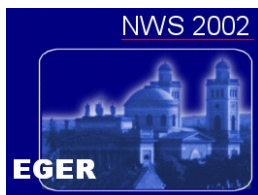


Networkshop, 2002



Dokumentumok és prezentációk környezetfüggetlensége és hordozhatósága

Holl András (1), Kadlecsik József (2), Telbisz Ferenc (2)

1 - MTA Konkoly Thege Miklós Csillagászati Kutatóintézete

2 - MTA KFKI RMKI SZHK

Felvetés

Mindennapi munkánk során gyakran kerülünk szembe számítógépes dokumentumaink cseréjének problémáival. Különösen figyelemre méltónak ítéljük a hivatalos kommunikáció kérdését. Két egyenrangú partner megállapodhat bármiféle szoftver, formátum használatában, a hatóságoknak, hivataloknak, intézményeknek azonban óvatosabban kell eljárniuk, ha ügyfeleikkel elektronikus dokumentumokat cserélnek. Az inkompatibilitás lehetősége növekszik országhatárokon átnyúló dokumentumcsere esetén, és a dokumentumok újrafelhasználásának esélyei csökkennek, ha évek/évtizedek múltak el a keletkezésük óta.

Előadásunkban a következő mindennapi dokumentumtípusokra térünk ki: e-mail üzenetek, szöveges dokumentumok, weboldalak, táblázatok illetve táblázatkezelő programok állományai, prezentációk és formanyomtatványok.

A csereszabatosság kérdését azzal az előfeltevéssel kell kezelni, hogy a felhasznált számítógépi hardverek, operációs rendszerek és alkalmazások sokfélék lehetnek. Manapság a PC-k, a Microsoft Windows operációs rendszerei, az MS Office csomag, valamint néhány további alkalmazás széleskörű elterjedtsége miatt gyakran tekintik egyes felhasználói programok "belső" dokumentumformátumait "de facto" szabványnak. Találkozhatunk azzal a törekvéssel, hogy más hardver- illetve operációs rendszer-környezetben is próbálják megteremteni e formátumok kezelhetőségét - egyelőre korlátozott sikerrel. A kérdés nehézsége nem kis részben e formátumok inherens problémáiban, illetve abban áll, hogy gyorsan változnak, s az alkalmazások (különböző verziói) önmagukkal sem tökéletesen kompatibilisek.

Helyzetfelmérés

A különböző számítógépes platformok - architektúrák és operációs rendszerek elterjedtségéről csupán meglehetősen pontatlan becslések állnak rendelkezésre. Érdemes lenne felmérést végezni a HUNGARNET közösségen belül! A szerzők szűk körben (az MTA KFKI-n és az MTA KTM CsKI-n belül) elindítottak egy cenzust.

Milyen a kívánatos formátum?

- A célnak megfelelő formátum

Célszerű először azt vizsgálnunk, mi az információátvitel célja? A szöveges

dokumentum szolgálhat rövid, gyors, informális tájékoztatásra. Lehet hivatalos levélről, dokumentumról szó - ez esetben célszerű olyan formátumot választani, amit szerkeszteni nem lehet, csupán a képernyőn megjeleníteni, vagy kinyomtatni. Ez esetben az is lényeges, hogy a dokumentum pontosan a feladó szándékainak megfelelően jelenjen meg (tördelés, karakterkészlet, esetleg tipográfia)! Végül előfordul, hogy a küldött szöveget a fogadó félnek szerkesztenie kell. Mindegyik eset különböző formátum használatát kívánja meg.

Azokban az esetekben, amikor manapság sokszor táblázatkezelő formátumot használnak, meg kellene vizsgálni, vajon egy program elküldésére van-e szükség (ami táblázatszerű adatokon operál), vagy táblázatos adatokat kívánunk összegyűjteni és összesíteni? Az utóbbi esetben célszerűbb egyszerű, szöveges formátumú táblázatot használni, mint spreadsheet-et.

Pályázati űrlapok esetén sem célszerű a spreadsheet-ek használata. Kötelezhető-e a pályázó egy bizonyos szoftver alkalmazására? Vajon ki tudja-e oldal-helyesen nyomtatni? Módosíthatja-e az űrlapot, akár csak egy szűknek talált "ablak" tágításával? Talán jobban megfelelne a célnak, ha az "űrlap" a pályázató szervezet web-szerverén elérhető HTML-forma lenne, amit a kiszolgáló gép dolgoz fel, esetleg szintaktikus ellenőrzés után, majd a felhasználónak letöltésre felkínál néhány nyomtatási formátumot? (A "beküldött" anyag rögtön digitális formában is rendelkezésre áll.)

Általában célszerűnek látjuk, ha a hivatalos kommunikációban a "hivatal" a szerveren kezeli a problémákat, a kliens-oldalon könnyű és univerzális program szükségeltetik csak - mint egy tetszőleges böngésző vagy e-mail kliens.

- Általános kívánalmak

Milyen lenne egy ideális formátum? Egyszerűbb néhány kívánatos tulajdonságot felsorolni, mint ezekből egy ideális formátumot összeállítani. Véleményünk szerint a jó formátum nyílt, közösségi erőfeszítéssel fejlődik ("bazár-modell"), valamiféle non-profit szervezet viseli gondját. Szabványos, stabil, visszafelé kompatibilis. Előny az egyszerűség: például szöveges (nem bináris) forma. Viszont rendelkeznie kell a célnak megfelelő funkcionalitással, elegendően "tetszetős" is kell legyen, s mindemellett elterjedtnek is kell lennie.

- Specifikus kívánalmak

Rövid, informális kommunikáció esetén érdemes csatolt állományok nélküli elektronikus levelet küldeni. Itt leginkább az ékezetes karakterek megfelelő reprezentációja fontos.

Hivatalos dokumentumoknál a PDF-et találjuk ajánlhatónak: lehet nem módosíthatóvá tenni, kompakt, ingyenes szoftverrel lehet olvasni/nyomtatni, sőt, előállítani is.

Szerkesztendő dokumentumoknál az adott helyzetnek megfelelően alkalmazható egyszerű szöveg, RTF, talán HTML is. Lehet, hogy majdan az XML lesz ebben az esetben is a legjobb formátum? Mindenképpen előny, ha a formátum megengedi akár egyszerű editorok használatát is. A .doc formátum kényelmes lehet, ha a cserepartner biztosan tudja olvasni, és bizonyosan ismeri az alkalmazott funkciókat. Hivatalos vagy archiválendő dokumentumoknál azonban használata nem ajánlható.

Néhány létező formátum

- .doc : nem választható külön a Word -től. Elterjedt, ám több különböző, nem teljesen kompatibilis verziója van. Többet tud, mint amire a legtöbbször

szükség van, mégsem annyit, hogy igazán bonyolult, igényes feladatokra, mint a kiadványszerkesztés, jól alkalmazni lehessen. Biztonsági szempontból aggályos. Lokális használatra jól alkalmazható, csereképessége, újrafelhasználhatósága korlátozott. Újabb fejlemény, hogy több, különböző platformon elérhető szabad szoftverrel is meg lehet kísérelni a kezelését.

- RTF : dokumentumcserére inkább alkalmazható.
- "sima" szöveg
- SGML, HTML, XML : az SGML-alapú formátumok a legígéretesebbek. Nem binárisak, szabványosak. A HTML korlátozott funkcionalitású, ám elterjedt. Az XML a jövő ígérete.
- TeX, LaTeX : tulajdonképpen nagyszerű formátumok, ám az ezeket kezelő programok csupán szűk körben használatosak, nem felhasználóbarátok. A fizikai-matematikai tudományokban viszont sok szakfolyóirat ezt a formátumot kéri.
- PostScript, PDF : Két kiváló dokumentumformátum. A PostScript elterjedt, sok szabad program írja/olvassa, ám kissé terjedős. A PDF modernebb, szintén kezelhető ingyenes programokkal, kompaktabb, sokat tud. Mindazonáltal óriási flexibilitásuk magában rejti az inkompatibilitás veszélyét. Korántsem biztos, hogy ezeket a dokumentumtípusokat mindig, mindenütt ki lehet nyomtatni!
- JPEG : szabad, elterjedt, kép formában megőrzendő dokumentumok tárolására is alkalmas.

Milyen az ideális szoftver?

Nem annyira a program a fontos, mint az alkalmazott formátum. Az adatstruktúrát kell definiálni, szoftvert ki-ki használjon ízlése, igényei, pénztárcája szerint. Örvendetes, ha egy jó szoftver több platformon is elérhető, s amennyiben pénzbe kerül, létezik korlátozottabb funkcionalitású, ingyenes változata is. Ugyancsak vonzó az alkalmazásról leválasztott felhasználói felület, "skin" gondolata.

Néhány létező szoftver

- MS Office
- StarOffice / OpenOffice / MagyarOffice
- AbiWord
- Netscape / Mozilla
- ghostscript / gsvie
- AcrobatReader
- TeX / LaTeX / LyX
- Irfanview / XnView
- Pine / Pegasus Mail / Eudora

Megjegyezzük, hogy a szabad szoftverek esetében néhol biztató a fejlődés. A LyX például könnyen kezelhetővé teheti a TeX/LaTeX formátumot. Általános elterjedésükhöz azonban még sok problémát meg kell oldani.

A részletek ördöge

A nehézségek gyakorta a részletekben bújnak meg. Az inkompatibilitási gondok egy része helytelen beállításokból adódik - maga a program képes lenne csereképesebb állományokat készíteni, csak az alapbeállítások nem megfelelőek, vagy a felhasználó nincs tisztában választásainak következményeivel. Talán a legtöbb nehézséget a rossz font-választás, illetve a megfelelő font-készletek hiánya okozza.

Praktikus tanácsok

Az olvasó figyelmébe ajánlunk néhány praktikus tanácsot.

- Általában nem célszerű link-eket használni. A mai dokumentumszerkesztők táblázatkezelők felkínálják a lehetőséget, hogy más dokumentumokat, ábrákat ne építsünk be, csak link-vel hivatkozzunk rájuk. Sajnos, amennyiben megváltozik a környezet (nincs hálózat, vagy csupán a könyvtár-struktúra más) ezek a dokumentumok teljesen használhatatlanná válhatnak.
- Minden dokumentumtípusnál célszerű az egyszerűbb és közismert lehetőségeket alkalmazni, annyit, amivel a kívánt végső formát jelezni lehet.

Tennivalók

Véleményünk szerint kívánatos lenne, ha a különböző szervezetek (HUNGARNET, MTA, minisztériumok, egyetemek) útmutatókat, ajánlásokat állítanának össze arról, különböző célokra milyen formátumokat célszerű alkalmazni, milyen programok illetve programverziók használatával, milyen beállításokkal lehet elősegíteni a dokumentumok cseréjét. Jó lenne, ha ezek az ajánlások legalább a tudományos/oktatási/közgyűjteményi körben egységesek lennének! Az ajánlások kidolgozására munkacsoportot/levelezőlistát kellene felállítani.

Szükség lenne magyar nyelvű dokumentációkra, szabad programok magyarítására, fontkészletekre: sokféle ingyenesen elérhető ISO Latin-2 fontot kellene készíteni.

Záró gondolatok

Meg vagyunk győződve arról, hogy a programokat és a formátumokat célszerű elválasztani egymástól. Az alkalmazások lehetnek bármilyenek, de az adatformátumoknak egyszerűeknek, szabványosoknak, "könnyűeknek" kell lenniük. Sok esetben előnyös lehet a "nehéz" központi szerverek és a "könnyű" univerzális kliensek használata, ám a protokollok és formátumok kérdése az, amire a leginkább figyelni kell.

URL-ek

Holl A., Kadlecsik J., Telbisz F.: Dokumentumok és prezentációk környezetfüggetlensége és hordozhatósága (előadás szöveges anyag)

<http://www.konkoly.hu/staff/holl/eger/eger.html>

Office File Format Migration

<http://www.microsoft.com/office/techinfo/enterprisezone/itcolumn02.htm>

RTF

http://www.biblioscape.com/rtf15_spec.htm

<http://support.microsoft.com/support/kb/articles/Q86/9/99.ASP>

SGML

<http://www.oasis-open.org/cover/general.html>

XML

<http://www.w3.org/TR/REC-xml>

HTML

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

TeX/LaTeX
(TeX Users Group)

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

(Comprehensive TeX Archive Network)

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

PostScript
(Encapsulated PostScript File)

http://partners.adobe.com/asn/developer/pdfs/tn/5002.EPSF_Spec.pdf

PDF

<http://partners.adobe.com/asn/developer/acrosdk/DOCS/pdfspec.pdf>

JPEG

<http://www.w3.org/Graphics/JPEG/jfif.txt>

Irfanview

<http://www.ryansimmons.com/users/irfanview/>

XnView

<http://perso.wanadoo.fr/pierre.g/index.html>

GsView, Ghostscript

<http://www.cs.wisc.edu/~ghost/>

Mozilla

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

Pine

<http://www.washington.edu/pine/>

Pegasus Mail

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

Eudora

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

AbiWord

<http://www.abisource.com>

MagyarOffice

<http://www.MagyarOffice.hu>

Dokumentumok és prezentációk környezetfüggetlensége és hordozhatósága

Holl András <holl@konkoly.hu>

MTA CsKI

Kadlecsik József <kadlec@sunserv.kfki.hu>

MTA KFKI RMKI SzHK

Telbisz Ferenc <Telbisz@sunserv.kfki.hu>

MTA KFKI RMKI SzHK

Ma már elengedhetetlen szükséglet a számítógéppel létrehozott dokumentumaink cseréje (másokkal vagy különböző rendszerek között), újrafelhasználhatóságuk és archiválhatóságuk nemkülönben. Rövid üzeneteket (e-mail), képeket, kinyomtatandó/megtekintendő, máskor szerkesztendő szövegeket küldünk, web oldalakon adunk közre információt, vagy ily módon kívánunk információhoz jutni, táblázatokat cserélünk, esetleg a rajtuk operáló programokkal együtt, prezentációinkat kívánjuk idegen környezetben bemutatni, elektronikus formanyomtatványokat kell kitöltenünk és kitöltetnünk. Nem szabadna előfordulnia, hogy egy dokumentumot valaki ne tudjon fogadni, más számítógép, más program, más környezet akadálya legyen ennek. Ugyancsak fontos szempont, hogy az imént felsorolt dokumentumokat évek, évtizedek múlva is elővehessük, esetleg teljesen más számítástechnikai környezetben újrafelhasználhassuk.

Manapság a Microsoft Windows operációs rendszerének, valamint az MS Office csomag változatai széles körben elterjedtek, némelyik dokumentumtípus esetén „de facto” szabványt képeznek, míg más típusokat több elterjedten használt rendszer segítségével is kezelhetünk. A fentebb felsorolt kívánalmak korántsem valósulnak meg teljesen, vagy akár csak elfogadható mértékben. A számítógépes kommunikációnak sok problémája van még – például az inkompatibilitás, nem megfelelő funkcionalitás. Előadásunkban ezt a helyzetet kívánjuk megvizsgálni, mind elvi, mind gyakorlati síkon.

Megpróbálunk bemutatni néhány, a számítógépes dokumentumok cseréjében szóbaeső formátumokkal szemben támasztható igényt, és megnézzük, a létező szabványok mennyiben felelnek meg ezeknek; a gyakorlatban választható formátumoknak milyen előnyei és hátrányai vannak. Felvázoljuk az ideális dokumentumkezelő szoftverek néhány vonását, és megvizsgáljuk, mire képes a jelenleg rendelkezésre álló szoftverek némelyike. Végül azt a következtetést vonjuk le, hogy a kívánalmaknak leginkább megfelelő szoftverek és formátumok használata esetén is céltudatosan kell megválasztani, mikor melyiket használjuk, valamint a szoftverek beállításait és a formátumok nyújtotta lehetőségeket is óvatosan kell alkalmazni.

Elektronikus adatgyűjtés a KSH-ban

Koltai Tamás <tamas.koltai@office.ksh.hu>

Központi Statisztikai Hivatal

. <eds862@mailop.ksh.hu>

Központi Statisztikai Hivatal

A KSH a MATÁV együttműködésével egy nem régen zárult kísérleti projekt keretében megindította Internetes adatgyűjtési rendszerét. A kísérletben 3 megye 172 cége vett részt, és ezek