

A DIGITÁLIS DOKUMENTUMTÁROLÁS KIALAKULÁSA, JELENE ÉS JÖVŐJE

Napjainkban mind nagyobb szerepet kap szöveges információk digitális formában való tárolása és terjesztése, legyen szó akár egy oktatási segédanyagról, egy irodalmi műről, vagy egy műszaki dokumentációról. Az elektronikus dokumentumtárolás számos előnnyel bír a hagyományos papíralapú információkezeléshez képest, gondoljunk például a szövegben való keresés lehetőségére, vagy a környezetvédelmi szempontokra: „*Ön elektronikus könyvet vásárolt: a fák köszönik*” – olvasható a Mekka elektronikus könyvtárház weblapján.

Az előnyök mellett azonban számtalan probléma vetődött fel a digitális dokumentumtárolás területén: Hogyan tehetjük operációsrendszer függetlenné a tárolt információkat? A használt formátumok garantálják-e a jó minőségű nyomtatás lehetőségét? Megoldható-e a konverzió a már létező formátumok között? Hogyan valószínűsíthető meg a szerzői jog védelme? Valóban az elektronikus könyvek a jövő? Hogyan juthatunk hozzá az a digitális formátumban tárolt könyvekhez? Milyen eszközök állnak rendelkezésünkre az elektronikusan tárolt információik asztali számítógépektől független olvasására (kézi számítógépekkel való kompatibilitás problémája, e-book reader eszközök terjedése, e-paper). Melyek azok a formátumok (.txt, .rtf, .doc, .html, .xml, .pdf, .lit, ...) és szoftverek, amelyek megoldást jelenthetnek a fenti kérdésekre?

Az elektronikus dokumentumtárolás első mérföldkövének tekinthetjük a Michael Hart által elindított Gutenberg Project-et. Hart 1971-ben határozta el, hogy a világ-irodalom alkotásait megpróbálja elektronikus formában hozzáférhetővé tenni az Internetet használók számára. Az ötlet onnan származott, hogy valamilyen okból szüksége lett volna a Függetlenségi Nyilatkozat szövegére, amit más megoldás nem lévén ő maga gépelt be. Mivel meglehetősen hosszadalmas munka volt, úgy döntött, hogy elküldi a kész szöveget minden ismerősének, hátha ők is hasznát veszik az elektronikus formában tárolt változatnak. Az e-mailben terjedő szöveg nagy sikert aratott, rohamosan terjedt és a visszajelzések ösztönözték Hart-ot a „Project Gutenberg” beindítására. A szövegek tárolására a .txt formátumot választotta, hiszen az ASCII (American Standard Code for Information Interchange) formátum mindenki számára olvasható volt, függetlenül az általa használt géptől és operációs rendszertől. Napjainkban már több ezerre tehető az ebben az „elektronikus szöveggyűjteményben” szereplő dokumentumok száma. (<http://promo.net/pg>)

Az ASCII formátumban tárolt szövegek ugyan többé-kevésbé teljesítik a platformfüggetlenség elvét, de nem alkalmasak például a szöveg megfelelő formázására és a jó minőségű nyomtatásra. Erre a problémára próbáltak a szoftvergyártó cégek reagálni a személyi számítógépekre kifejlesztett szövegszerkesztő programok meg-

jelenésével. Az elmúlt 20 évben sok hasonló szövegszerkesztő alkalmazás született, de mindközül „legnépszerűbb” a Microsoft Word. Széleskörű elterjedése alapján az elektronikus dokumentumtárolás de facto szabványának tekinthető a Microsoft .doc formátuma. A .doc fájlok széleskörű formázási és jó minőségű nyomtatási lehetőséget biztosítanak, azonban komoly hiányosságokat mutatnak a hordozhatóság terén. A különböző verziók egymással sem teljesen kompatibilisek, és egészen a közelmúltig kizárólag a Windows-os platformon voltak használhatók. Mára már lehetővé vált a formátum más operációs rendszerek alatt történő felhasználása is (MS Office for Macintosh, StarOffice, OpenOffice.org, ...), de továbbra is problémát jelent a dokumentumok „formátummegőrzése”.

Elsősorban a hordozhatóság megkönnyítésére fejlesztette ki a Microsoft az RTF (Rich Text Format) formátumot. Ez a fájlformátum képes a különböző operációs rendszerek, különböző szövegszerkesztő programjai közötti adatcserére, mind szöveges, mind grafikus információk tárolására. Legfrissebb verziója az RTF 1.6, amelynek leírása a Microsoft weboldalán található meg.

<http://msdn.microsoft.com/library/en-us/dnrftspec/html/rftspec.asp>

Sajnos sem a .doc, sem az .rtf formátum nem nyújt megbízható megoldást az eszközfüggetlen nyomtatásra, ugyanis nyomtatásra szánt, lezárt dokumentum egy másik gépen valószínűleg kissé áttördelve jelenik majd meg.

A nyomtatásra szánt információk hordozhatóságára a PostScript formátum jelentett megoldást. A PostScript az Adobe cég szabványos platformfüggetlen formátuma. Olyan lapleíró nyelv, amely elsősorban nyomtatóra szánt információk tárolására szolgál, azonban bizonyos segédprogramokkal (pl: Ghostscript, Ghostview) képernyőn való megjelenítésre is alkalmas.

A PostScript kifejlesztésében jelentős szerepe volt John Warnocknak, aki, miután elhagyta a Xerox céget, 1982-ben megalapította az Adobe-t. Ebben az időszakban kísérletezett az Apple az első lézernyomtatójának kifejlesztésével, és Steve Jobs felkérte az Adobe-t egy nyomtató-vezérlőnyelv kifejlesztésére. 1985-ben piacra került az első PostScript lézernyomtató Apple Laserwriter néven.

A PostScript kifejlesztésének tapasztalatira építve alkotta meg az Adobe cég a PDF (Portable Document Format) formátumot. A tervezésben a következő szempontokat vették figyelembe:

- legyen hordozható (amint a neve is mutatja),
- egyaránt legyen alkalmas monitoron való megjelenítésre és jó minőségű nyomtatásra,
- kezelje a hyperhivatkozásokat,
- on-line, egy böngésző segítségével is olvasható legyen a megfelelő plug-in segítségével,
- rendelkezzen különböző dokumentumvédelmi szintekkel.

A .pdf állományokat tetszőleges forrásállományból elő lehet állítani az Acrobat PDFWriter (vagy Distiller) segítségével, amely mint nyomtató jelenik meg a rendszerben. Az elkészült dokumentumok az ingyenes Acrobat Reader szoftver segítségével olvashatók, nyomtathatók. A Reader tartalmazza azt a plug-int, amely a böngészők ablakában is képes megjeleníteni a .pdf fájlokat.

Mára ez a formátum vált a műszaki dokumentációk és a nyomtatásra szánt anyagok terjesztésének elsődleges szabványává. Népszerűségének legfőbb oka talán a dokumentumok védelmének biztosítása. Az elkészült dokumentumban köthetjük jelszóhoz a megnyitást, az esetleges módosításokat, tilthatjuk a nyomtatást, a szövegrészek kimásolását, az űrlapokhoz való hozzáférést.

A .pdf a felsorolt tulajdonságoknak köszönhetően az egyik legalkalmasabb formátum az elektronikus könyvek számára is, éppen ezért az Adobe mind több kézi számítógépre (Acrobat Reader for Symbian OS, Acrobat Reader for Palm OS, Acrobat Reader for Pocket PC) elkészíti az Acrobat Reader speciális változatát, amely alkalmassá teszi ezeket az eszközöket az e-könyvek asztali számítógéptől független olvasására.

Az e-book piac sikerét látva a Microsoft is kifejlesztette a saját formátumát (.lit) és hozzá az olvasó szoftvert (Microsoft Reader). Az elektronikus könyvpiacnak két fő vetélytársa a formátumokat tekintve tehát az Adobe és a Microsoft.

Az elektronikus dokumentumtárolás történetében egy új fejezetet nyitott a World Wide Web Consortium (W3C) által 1998-ban bejelentett XML (Extensible Markup Language) szabvány.

Az XML a jelölőnyelvek családjába tartozik, akárcsak az SGML és a HTML, egyre több alkalmazás (szövegszerkesztők, böngészők) támogatja és várhatóan kiemelkedő szerepet játszik majd az információtárolásban.