

# TARTALOMKEZELÉS A MÉDIAKONVERGENCIÁBAN

Magyar Gábor

PhD, egyetemi docens  
BME Távközlési és Médiainformatikai Tanszék  
magyar@mail.bme.hu

## 1. Bevezetés

Az infokommunikációs hálózatok *online* hozzáférést biztosítanak a tartalomhoz. A tartalomkezelés dolga, hogy tároljuk, kezeljük, megtaláljuk azt, ami ebből bennünket érdekel. A tartalom továbbítása és kezelése napjainkban közös technológiai alapra kerül. Ez új lehetőségeket teremt, ami megszünteti a kétfajta tevékenység elkülönültségét.

A távközlés és az informatika technológiai konvergenciája ma már valóság, s kortársai vagyunk a médiatechnológia becsatlakozási folyamatának. Olyan digitális médiaformátumok váltak szabványossá, amelyek illeszkednek az infokommunikációs hálózatok műszaki jellemzőihez. Az átalakulásnak azonban csupán első felvonása, hogy a rádió-tévé műsorszórás egyeduralma megtört, és felzárkózik mellé az internet. A folyamat nem áll meg a tévéstudiók digitalizálásánál. Az új rádiós, tévés, internetes műszaki rendszerek nem egyszerűen kapcsolódnak a hálózathoz, nem egyszerűen kibocsátanak valamilyen médiatartalmat magukból, amit aztán a hálózat elszállít valahová. Sokkal mélyrehatóbb változásról van szó: a „tartalomgyárak” és a műsorszolgáltatók egy nagy elosztott rendszer részeivé válnak. Mindehhez a tartalomkezelési tech-

nológiáknak is be kell tagozódnuk a konvergenciába.

Az internet korában minden szervezet és az egyének is szembekerülnek azzal a szükséglettel, hogy adott feladat elvégzéséhez adatbázisból (például egy vállalati pénzügyi rendszerből), dokumentumokból (például egy iktatott jogi állásfoglalásból és régi elektronikus levelekből), valamint internetes forrásokból (például egy másik vállalat portálján feltehető árlistából) származó adatokat is fel kell használniuk. Az említett adatok már jó ideje elektronikusan keletkeznek (a régiek egy részét is ilyenné alakították), és informatikai eszközökkel rendezzük, tároljuk, továbbítjuk, keressük és használjuk fel őket. Az elkülönült vállalati, könyvtári és más adatkezelő rendszerek miatt az eltérő logikájú sokféleségből nehezen kinyert adatok csak az ember fejében állhatnak össze egységes tartalommal. A helyzet azonban megváltozik.

A felnövekvő generációk már nem veszik természetesnek, hogy különálló szövegszerkesztő, táblázatkezelő, naptári, kommunikációs és más alkalmazásokat használnak, legfeljebb állományszintű (export-import) adatkapcsolattal. Ehelyett egyetlen számítógépi környezetet igényelnek levelezéshez, azonnali üzenet-cseréhez („cseteléshez”), leckeíráshoz, tanulás-

hoz, szörföléshez, szórakozáshoz. Az egységes infokommunikációs környezet egyre átfogóbb, rövidesen ki fog terjedni a ma műsorszórás-tartalmaira is.

A felhasználási trend a mögöttes adat- és információkezelés megváltozásával függ össze, két értelemben is. Egyrészt, a multimédia-tartalom egységes szemléletet kíván és kínál: hétköznapi eset, hogy e-level, Word fájl, PPT prezentáció keverten tartalmaz szöveges, képi és hangelemeket. Előnyös, ha a multimédia-tartalomhoz egységes modellezéssel és technológiákkal közelítünk. Másrészt sokféle és gazdag szerkezeti kapcsolatrendszer alakul ki az adatelemek között. A tartalom gépi kezelésének egyik legizgalmasabb problémája a *tartalom – forma – szerkezet* összefüggésének figyelembe vétele. A tisztán szöveges tartalom digitális ábrázolása, tárolása, előkeresése viszonylag egyszerű: az írás mint információ-kódolás alapegységeivel (a mi írásunk esetében karakterekkel) végzünk műveleteket. Ez esetben is lényegi információt hordoznak azonban a formai elemek (pl. egy cím elkülönítése, kulcskifejezések kiemelése aláhúzással vagy eltérő színnel) és a szerkezeti összefüggések (triviális mai példa a hiperlink, továbbá idézzük fel itt a lexikon szócikkeinek egymásra utalását is, de a nyelvtan, sőt a szöveg mondanivalójában kifejezett utalások is szerkezetet hordoznak). Bonyolultabb formai és szerkezeti viszonyokkal találkozunk a multimédiatartalom (pl. egy hírportál) esetében.

A médiakonvergencia természetének megértéséhez először át kell tekintenünk a tartalomkezelés alapfogalmait.

## 2. Adat, információ, tudás, tartalom

Az *adat* és az *információ* szavakat gyakran szinonimaként használják, pedig jelentésük eltérő. Adatnak a valóság nem értelmezett

tükörképét nevezzük. Az adat nyers (feldolgozatlan) tény, ami valakinek vagy valaminek a jellemzéséhez hozzásegít. A felhasználónak azonban végső soron nem adatokra, hanem információra van szüksége. Az információ értelmezett adat. Mai tudásunk szerint csakis az értelemmel bíró ember képes adatból információt értelmezni. A számítógépi reprezentációkba a szemantikai összefüggéseket a programozó ember viszi be. Az adat formális közlési módjait szintaktikai konvenciók írják le.

A *tudás* az ember által megfelelő környezetbe, kontextusba helyezett információ (adott információt hol, mikor, mi célból használunk fel). A tudás felépülése szubjektív, egy ember vagy egy csoport háttértudása, előítéletei és érzelmei is szerepet játszanak abban.

A *tartalom* szót sokféleképp használjuk a hétköznapi életben, például a művészetekben és a jogban. Műszaki meghatározáshoz az információból nem indulhatunk ki, mert a médiában, azaz a kommunikációs közegben adatok vannak jelen, és nem információk, hiszen a közegben (ma) nincs feldolgozás vagy értelmezés. Ezért a mi definíciónk így szól: a tartalom ember általi érzékelésre egyidejűleg felkínált adatok összessége. A tartalomnak van (megjelenítési és tárolási) formája, valamint belső struktúrája is.

Az *adatkezelés* a nyers adatok tárolásának, lekérdezésének és manipulálásának kérdéseivel (például: adatmodellezés, adatbázisok) foglalkozik. Az *információmenedzsment* feladatköre az információk előállítása, rendszerezése, értékelése és visszakeresése. Az információvá értelmezésben elengedhetetlen szerepe van az adatszerkezet leírásának és felhasználásának. Az adatbázis-kezelőkben erre adatmodellt használunk, ami adatkapcsolatok leírására szolgáló formalizált jelölésrendszerből és az adatokon végrehajtható műveletek-

ből áll. Az internet térhódításával egyre több olyan adathoz férünk hozzá, amelyekhez nem készült adatmodell, holott az adatok között szerkezeti információk keletkeznek (például hiperlinkekkel, metaadatokkal). Adatmodell nélkül azonban a hasznos információ fellelése pontatlan és esetleges.

Az adatbázisban az adatok tárolása által meghatározott szerkezet jól illeszkedik az adatok információtartalma által meghatározott struktúrához (az adatok értelméhez, azaz szemantikájához). Amennyiben ez az illeszkedés hiányos vagy szabálytalan, sőt időben változó, akkor félstrukturált adatkezelésről beszélünk. Egyszerű példa erre egy weboldalon elhelyezett táblázat. Az információ kinyeréséhez itt először azonosítani kell az információ szerkezetét is, azaz meg kell találni a táblázatot, és azonosítani az oszlopok/sorok jelentését, hiszen ezt a tárolás struktúrája nem rögzíti. Az ember számára ez egyszerű, ám a gépnek nem az. Klasszikus mérnöki értelemben itt nincs adatmodell, az adatokon végzett műveletek halmaza nincs egyértelműen meghatározva. E kvázi-adatmodellek leírására komoly tudományos apparátus épült ki annak érdekében, hogy az interneten továbbra is mérnöki szakértelem nélkül szaporodhasson a tartalom, ám ennek szervezése és felhasználása jobb legyen.

### 3. A tartalomkezelő rendszer

A *tartalomkezelés* a tartalommal foglalkozó emberi munkafolyamatok gépi támogatását jelenti, ezért magában foglalja az adatkezelést és az információmenedzsmentet is. A tartalom előállításának, keresésének, feltárásának, megjelenítésének, elemzésének, szerkesztésének, terjesztésének és archiválásának legalább egy részét megvalósító műszaki alkotást *tartalomkezelő rendszernek* (CMS – Content Management System) nevezzük.

A CMS kifejezést kezdetben az internetes portálokra használták, azaz a félstrukturált adatok világára. Csakhogy a ma trendje éppen az elkülönült rendszerek egységesítése, a felhasználó információ igényének „egyablakos kiszolgálása”. Olyan tartalomkezelő rendszerek jönnek létre, amelyekben strukturált, félstrukturált (és strukturálatlan) adatokat egyaránt kezelni leszünk képesek, mégpedig közös technológiával. A kifejezetten webes tartalmakat kezelő rendszer a WCMS (Web Content Management System), a vállalati tartalomkezelő az ECMS (Enterprise Content Management System). Elterjedt, kevésbé specifikus tartalmú rövidítés a DAM (Digital Asset Management – digitális vagyongazdálkodás) rendszer.

A tartalomkezelő rendszer három fő funkcionális részre osztható: *gyűjtő, tároló és publikációs* alrendszerekre.

A *gyűjtő alrendszer* feladata a tartalom (szöveg, kép, hang, mozgókép s ezek kombinációi) előállítása és/vagy bejuttatása a rendszerbe. Szerzői eszközök segítik a tartalom előállítását, más adatforrásból formai és szerkezeti ellenőrzésre képes interfészen keresztül vehető át adat. A gyűjtő alrendszer végezheti el a szükséges formátum- és szerkezetátalakításokat, csoportosításokat, válogatásokat is. Igen fontos feladata a tartalom komponensekbe (tartalomegységekbe) szervezése (például tévéműsor mozgóképfolyamának élet-szerű egységekre – filmek, hírek, reklámok stb. – darabolása). Ebben az alrendszerben a tartalmat leíró adatokkal láthatjuk el. A tartalom jellemzésére ma a legfőbb eszközünk: a *metaadat*. Metaadatként rögzítjük egy tartalomrész címét, alkotóját, keletkezésének időpontját, formátumát, elérhetőségét és sok minden mást. A metaadat egy másik adatot jellemez, „adat az adatról”. A metaadatok

önmaguk is adatok, róluk is lehetnek további metaadatok. A metaadatot az eredeti (még nem jellemzett) tartalomhoz csatoljuk valamilyen módon (például a digitális tartalom állományába kódoljuk, vagy mutatók, linkek teremtenek kapcsolatot a tartalom és a leíró adatok között). Metaadatot ír a könyvtáros, amikor katalogizálja az állományba vett könyvet. Metaadatot rögzít a digitális fényképezőgép, amikor a képhez elmenti a keletkezés időpontját, a kép felbontását, kódolási módját, a zoom állását stb. Metaadatot ír az ember, amikor a fényképezőgépből a számítógépbe másolt képhez szavakat csatol: „Ági 2000-ben”. A metaadatok származtatása ma nagyon élőmunka-igényes, ezért kutatásokat végzünk ennek automatizálására. A metaadatok leírására döntően jelölő nyelveket (például: HTML, XML) használunk.

A *tároló alrendszerben* történik a tartalomkomponensek (valamint metaadatok) hosszú távú tárolása és az erőforrások kezelése. Ide csatlakozik a CMS adminisztráció, a belső munkafolyamatok ütemezése, vezérlése és követése. E munkafolyamatok műszakilag is igen bonyolultak lehetnek. A megőrzés ugyanis összetett feladat. A komponenseket oly módon kell (a hozzájuk – esetleg sokdimenziós relációkban – rendelt leíró adatokkal együtt) „eltenni”, hogy több szempontból (például: méret, hozzáférési idő, költség) optimalizálható legyen, arra is tekintettel, hogy a megőrzés értelme végső soron az, hogy a tartalom később hozzáférhető, fellelhető legyen. Ennek műszaki környezete a mai hálózatos, esetleg elosztott erőforrású világban önmagában is komplex. A tároló alrendszer műszaki alapja akár egy *peer-to-peer* állománycserélő hálózat is lehet. A hosszú távú megőrzésnek a digitális világban több problémája is lehet: a fizikai romlás, az elavulás (példá-

ul: régi formátumok lejátszása), a megbízhatóság és a biztonság. Elvi kérdés is felmerül: míg egy régi könyv esetében világos, hogy mi számít eredeti példánynak, egy digitális dokumentum esetében ez nem egyértelmű.

A *megjelenítő alrendszer* a tartalomkomponensekből publikációkat, a felhasználók számára szóló közléseket készít. Egyszerű esetben ez a tárolt komponensek kijátszása, gyakran azonban formátum átalakítása, átcsomagolása/átszerkesztése is szükséges. A publikáció különböző közlési csatornákra készülhet (például: nyomtatott könyv, újság, online portál, hírlevél, tévéműsor). A korszerű tartalomkezelő rendszerek többplatformos kijátszásra képesek. Egyazon CMS-ből lehetséges nyomtatott, sugárzott és online közlést is megvalósítani, ezen belül is többféle felhasználói készülékre (például: médiatartalmat tévére, PC-re, PDA-ra, mobiltelefonra).

A metaadatok, a belőlük épített struktúrák és főleg ezek egyezményessé válása (szabványosítása) ma a tartalomintegráció kulcsa. A kereskedelemben is kapható sokféle tartalomkezelő termékek legjobbjai gazdag metaadat-kezelést kínálnak. Említtettük már, hogy a metaadatok többsége emberi közreműködéssel keletkezik. Virágzó módja ennek az online közösségek működése. A „blogterekben” a szerzők megadják az írásukra vonatkozó leíró adatokat, hiszen ez elősegíti láthatóságukat, közlésük értelmét szolgálja. A wiki közösségek (lásd Wikipedia) alkotói is megadnak metaadatokat és szerkezeti adatokat. E sajátos online közösségek működési közeget is CMS rendszerek valósítják meg.

#### 4. A szemantikus adatok beépítése

A tartalomhoz rendelt metaadat lehet *leíró* (például: formátum, a keletkezés időpontja), vagy *szemantikus* (ami a dokumentum jelenté-

sére vonatkozik; például egy cikk kulcsszavai). Ha kellően sok ilyen metaadattal látjuk el az információs rendszerünkben elérhető tartalmakat, akkor olyan rutinfeladatokat is gépesíthetünk, amelyekhez az emberi intelligencia nélkülözhetetlen. Például sokkal pontosabban és gyorsabban kielégíthetjük információs igényünket. Azt is mondhatjuk, hogy a gazdagon összekapcsolt szemantikus adatok rendszere az emberi tudás egyfajta reprezentációja. Ez a tudásreprezentáció nyilvánvalóan szűkebb, mint az emberi tudás maga (hiszen kizárólag a már használatba vett metaadatokkal írja le a világ dolgait). E szemantikus veszteséggel szemben nagy előnye azonban, hogy nemcsak emberi, hanem gépi feldolgozásra is alkalmas.

Ezt a megközelítést egy közösség, egy vállalat tartalomkezelő rendszeréről az egész internetre kiterjesztve a szemantikus világháló elképzeléséhez jutunk. Ez egy réteges felépítésű tudáskezelési koncepció. Minden adatot, amit az interneten kezelni akarunk, egyedi azonosítóval kell ellátni. Ez az ún. URI (Unified Resource Identifier, például egy internetcím). Az adatszerkezet leírására XML nyelvet használunk. Utaltunk már arra, hogy egy adathalmaz értelmezéséhez az adatok közötti viszonyokat ismernünk kell. Ennek modellezésére a szemantikus hálóban elsősorban egy erőforrás-leíró keretrendszert (Resource Description Framework – RDF) alkalmazunk.<sup>1</sup> Az RDF egy absztrakt adatmodellezési technika, amely bármilyen típusú metaadat leírására alkalmas bármely (egyedi azonosítóval rendelkező) internetes erőforrásról. Az így rendszerezett adatok jelentésének gépi feldolgozhatóságához szükséges még a jelentéssel bíró fogalmak közötti össze-

<sup>1</sup> Vannak más technikák is, például a tématerkép (Topic Map).

függések precíz leírása is. A gépi feldolgozás számára rögzítjük a fogalmakat, a fogalmak közti kapcsolatokat és az adott témakör alapigazságait kifejező axiómákat. Ezzel ún. *ontológiát* készítünk, ami adott felhasználói csoport által egy adott témakörben közösen használt világkép formális specifikációja. A formális tudásleírás elemein akár gépi következtető, logikai algoritmusokat is végrehajthatunk. Ez a formálisan leírt tudást immár gépi úton gyarapítja. Ezt a bonyolult konstrukciót azért fejlesztjük, hogy előre feldolgozott információt nyújtson számunkra, minél kevesebb emberi munkával. A szemantikus háló a webet egy önleíró adatbázissá formálja.

Akár az emberi tartalomkezelésnél, itt is gondot kell fordítani a megbízhatóságra és a hitelességre. A feldolgozást végző szoftverügynökök számára ezért külön rétegekben kell kapaszkodókat nyújtani annak megítéléséhez, hogy mely adatok minősége bizonytalan, esetleg téves vagy félrevezető.

### 5. *A médiakonvergencia hatása az infokommunikációra*

A médiakonvergencia első fázisa már valóság: az új médiatartalom-kezelő rendszerek olyan médiafolyamokat (stream) és állományokat (file) állítanak elő, amelyek illeszkednek a – vezetékes és vezeték nélküli – infokommunikációs hálózatokhoz.<sup>2</sup> Ily módon lineáris műsorközvetítés és kívánságalapú (on-demand) tartalomszolgáltatás egyaránt biztosítható. Vannak szabványok metaadatelemeire és azok szerkezetét előíró sémák számára is. Ezek nélkül nem lehetséges sem az elektronikus műsorújság, sem az *on-demand* katalógus.

<sup>2</sup> Ilyen alapokon működik a digitális tv, a kábel-tv, az IPTV, a filmek és a zeneszámok letöltése, részben azok lemezen való terjesztése is.

Ez azonban csak a kezdet. A hálózat számára a médiatartalom nagy adatmennyiségével, sávszélességigényével és forgalmi jellemzőivel is új kihívást jelent. A tartalom egyre számosabb ponton keletkezik a tartalomgyáráktól a háztartásokig (a felhasználói tartalomig – user generated content, például: interneten közzétett saját készítésű filmek, mint születésnap videó, művészkedés; mások tartalmához való „képi hozzászólás”, tovább szerkesztés stb.). A terjesztés a műsorszórástól az egyenrangú pontok (peer-to-peer) hálózatáig többféle modellben történhet.

A médiakonvergencia idején is természetesen gondoljuk megkülönböztetni az *előállítás–továbbítás–felhasználás* munkafolyamat lépéseit. Lássunk egy korszerű példát! Egy videó állománycserélő tartalmait (videoklipeket) a felhasználó feltölti-letölti: a tartalom előállítása és felhasználása egyértelműen elhatárolható. Csakhogy: a klipek címkézésébe (gazdagabb információvá értelmezésébe, azaz: megtalálhatóbbá, jobban értékesíthetőbbé tételébe) a közönség is bekapcsolódhat (a saját videoklipjein kívül másokéba is!). Az eredmény: az előállítás és a felhasználás összekeveredik. Ugyanez tapasztalható az ún. közösségi portálokon (például iwiw) és a blogterekben is.

Úgy tűnik, hogy az előállítás–továbbítás–felhasználás lépéseiből csakis a továbbítás (a klasszikus *távközlési* funkció) marad élesen elkülönült. A médiatartalom terjesztő hálózatok azonban elosztott tartalomtároló és -kezelő hálózatokká fejlődnek. Ma még egyszerűbb centrális architektúrában gondolkodni,<sup>3</sup> amikor azonban majd a végpontok száma nagyon nagy lesz, akkor egy centrális architektúra hatalmas „túlmeretezést” igényelne.

<sup>3</sup> A centrális modellben egy központi tartalomforrás (médiaszerver) bocsátja a tartalmat a terjesztőhálózatba.

Miért? Egyrészt, az ún. *multicast*<sup>4</sup> elosztás akkor lesz gazdaságos, ha a népszerű tartalmakat több példányban tárolva a fogyasztói berendezésekhez közelebbről juttatjuk ki, másrészt, a teljes forgalom nem elhanyagolható része lesz felhasználói tartalom. Ez az előzetesen nem méretezhető forgalom gazdaságosabban kezelhetőnek tűnik az erősen elosztott architektúrában. Jegyezzük itt meg: az elosztott architektúra egészen a fogyasztó *set-top-box*áig<sup>5</sup> terjedhet, mert lesz abban tároló- és feldolgozókapacitás is. Miért ne lehetne része az a hálózatos erőforrás-kezelésnek? Kevésbé forgalmas órákban például bizonyos tartalmakat – a várhatóan legkurrensebb filmeket, a legfrissebb híreket, lehetőleg a felhasználó érdeklődési profiljának megfelelően kiválogatva a szolgáltató letöltheti a set-top-box tárolójába, így csökkentve a nagy forgalmú órák terhelését, ezzel a gerinchálózat csúscapacitását, tehát a szükséges beruházás nagyságát. A fenti gondolatmenetben a set-top-box helyére bármilyen (fix vagy mobil) előfizetői berendezést behelyettesíthetünk.

Ebben az architektúrában a tartalmat előállító, feldolgozó és használó rendszerek nem csupán fizikai, hálózati és alkalmazási szinten, hanem tartalom szinten is együtt fognak működni. A tartalomfelhasználó (a tévé szerkesztője vagy egy blogger) ugyanis ebben gondolkodik: műsorszámokban, dalokban, klipekben, dokumentumokban. A tartalomgyárák állandó kapcsolatban fognak állni más produkciós forrásokkal, továbbá az

<sup>4</sup> Többesadás: amikor a forgalmi hálózat csomópontja a jelet a hálózat több kimeneti útjára továbbítja.

<sup>5</sup> Szűk értelmezése: eszköz, adapter, amely az analóg műsorjelek vételére alkalmas tévékészülékeket alkalmassá teszi digitális műsorjelek vételére is. Tágabb értelemben: eszköz, amely a bármely eltérő formátumú, illetve kódolású műsorjeleket alkalmassá teszi tévékészülékben való felhasználásra.

archívumokkal. Együttműködésüket (a kívánt tartalmak megtalálását és a megfelelő formátumra alakítását) metakeresők, tematikus és más alapokon szerveződő szolgáltatók fogják segíteni.

Lehetséges ez? Bemutattuk, hogy a tartalomkezelés a tartomelemek milyen sokréttű kapcsolatrendszerét hozza létre, beleértve még azok jelentéstani rétegét is. Vajon miért ne használnák ki az így létrejövő lehetőségeket nem csupán a médiatartalom feldolgozásában, de terjesztésében is. Mindez ma még talán kétkedő mosolyt csal a mérnökök arcá-

ra, túl bonyolultnak, menedzselhetetlennek tűnik. A szerző szerint e vízióban a különösen érdekes kérdés az, hogy vajon valóban összefüggő és átjárható infrastruktúrájú infokommunikációs hálózat alakul-e ki? Ha igen, akkor egyáltalán nem irreális feltételeznünk, hogy az infrastruktúra nem pusztán a jelfolyamok továbbítására, hanem tárolására, átalakítására, sőt *tartalom szerinti* terjedésére (útválasztására és követésére) is kiterjedhet.

Kulcsszavak: *tartalomkezelés, adat, információ, tudás, média, szemantikus háló, wiki, blog*

## IRODALOM

- Baeza-Yates, Ricardo - Riberio-Neto, Berthier (1999): *Modern Information Retrieval*. Addison-Wesley
- Bergeron, Bryan (2003): *Essentials of Knowledge Management*. John Wiley & Sons, Inc.
- Berners-Lee, Tim - Hendler, J. - Lassila, O. (2001): The Semantic Web. *Scientific American*. 17 May
- Boiko, Bob (2001): *Content Management Bible*. Hungry Minds, New York
- Fraser, Stephen (2002): *Building a Content Management System*. Apress, Springer Verlag, New York..
- Guarino, Nicola - Masolo, C. - Vetere, G. (1999): OntoSeek. Content-Based Access to the Web. *IEEE Intelligent Systems*. 14, 3, 70–80.
- Magyar Gábor – Szakadát István (2001): *Metadata System of National Audiovisual Archive in Hungary*. Invited Paper. 20<sup>th</sup> Conference of the Audio Engineering Society: Archiving: Restoration and New Methods of Recording. Budapest, 5-7 October 2001. CD-ROM: Mira Digital Publishing Inc.

- Magyar Gábor (2005): *Key Technical Problems of Long-term Digital Archiving of Documents*. Invited Paper. 3<sup>rd</sup> Conference of Documentation & Electronic Archiving, Dubai, 18-20. Sept. 2005.
- Magyar Gábor – Knapp Gábor (2004): Building an Open Document Management System with Components for Trust. In: Linger, Henry – Fisher Julie (eds.): *Constructing the Infrastructure for the Knowledge Economy*. Kluwer Academic–Plenum Publishers, New York, USA, 430–442.
- Magyar Gábor – Hutter O. – Mlinarits J. (2005): *e-Learning*. Műszaki, Budapest
- Magyar Gábor (2006): *A média konvergenciája*. Előadás a 6. HTE Kongresszuson (Áttörés az infokommunikációban) Balatonkenese, 2006. október 5.
- Sherman, Price (2001): *The Invisible Web (Uncovering Information Sources Search Engines Can't See)*. CyberAge Books,
- Sowa, John F. (2000): *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. Brooks Cole Publishing Co.