictionaries</u> , <u>Quotations</u> ...	Inside Yahoo! · <u>Y! Movies</u> - showtimes, reviews · <u>Y! Clubs</u> - create your own · <u>Y! Visa</u> - instant credit while you wait <u>more...</u>
Education <u>College and University</u> , <u>K-12</u> ...	Regional <u>Countries</u> , <u>Regions</u> , <u>US States</u> ...	
Entertainment <u>Cool Links</u> , <u>Movies</u> , <u>Humor</u> , <u>Music</u> ...	Science <u>Biology</u> , <u>Astronomy</u> , <u>Engineering</u> ...	
Government <u>Military</u> , <u>Politics</u> , <u>Law</u> , <u>Taxes</u> ...	Social Science <u>Archaeology</u> , <u>Economics</u> , <u>Languages</u> ...	
Health <u>Medicine</u> , <u>Diseases</u> , <u>Drugs</u> , <u>Fitness</u> ...	Society & Culture <u>People</u> , <u>Environment</u> , <u>Religion</u> ...	

1. ábra. A Yahoo Internet katalógus belépőlapjának részlete, melyen az osztályozási rendszer legfelső szintje látható.

Ezeknek az osztályozási rendszereknek az osztályait, függetlenül azok szintjétől, a szolgáltatók általában „kategóriáknak” nevezik. Ez és sok más elnevezéssel eltérés a hagyományostól feltehetően éppen abból ered, hogy a készítőknél nem is tudatosult: olyan rendező rendszert terveztek és használnak, melybe besorolják, osztályozzák az információtöteleteket. Innen nézve nem a rendszer logikai (kategoriális), hanem besoroló, „tartalmazó” szerepéről van szó, azaz dolgok (html rekordok) osztályairól (nem pedig html rekordok „kategóriáiról”).

A szisztematikus rendezés hiánya

A tervezők hagyományoktól való érintetlensége abban is megmutatkozik, hogy az egyes szinteken az ilyen típusú rendszerek többségében az osztályokat nem szisztematikus, hanem betűrendben jelenítik meg. Indokaik kétségtelenül

nyomósok: a lehető legkevesebb szellemi erőfeszítést szeretnének okozni a végfelhasználónak. A legfelső szinten még nem annyira feltűnő, hogy a hierarchikus rendszer adott szintjén össze nem tartozó osztályok kerülnek egymás mellé, mert ezen a szinten minden keresőszolgáltatásban a lehető leggyorsabb áttekintésre töreksenek: egy pillantással lehessen fölmérni, hogy a rendszer lényegében mit és hol tartalmaz. Az alsóbb szinteken azonban szokatlan találkozások adódnak. A Science (Tudomány) második szintjének több mind 60 osztálya például így kezdődik: Acoustics (Akusztika), Agriculture (Mezőgazdaság), Alternative (Alternatív technikák), Amateur science (Amatőrök által művelt szakterületek), Anthropology and Archeology (Embantan és régészet), Artificial Life (Mesterséges élet) stb.

A hierarchikus rendszer nem különösen „mély”: alig 3-4 szintet tartalmaz. Ezért jelenik meg a második és a harmadik szinten olykor nagyon sok osztály. A szerkesztők valószínűleg nem mernek a már széles körben megismert főszerkezeten változtatni; ilyen változtatás nélkül azonban nem oldható már meg, hogy az egyes szinteken az osztályok számát csökkentse. Az egész emlékeztet a természetek hangyaboly-építményeire: a fejlődés szerves és nagyon gyakorlatias, mindig kizárólag a lehetőségekhez igazodik, sohasem elvekhez. Kétségtelen, hogy az elvek alkalmazásának vannak praktikus határai. De az is igaz, hogy a prakticitás túlfeszítéséből is adódnak határok. Van, amikor már nincs megtevesztőbb, mint a realitás.

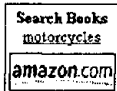
A polihierarchikus szerkezet

Az eddig megjelent átfogó Internet katalógusok egyetemes célú osztályozási rendszereit nem jellemzi a felosztási szempontok következetessége. Érezhető, hogy kereskedelmi szempontok érvényesülnek az osztályok fölvételében: az a felfogás, hogy „mi van azon a szakterületen eladható információ”. Csak feltételezzük, hogy a keresőszolgáltatások gépei által feldolgozott információtételek mennyiségének növekedésével a rendszerek finomszerkezete tartalmilag fokozatosan koherensebbé válik. Ugyanakkor az alkotók szakmai érintetlenségének előnyei is vannak: friss szemmel vágta neki a világ rendező célú felosztásának, s ez hosszabb távon nem maradhat következmények nélkül a hagyományos könyvtári és dokumentációs osztályozásra sem.

Különösen hasznos megoldások születtek az ilyen osztályozási rendszerek hierarchialáncai között. Ennek alapja, hogy a hypertext a kereszthivatkozások eszményi rendszere, és ezt hasznosítják a hierarchikus szerkezeten belül is. Itt is létrehozhatnak keresztirányú összefüggéseket. Ez abban nyilvánul meg, hogy egy osztály egyszerre több magasabb szintű osztály alárendeltje is lehet, az osztályozási rendszerek tehát – szemben a hagyományos egyetemes könyvtári rendszerekkel – polihierarchikusak. Ez olykor rendkívül bonyolult, néha már lehetetlennek tűnő struktúrákat eredményez, de a felhasználót nagyon jól szolgálja, mert az ismétlődések következtében a hierarchikus rendszer redundnáns.

A 2. ábrán azt láthatjuk, hogy például a Motorcycles (Motorkerékpárok) hány különféle hierarchialáncon belül jelenik meg. Mindig van egy „gazdaosztály” („szülőosztály”), melyhez a polihierarchikusan alárendelt alosztály kapcsolódik (a többi előfordulást a megjelenítésben a @ jellel jelölik).

Categories	Web Sites	Web Pages	Related News	Net Events
Yahoo! Category Matches (1 - 15 of 15)				
Recreation > Automotive > Motorcycles				
Business and Economy > Companies > Automotive > Motorcycles				
Net Events > Recreation > Automotive > Motorcycles				
Recreation > Automotive > Motorcycles > Feet Forwards Motorcycles				
Business and Economy > Companies > Financial Services > Insurance > Automotive > Motorcycles				
Recreation > Hobbies > Models > Motorcycles				
Business and Economy > Companies > Automotive > Motorcycles > Manufacturers > Honda Motorcycles				
Recreation > Sports > Motorcycle Racing				



2. ábra. A Motorkerékpárok (Motorcycles) polihierarchikus előfordulása a Yahoo osztályozási rendszerében.

A helyzet azonban ennél bonyolultabb. A szerkesztők friss szemléletét minden jel szerint nyelvészeti szempontok sem kötik gúzsba: nem sokat foglalkoznak például a homonímák megkülönböztetésével. Gyakori, hogy ugyanazzal a névvel nem ugyanazt az osztályt jelölik, hanem másikat, amelynek vagy nem ugyanaz a terjedelme (nem azonosak a hozzá besorolt információtételek), vagy nem ugyanaz a felosztása (nem azonosak az alatta megjelenő alosztályok). Például a Motorkerékpárok osztályai között van olyan, amelynek sokkal részletesebb a felosztása, mint a többinek. A „Recreation – Automotive – Motorcycles” és a „Recreation – Hobbies – Models – Motorcycles” láncban a Motorkerékpárok osztálya sokkal több alosztályt tartalmaz, és ezek elemzése alapján e két osztályelőfordulást a következőképpen kellene megkülönböztetni a többi, ugyanilyen nevű osztálytól: „Motorkerékpárok (a szabadidő és a barkácsolás szempontjából)”. A szerkesztők nyilván abból indulnak ki, hogy maga a hierarchialánc is definiálja a jelentést. Hozzá kell azonban tenni, hogy „adott esetben”. Más esetekben ugyanis eltérő hierarchialáncokban ugyanaz az osztály szerepel (pl. Motorkerékpárként), azaz az eltérő hierarchialánc nem definiál eltérően.

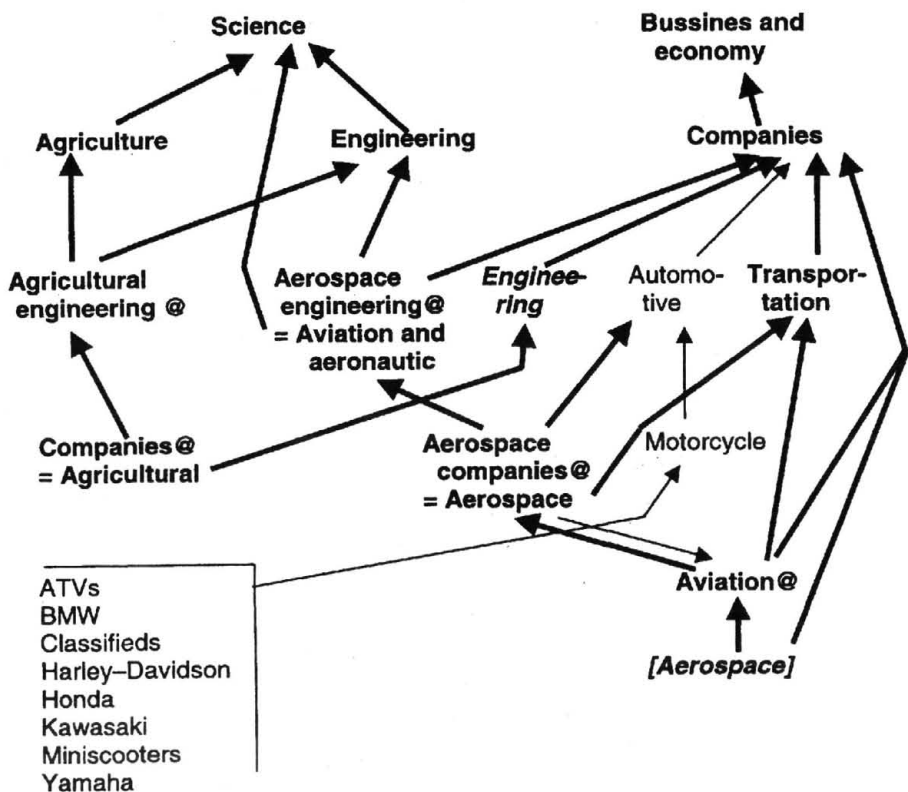
A struktúrák gazdagsága

Hogy ezeknek az osztályozási rendszereknek a rejtett szerkezeti bonyolultságát jobban lássuk, a 3. ábrán a Yahoo osztályozási rendszerének egy részletét kiemeltük, és címkézett irányított gráffal ábrázolva mutatjuk meg.

Az előbbieken tárgyalt Motorkerékpárok osztály összefüggéseit a jobb elkülöníthetőség kedvéért nem félkövéren jelentettük meg.

A gráf alapján a következők ismerhetők fel.

- Az Agricultural engineering (Agrotechnika) egyrészt az Agriculture (Mezőgazdaság), másrészt – @ jelöléssel – az Engineering (Mérnöki tudományok/Technika) alosztálya.



3. ábra. A Yahoo polihierarchikus osztályozási rendszerének részlete címkézett, irányított gráf formájában.

- Az Aerospace engineering (Repüléstechnika) az Engineering és a Companies (Cégek), továbbá Aviation and aeronautic (Légügy/Repüléstan) néven a Science (Természettudomány) alosztálya, mely utóbbinak ugyanakkor tranzitív alárendeltje.

Az, hogy ugyanazt az osztályt más néven a tranzitív fölrendelt alá rendeljék, hajmeresztő a hagyományos osztályozási rendszerek ismerőjének (olyan ez, mintha a Kutyát egyrészt alárendelik a Háziállatnak, ugyanakkor Eb néven az Állatnak, melynek ugyanakkor a Háziállat a közvetlen alárendeltje). A piaci viszonyok terén iskolázott rendszertervező viszont abból indulhatott ki, hogy a Természettudományok felől nézve jobban fest az általánosabban megfogalmazott osztálymegnevezés (Légügy...), nem pedig a Repüléstechnika, amely viszont a Technika felől nézve adekvátabb osztálynév.

Azt is észre kell venni, hogy az Aerospace engineering az Engineering alá rendelve valójában olyan osztályt képvisel, amely a repüléstechnikára vonatkozó *információk* tételeit tartalmazza, a Companies alá rendelve pedig azt, amely a repüléstechnikával foglalkozó *cégek* információit tartalmazza. Ennek a példának az esetében nincs a Yahooban különbség a két osztály terjedelme (információtételei) között.

- Az Engineering esetében azonban van. Ebből ugyanis két osztályt találunk, de ez a két osztály nem ugyanaz: a Cégeknek alárendelt osztály ugyanis – melyet dőlt betűvel jelenítettünk meg – csak a műszaki tevékenységeket végző cégek információit tartalmazza, a Természettudományoknak alárendelt Engineering ezzel szemben minden, a technikára és a műszaki tudományokra vonatkozó információit tartalmazza.
- A dőlt betűvel megjelölt *Engineering* alárendeltje az Agricultural (Mezőgazdasági) [így, jelzősen], mely az agrotechnikai cégek információit tartalmazza. Ugyanennek az osztálynak az Agrotechnika alárendeltségében viszont Companies (Cégek) a neve. Ha belegondolunk, ez egész logikus: az Agrotechnika felől nézve cégekről, a műszaki cégek felől nézve meg „mezőgazdaságról”, azaz Agrotechnikai (cégekről) van szó.
- Talán a legmerészebb húzás, amikor ugyanazt az osztályt alárendelik egy másiknak, ugyanakkor fölérendelik neki. Ez a helyzet az Aerospace (= Aerospace companies @) és az Aviation között. De ha meggondoljuk, hogy ezekben az osztályozási rendszerekben egyáltalán nincs pontosan meghatározva, hogy mit is értünk tulajdonképpen azon a reláción, amely az egyes osztályokat összekapcsolja, ez a megoldás korántsem olyan hajmeresztő, mint ahogy logikai szempontból látszik. Eddig ugyanis abból indultunk ki, hogy az Internet katalógusok osztályozási rendszerei hierarchikusak, és alapvetően csak alá-fölérendeltségi kapcsolatok vannak. Valójában azonban olyan rendező rendszerekről van szó, melyekben nincs egyértelműen definiálva a kapcsolat: lehet hierarchikus (az esetek többségében), de van, amikor egyszerűen csak annyit jelent, hogy „lásd még”. Az Aerospace és az Aviation között valójában az utóbbi összefüggésről lehet szó, és ez logikailag teljesen megengedett. Más lapra tartozik, hogy ezekben az osztályozási rendszerekben a mindenkori, definiálatlan relációt csak az jelöli, hogy „az egyik következik a másik után”. Ha a teauruszszabvány szerint pontosan jelölnénk a tárgyalt esetet, a 4. ábrán látható szócikket kapnánk:

Transportation		Aerospace		Aviation	
A	Aerospace	F	Transportation	F	Transportation
	Aviation	X	Aviation	X	Aerospace

4. ábra. Yahoo összefüggések szabványos teauruszcikk formában

Az osztályok típusai

A Yahoo belépőoldalát vizsgálva többféle osztálytípus különíthető el. Könyvtárszervezési szempontból azt mondanánk, hogy ahány osztálytípus, annyiféle gyűjtőköri forrástípus.

Szakterületek, tudományok, tevékenységi körök

Arts & Humanities (Művészet és társadalomtudomány)

Bussines & Economy (Kereskedelem és gazdaság)

Computers & Internet (Számítástechnika & Internet)

Education (Oktatás-művelődés)

Stb.

Ezek az osztályok felelnek meg a dokumentumok hagyományos osztályozási rendszereiben alkalmazott osztályoknak, de itt is lépten-nyomon érheti az embert meglepetés: valamelyik szakterületen belül felbukkanhat apróhirdetéseket tartalmazó osztály vagy tényadatokat tartalmazó osztály stb.

Kereskedelmi jellegű osztályok:

Shopping (Bevásárlás)

Travel Agent (Utazási irodák)

Classified (Apróhirdetések, üzleti)

Personals (Apróhirdetések, személyi)

Ezek elsősorban arra valók, hogy az adás-vételt támogassák. Az osztályok erősen válogatott, csak a rendeléssel szembeni igényekből szűrődő szakterületek. Ezekben belül a besorolt információtetelekből kiindulva megrendelhető árucikkek, utazáshoz jegyek, elérhetőek a hirdetések feladói.

Adattárak, címek, helyek osztályai

Yellow Pages (Szakmai telefonkönyv)

People Search (Drótpostacím és személykeresés)

Maps (Térképek)

Ezekben az osztályokban cégek, személyek adatait tartalmazó információtetelek találhatóak, a térképek esetében pedig helyek azonosíthatók vizuálisan.

Hírek, tényadatok

Today's news (Aktuális hírek)

Stock Quotes (Tőzsdehírek)

Sports (Sporthírek)

Weather (Időjárásjelentés)

TV (Tévéadás)

Ezekben az osztályokban tényadatok szerepelnek.

Segítségek, gondúzó

Calendar (Naptár, események)

Games (Játékok)

Pager (Letöltő)

My Yahoo (Testre szabott Yahoo)

Yhooligans (Kapcsolatok)

E-mail (Drótposta bejelentkezés)

Ezekben az osztályokban a mindennapokban hasznos eszközök találhatóak meg játékok. Valójában nem is tekinthetők osztályoknak, hanem speciális szolgáltatások belépőpontjainak.

Külön említést érdemel a Yahoo esetében, hogy már nagyon sok nyelven elérhető (a World Yahoo osztály alatt található meg az egyes nyelvi változatok, melyek nem pontos másolatai az angolnak, hanem az adott ország körülményeihez alkalmazkodó variációk (van már kínai nyelvű változat is).

Az elérhetőség nyelvi meghatározása mellett a regionális finomítás is lehetséges (Yahoo Get Local).

3. Összegezés

Összefoglalva azt mondhatjuk, hogy ezekben az osztályozási rendszerekben számos rendkívül rugalmasan alkalmazott megoldásra bukkanunk, ezek jelentős része a Web körülményei között akkor is beválik, ha logikailag ellentmondásos, de előfordulnak a rendszer koherenciáját gyengítő megoldások is, melyek a későbbi fejlődés során bonyodalmakat okozhatnak. Például akkor, ha a hierarchikus összefüggések mentén automatikus találatbővítést vagy -szűkítést kellene megvalósítani. Egy ilyen esetben nem kerülhető meg az összefüggések ellentmondásmentes megvalósítása. A Web osztályozási rendszereinek fejlődésében jelentős tartalékok vannak, és előbb-utóbb megtermékenyítő hatásuk lesz a hagyományos osztályozási rendszerekre.

Ungváry Rudolf

