

Dr. NÉGYESI Imre

AZ INFORMÁCIÓGYŰJTÉS JÖVŐKÉPE

THE VISION OF THE INFORMATION COLLECTON

Évszázadokon keresztül, ha valaki információt akart gyűjteni valamilyen témában, akkor elsősorban a papír alapú dokumentumokra támaszkodhatott. A tudományok haladásával szükségszerűvé vált, és lehetőség is nyílt arra, hogy elsősorban a világháló segítségével, gyorsabban illetve kényelmesebben szerezzünk új ismereteket bármely témáról is legyen szó. A számítógépes hálózatok — elsősorban az Internet elterjedése — azonban nem csak lehetőségeket, hanem „veszélyeket” is magában rejtnek az emberekre nézve, amelyeket sohasem hagyhatunk figyelmen kívül az információk keresése közben.

Through centuries, if somebody wanted to collect information in some a kind of topic, then primarily the paper basis may have leant on documents. Became necessary with the progress of the science and an opportunity overlooked it in order for us to obtain new knowledge on any topics with the help of the web, more quickly and more comfortably primarily let him be a word. The computerized networks and primarily the internet spreading however not only opportunities, but it hides dangers alone with regard to peoples, that we may never leave dangers out of consideration during the search of the information.

BEVEZETÉS

Napjaink egyik vezető kutatási témája az e-társadalom, az e-kormányzat fejlesztési lehetőségeinek vizsgálata, az állampolgárok számára nyújtandó e-szolgáltatások lehetőségeinek kiszélesítése. Ugyanakkor az új rendszerek kifejlesztése során kiemelt figyelmet kell fordítani az információ biztonságra, mert e nélkül egyetlen új rendszer bevetése sem lehetséges. A 2000. év egyik vezető híre volt egy az egész Földet behálózó földkörüli „információs pajzs” létének beismerése. Ennek a lehallgató rendszernek az ismertetése csak az indító gondolatát képezi ennek a cikknek, ezért nem térek ki politikai kérdésekre, és informatikai szempontból is csak egy rövid ismertetésre szorítkozom. Nem céлом tehát a politizálás, nem kívánok beleszólni nemzetbiztonsági szakmai kérdésekbe, csak szeretném a már ismert tényeket összekapcsolni az informatikai szakmai háttér elemeivel, és ezáltal egy rövid helyzetképet adni a dokumentumkezelés jelenlegi helyzetéről. Kiindulásként természetesen nem kerülhetjük meg a kezdeteket, ezért kezdjük egy rövid történelmi áttekintéssel.

EGY KIS TÖRTÉNELEM

Az egész cikk mottója lehetne az a két idézet, amely bár ellentmond egymásnak mégis, mintegy keretbe foglalja a jelenlegi helyzetet, és ezért használják előszeretettel a témával foglalkozó írásokban. Az egyik Henry L. Stimsonnak — aki az Amerikai Egyesült Államok külügyminisztere volt — még az 1920-as évek végén elhangzott mondata, mely szerint: *„úriember nem olvassa mások levelét”*. A szállóigévé vált mondas alapján 1929. október 31-én bezárta kapuit az USA rejtjelző szolgálata, de sajnos a történelem nem igazolta Stimson. Nem sokkal Stimson *„úriember”* társadalmi elképzelése után, 1949-ben jelent meg Georg Arthur Orwell angol író *„Ezerkilencszáznyolcvannégy”* című regénye. Ebben a műben a szerző — az irodalomtörténészek meghatározása szerint — egy 20. századi utópiát vetít elénk, mely szerint *„az embernek, annak tudatában kellett élnie, hogy lehallgattak minden hangot, amit kiadott, s az állampolgárok minden mozdulatát megfigyelték”*. [1] Mára már bebizonyosodott, hogy inkább Stimson mondatai tekinthetők utópiának és Orwell fantáziája nem szárnyalta túl napjaink valóságát.

Az egész világot behálózó műholdas figyelő rendszer létének beismerése csak a sajtó nyomására napvilágra került dokumentumokból vált ismerté a közelmúltban. Ez az ECHELON- (magyarul: harcvonat) rendszer mai is a legnagyobb titokban, a titkosszolgálatok felügyelete alatt működik. A rendszer kiépítésének kezdetét a II. világháború idején megszületett BRUSA COMINT (communications intelligence) egyezmény képezte, amelyet 1943. május 17-én az Egyesült Királyság valamint az USA titkosszolgálatai ratifikáltak. Így jött létre 1948-ban a titkos UKUSA megállapodás, illetve szövetség, amelynek tartalma és hatálya napjainkban is érvényes. A szerződés alapító szervezetei (NSA: USA-Nemzetbiztonsági Szolgálata, GCHQ: Anglia-Government Communications Head Quarters, CSE: Canada-Communications Security Establishment, GCSB: Új-Zéland-General Communications Security Bureau) később további országokkal bővült. Jelenleg az UKUSA szövetségesek közé tartozik Németország, Japán, Norvégia és Törökország is. [2]

AZ ECHELON RENDSZER

Az ECHELON-rendszer terve egyszerű és világos: földi megfigyelő bázisok, kémhajók (hét tengeren), tengeralattjárók továbbá szigorúan titkos műholdak figyelik az egész Föld, az egész emberiség globális kommunikációs hálózatának forgalmát. Lehallgató állomásokat létesítettek a Glóbuszon valamint a világűrben, amelyek lehallgatják az összes műholdas, mikrohullámú, mobil illetve optikai-szálalás kommunikációs forgalmat, majd továbbítják mindezt a mérhetetlen mennyiségű információt az ECHELON számítógépes rendszerébe. Ez a rendszer a legkorszerűbb hang és optikai karakterfelismerő (OCR) programokat, valamint olyan kódszavas illetve kifejezés szótáron alapuló (ennek neve ECHELON Dictionary) szövegfelismerő rendszert tartalmaz, amely kiválogatja a kívánt üzeneteket, majd kódolt jelzés kíséretében rögzíti azokat további elemzés céljára. A lehallgató állomásokon működő intelligens analízátorok a rögzített beszélgetést vagy dokumentumot összevetik a kulcsszó illetve kifejezés listával, és ez alapján továbbítják (vagy nem) a megfelelő hírszerző központba, ahol az „illetékesek” eldöntik, hogy szükséges-e a további lehallgatás? Az ECHELON hálózat gerincét képező lehallgató állomásokat az Intelsat és Inmarsat műholdakról irányítják, amelyek így képesek kezelni azt az óriási mennyiségű telefon, fax, stb. kommunikációt, amely az országok illetve kontinensek között történik. Különös jelentősége van a számtalan rádió-frekvenciás lehallgató állomásnak, mivel napjaink katonai és civil kommunikációjának jelentős része rádió frekvencián történik.

A lehallgatott üzenetek elektronikus úton kerülnek az óriási számítógépes rendszerbe, ahol a legmodernebb hang és optikai karakterfelismerők, valamint tartalomelemző rendszerek elemzik a rendezetlen információ halmazt. Ezek a programok és számítógép-rendszerek sok szempontból már a jövőbe vezetnek, továbbá olyan alrendszereket tartalmaznak, amelyek *egyrésztől* vezérik a kulcsszó kereső programokat, *másrésztől* tartalomelemzéssel válogatják szét az üzeneteket, és egy óriási szöveges adatbázisba rendezik azokat, amelyből a kívánt dokumentumok (üzenetek) kulcsszavak alapján már könnyen hozzáférhetőek. A legerőteljesebben a jövőbe mutató alrendszer pedig az, amely hangfelismerő programok segítségével konvertálja a beszélgetéseket szöveges üzenetké, továbbá képes egyéni hangminták alapján a beszélőt azonosítani, és tárolni a jövőbeli elemzés céljára.

INFORMATIKAI ALAPOK, ÉS A JELEN [3]

A röviden ismertetett rendszerről leegyszerűsítve kimondhatjuk, hogy üzenetek lehallgatását végzi, majd a lehallgatott dokumentumokat feldolgozza. A dokumentumfeldolgozás jelenleg a papír alapú dokumentumokra épül, az aktákat beszkennek vagy elektronikusan konvertálják, és egy nagyfelbontású kép formájában tárolják merevlemezeken illetve optikai adattárolón. A fájlt ezután is meg lehet nézni, nyomtatható, megosztható és tárolható, de a képfeldolgozás óriási lehetőséget biztosít azzal, hogy a dokumentum tartalma aktív lesz. A betűk így nem csak festéket jelentenek a papíron, mert az optikai karakter felismerés (OCR) segítségével a szöveg a számítógépes rendszer által is olvasható és szerkeszthető lehet.

Dokumentumot háromféleképpen lehet bevinni a rendszerbe: beolvasással (szkenneléssel), konverzióval, importálással. Szkennelés során a dokumentumról raszteres kép készül, mely a számítógépen eltárolható. A dokumen-

tumok konvertálása azt jelenti, hogy az elektronikus dokumentumokat az alkalmazás nem módosítható képpé alakítja át azért, hogy a képfeldolgozó (imaging) rendszeren belül tárolja. A konverzió során a dokumentumról készül egy szöveg fájl is, hogy az alapján is vissza lehessen rá keresni (full-text indexelés). Az importálás a másik módszer elektronikus fájloknak a dokumentum menedzsmenet rendszerbe juttatására.

A karakterfelismerésnek három fő csoportja van:

1. *OCR (Optical Character Recognition* – optikai karakter felismerés), amely kifejezést a karakterfelismerésre általánosságban szokták használni, de leginkább a nyomtatott szövegek felismerését értjük alatta.
2. *ICR (Intelligent Character Recognition* – intelligens karakter felismerés), amely a nyomtatott betűs kézírás felismerésére szolgáló technológia.
3. *OMR (Optical Mark Recognition* – optikai jel felismerés), amely különböző jelölések, jelölőnégyzetek felismerésére használható.

Alapvetően kétféle elv ismeretes a karakterfelismerésre:

1. a bittérképes betűfelismerés vagy mintaillesztés (matrix matching): a program körülhatárol egy betűt, és ezt numerikus számsorozatként értelmezi a tulajdonságvektorai alapján. Ennek a képét összeveti az eltárolt képekkel pontról-pontra. Ha valamelyik eltárolt képpel kellően nagy százalékban egyezik az összehasonlított kép, akkor elfogadja annak a betűnek;
2. az objektum orientált felismerés, más néven kontúranalízis. Itt a szoftver az egyes betűk vonalait, vonalkapcsolatait elemzi, tulajdonképpen a bittérképet vonalakra bontja. Ezeket a vonalakat azután összeveti az általa ismerttel, és ennek alapján azonosítja be az egyes karaktereket. Ez a módszer sokkal közelebb áll az emberi gondolkodáshoz és felismeréshez. Itt a betűtípus és betűméret nem jelent komoly akadályt, hiszen a betűket jellemző alakjuk alapján ismeri fel. Annál nehezebb viszont a tanítás és az újabb karakterek megadása.

A karakterfelismerés folyamata során először az egyes karaktereket tartalmazó képpontokat csoportosítja egybe a szoftver (szegmentálás). Ez nem egyszerű feladat, mivel egy karakter nem mindig egy elemből áll (például az ékezetes karakterek, felkiáltójel, stb.). Ha a beolvasott kép túl világos, az csak fokozza a problémát, mert a karakterek széttöredezhethetnek, és így nem lesznek felismerhetőek. Ellenkező esetben, ha a kép túl sötét, vagy vastag betűtípusról van szó, akkor pedig összeérnek a karakterek, és így nehezíti meg a felismerést. Igaz, hogy folyamatosan karakterfelismerésről beszélünk, de a lényeg az, hogy ez alapján a szöveg tartalma legyen ismert. Ehhez nagy segítséget nyújtanak a nyelv-specifikus információk, amelyeket a beépített felismerési nyelvek tartalmaznak. Így abban az esetben, ha a program döntés előtt áll, hogy vajon milyen betűnek a képpontjait dolgozza fel, minden lehetséges megoldást értékel.

A JÖVŐ LEHETŐSÉGEI

Ha a karakterfelismerést, mint módszert tekintjük, ebből a rövid leírásból is látható, hogy számos nehézséget rejt. Ebből egyenesen következik, hogy keresni kell azokat a módszereket, amelyek segítségével az információgyűjtés módszere javítható. Ez a jövőben használható módszer lehet a szemantikus, tartalom alapú keresés, amely forradalmasíthatja az információgyűjtést. [4]

A hagyományos keresőrendszerek ugyanazon elv alapján működnek. A keresőrobotok meglátogatják és összegyűjtik az egyes oldalakat, majd a megtalált oldalakat egy indexelő elemzi, kigyűjti belőlük a releváns — azaz az indexelendő — kifejezéseket, majd listát készít belőlük. A felhasználó által feltett keresőkérdést a keresőrendszer elemzi, feldolgozza és a korábban beindexelt oldalak között elvégzi a keresést, ezután sorrendbe helyezi a találatokat. Az ilyen rendszerek „lelke” tehát az indexelésben valamint a találatok sorrendezésében rejlik. A hagyományos keresés legnagyobb problémájának így az tekinthető, hogy az ilyen jellegű keresők nem foglalkoznak a talált dokumentumok valamint a keresőkérdés jelentésével, hanem kizárólag annak szöveges alakjával. Ebből több probléma is fakadhat, mivel az információ-visszakeresés túlzottan a nyelvi információkra hagyatkozik. Így egy bizonyos nyelvű keresésre csak az adott a nyelven ad találatokat (szöveges találatokat) miközben bizonyos esetek-

ben egy-egy képi találat hasznosabb lehet, mint a szöveges. (Ráadásul a képek által tárolt információkat sem képek a keresők automatikusan kinyerni.)

A kulcsszó alapú keresés tehát a hagyományos keresőmotorok használatakor gyakran nem vezet eredményre. A felhasználónak ugyanis olyan keresőkifejezést kell megadni, amelynek elemei (szavai) „véltetően” szerepelnek majd azon az oldalon, ahol az információigényt kielégítő tartalom is szerepel. Ehhez a felhasználónak ki kell találnia, hogy milyen szöveggörnyezetben szerepelhet a keresett tartalom, ennek hiányában ugyanis a keresése sikertelen lesz. További gondot jelenthet, hogy túl általános kifejezéseket használva feldolgozhatatlanul nagy mennyiségű válaszdalt ad vissza a kereső, míg pontosan specifikált keresőkifejezések szavai együttesen gyakran egyetlen dokumentumban sem fordulnak elő. A felhasználónak alkalmazkodnia kell a gépi keresés technológiájához, igazán eredményesen csak akkor tudja a keresőket használni, ha megérti azok működési elvét, és sajátjává teszi ezt a „gondolkodásmódot”.

A megoldást a szemantikus világháló irányzat jelentheti, amelynek lényege és célja, hogy jelentést vigyen a webre úgy, hogy a webes tartalmakhoz metainformációkat (azaz információkat a dokumentumban található információról) társítson, és lehetővé tegye, hogy ezeken az információkon következtetéseket is végezzünk. A szemantikus világháló-egondolás két alapötleten nyugszik: *egyrészt* a metaadatok erőforrásokhoz való kapcsolásán, *másrészt* azon az elgondoláson, hogy ezen metaadatok segítségével következtetni is kell tudni az adott dokumentum tartalmára. A metaadatok használata általános eszköz lehet arra, hogy a különböző erőforrásokhoz jelentést társítsunk, így egy-egy erőforrás jellemzőit meg tudjuk adni úgy, hogy az gépek (keresők) számára feltérképezhető legyen, óriási lépést tehetnénk az intelligensebb kereshetőség felé.

Mindezek után nézzünk egy alternatívát, amely megoldást jelenthet a jövőben. A felmerült problémák feloldását a természetes nyelvű kérdésfeltevés megengedése, az ún. *szabadszavas keresés* jelentheti. Nyilvánvaló, hogy a természetes nyelvű kérdések gépi „megértése” csak a nyelv szintaktikai szabályainak valamint szemantikai elemeinek bizonyos részét tartalmazó tudáskomponensek birtokában lehetséges. A mélyháló tartalmában való keresés esetén a sémainformációba kódolt szemantikus adatok már kiindulást jelenthetnek a megfelelő tudásbázis felépítésére. [5]

A *szabadszavas keresés* a szövegek keresése esetén a felhasználó számára lehetővé teszi az internetes adatbázisok tartalmában, a mélyhálón való szabadszavas, azaz magyar nyelvű, kerek egész kérdő mondatokkal történő keresést. Ez a keresési mód a felhasználó számára két jelentős előnnyel jár. *Egyrészt* lehetőséget ad a jó minőségű adatokat tartalmazó, hagyományos kereső-motorok segítségével nem elérhető tartalmak keresésére közös kiinduló pontból. *Másrészt* a szabadszavas kereséssel jelentősen egyszerűsödik az információigényt megfelelően reprezentáló keresőkifejezések megadása.

A megoldásként alkalmazható szabadszavas mélyháló kereső-program ismertetése előtt tisztázni kell mi is az a mélyháló. A portálok oldalai „mögött” található, strukturált, jellemzően adatbázisokban tárolt, dinamikusan elérhető tartalmak összességét mélyhálónak nevezzük (angol terminológia: deep web, DW). Az elnevezés a tartalom nehezebb elérhetőségére utal, szembeállítva azt a klasszikus, felszínen található tartalommal. A mélyháló nem azonos a láthatatlan vagy fekete hálóval. A láthatatlan háló részét képezik azok az oldalak is, amelyek tűzfal mögött, intraweben, jelszóval védett vagy más, általánosan meg nem közelíthető módon érhetőek el. A mélyháló jellemzője, hogy elvben bárki hozzáférhet ezekhez az információkhoz, de szisztematikusan, keresőmotorok általi bejárása nem volt lehetséges. Lehetővé válik tehát a gépek által készített, illetve előállított portálooldal rendszerének, a teljesen dinamikusan előálló oldalaknak a megtekintése is.

Nézzük meg röviden, hogyan működik ez a keresőrendszer. A felhasználó kérdése először a természetes nyelvi feldolgozó modulhoz kerül feldolgozásra, tehát ennek a modulnak a bemenete egyúttal az egész mélyháló kereső-szoftverrendszer általános bemenete is. A nyelvi feldolgozó modul a kérdés nyelvi feldolgozását, releváns lexikai egységekre való bontását (tokenizáció), valamint morfológiai és szintaktikai elemzését végzi. (Természetesen e feladatok elvégzéséhez különböző tudáskomponensekre és segédeszközökre van szükség.) Kimenete szintaktikailag elemzett, zárójelezett mondatalternatívák listája. A szintaktikailag elemzett mondatokat a kontextus felis-

merő modul átalakítja az általunk definiált kifejezésekké, amely már a kontextusra vonatkozó információkat is tartalmazza. Ez képezi a mélyháló-kereső motorjának a bemenetét, amelynek feladata többrétű. Egyrészt a kontextus-információk, az aktuális kérdés tárgya, a kérdésben szereplő tulajdonságok és a rendelkezésre álló adatbázis-leírók segítségével a kérdés megválaszolására alkalmasnak tartott adatbázisok meghatározása. Másrészt a bemenetére érkező formalizált mondatokból a megfelelő adatbázisok felé küldendő, az adatbázisra jellemző, de szabványos sémákra illeszkedő speciális SQL lekérdezések (DWL nyelvű lekérdezések) előállítás és ezek továbbítása. Ez utóbbi feladat magában foglalja az adott adatbázisra vonatkozó DB Info alapján a történő átalakításokat is.

A tartalomszolgáltatótól kapott választ (amely több találatot is tartalmazhat) továbbításra kerül a válaszfeldolgozó modulnak, amely összegyűjti az egyes adatbázisoktól beérkező eredményeket, és azokat különböző szempontok szerint rangsorolja, majd megjeleníti a felhasználó számára. A válaszok helyességének elemzése rendkívül időigényes feladat lenne, ezért a válaszokat az alábbiak szerint csoportosíthatjuk, illetve rangsorolhatjuk:

1. amennyiben az eredeti kérdés értelmezése nem egyértelmű, akkor az abból generált különböző kérdés-reprezentációk szerint;
2. adatforrás és azon belül séma szerint;
3. a válaszok beérkezési sorrendje szerint;
4. felhasználói profil szerint;
5. és végül a korábbi keresések felhasználói szokások alapján történő kiértékelése nyomán kialakult forrás--kontextus relevancia alapján.

A röviden ismertetett szabadszavas mélyháló kereső-alkalmazás nemzetközi viszonylatban is élenjáró technológiákat tartalmaz. Egyedülálló módon széleskörű természetes nyelvű feldolgozást valósít meg, aminek segítségével képes magyar nyelvű kérdőmondatokat SQL lekérdezésekké átalakítani, továbbá egy olyan komplex internetes keresőszolgáltatást javasol, amely három keresési technológiát integrál (felszíni, mélyhálós, illetve a nem részletezett vizuális tezaurusszal indexelt képi technológiát). Az integráltan kezelt feladatokra különállóan már léteznek technológiák. Az Interneten több mélyháló-kereső is található, amely angol nyelvű adatbázisokkal van kapcsolatban. Ilyen például a *BrightPlanet CompletePlanet* nevű keresője, amely a mélyhálós oldalak több mint felét indexeli, de ez csak kulcsszó alapú keresést támogat. Hasonlóan kulcsszó alapú keresőszolgáltatást nyújt még a *Pro-Fusion* és a *Copernic* is. Az integrált, képi keresést is támogató szolgáltatások egyre elterjedtebbé válnak, hiszen újabban a *Google* és *Yahoo!* is támogatja ezeket, míg korábban csak olyan kisebb kereső oldalak nyújtották, mint pl. az *iBoogie*. A szolgáltatások a képeket fájlnevek, illetve az esetleges egyéb képhez csatolt információk alapján indexelik, a képi tartalomban való keresést, annak bonyolultsága miatt egyik sem teszi lehetővé.

ÖSSZEFOGLALÁS, KÖVETKEZTETÉSEK

A cikk kiindulási alapját az „információs pajzs” létének beismerése jelentette. Az egész Földet behálózó lehallgatórendszer politikai, nemzetbiztonsági vizsgálata nem volt célom. Az alapvető cél az volt, hogy ráirányítsam a figyelmet arra, hogy az információgyűjtés már a jelenlegi módszerekkel, a karakterfigyelő rendszerek segítségével is milyen veszélyeket rejt magában. Azonban ha még jobban elterjednek a mélyháló keresésére is alkalmas keresőprogramok, teret nyer a szemantikus keresés, akkor annak nemzetbiztonsági vonatkozásainak vizsgálata szükségesszerűvé válik. (Természetesen nem állítom, hogy jelenleg nem folynak kutatások ebben az irányban a nemzetbiztonsági kérdésekkel foglalkozó informatikusok körében.) A fő kérdések a következők lehetnek:

1. Valóban együtt kell élnünk az ECHELON mindent behálózó „információs pajzsával”?
2. Az „információs pajzs” milyen módon viszonyul a civil társadalomhoz?
3. Lehetséges-e a védekezés a totális információs kiszolgáltatottság ellen?

Az ismertetett Interneten használható keresőrendszerek, illetve a felvázolt jövőképben jelzett megoldások mindegyike szolgálhatja az információs társadalom minden polgárának életét, de lehetőséget ad a visszaélésekre is.

Amennyire lehetséges volt elkerültem az informatikai kifejezéseket valamint a szigorúan szakmai megfogalmazásokat, hogy minnél nagyobb tömegek figyelmét ráirányíthassam a megfogalmazott problémára. Aki további információkat szeretne a témával kapcsolatban annak ajánlom a hivatkozott dokumentumokat, továbbá a keresést az *Interneten*.

Azért, hogy ne csak a negatívumok jelenjenek meg nézzünk egy példát arra, hogy milyen tévedések történtek a múltban, amelyek már az új keresőrendszerekkel elkerülhetővé válhatnak, és így akár pozitívumnak is tekinthetők felismerésük.

Mrs. Smith telefonon elmesélte barátnőjének, hogy kisfia bombát készített játszótársaival a játszótéren. Mrs. Smith sosem gondolta volna, hogy a gyermekjáték nem telefontéma. A brit elhárítás számítógépe (összekötve egy telefonvonalakat ellenőrző műholddal) — az ECHELON Dictionary alapján — felfigyelt a bomba szóra. Beindult az automatikus gépezet, és a beszélgetést rögzítette, majd szöveges formára konvertálta, ezután megjelenítette a GCHQ elemzőtisztjének képernyőjén. A tiszt mivel nem volt biztos abban, hogy a hölgyek nem terrorcselekményt készítenek-e elő virágnyelven beszélgetve, felvette Mrs. Smith nevét és telefonszámát a hírszerzés adatbázisába. Az asszony neve mellé ez a megjelölés került: LEHETSÉGES TERRORISTA.

Kulcsszavak: Internet, Informatika, Információs társadalom, Információ

Keywords: internet, informatics, information society, information

FELHASZNÁLT IRODALOM

- [1] ORWELL, G., A.: 1984. -Bp.: Európa Könyvkiadó, 2004. (eredeti cím : *Nineteen Eighty-Four*)
- [2] DÉNES Tamás: *ECHELON az e-társadalom információ-pajzsa?*
(forrás:http://www.titoktan.hu/index_b.php, DT Dénes tamás matematikus DT.htm, 2008.)
- [3] RÓNAI Tamás: *Digitális archiválás, adattömegek végleges megőrzésének problémái.* -in.:*Informatika a felsőoktatásban.*
-Debrecen, 2005.
- [4] SZEREDI Péter, LUKÁCSY Gergely, BENKŐ Tamás: *A szemantikus világháló elmélete és gyakorlata.*
-Bp.:Typotex Kiadó, 2005.
- [5] TIKK Domokos, KARDKOVÁCS Zsolt, MAGYAR Gábor: *A szavak hálójában: szabadszavas mélyháló kereső program.* (forrás: www.categorizer.tmit.bme.hu/~domi/publications/htechnika.pdf, 2005.)