

WEB-BÁNYÁSZATI MÓDSZEREK ALKALMAZÁSA INTERNET-ALAPÚ TANANYAGOK MINŐSÉGVIZSGÁLATÁRA

(A „13. MultiMédia az oktatásban” konferencián elhangzott „Web-bányászati módszerek alkalmazása Internet-alapú tananyagok minőségvizsgálatára” c. plenáris előadás bővített és részben átdolgozott változata)

*Izsó Lajos, izsolajos@erg.bme.hu
BME, GTK Alkalmazott Pedagógia és Pszichológia Intézet*

1. Bevezető

Az elmúlt néhány évben az internet-alapú tananyagok – a gyakran tapasztalható ellentmondások és átmeneti visszaesések ellenére – jelentős mértékben terjedtek. Az oktatási webhelyek gyakran csupán közvetítik a különböző formátumú (doc, ppt, pdf, etc.), alapvetően sztatikus tananyagok hozzáférését, letöltését. Újabban azonban már határozottan nő azon webhelyek száma is, amelyek valódi interaktív tananyagfeldolgozási lehetőségeket biztosítanak (pl. anyagfeldolgozás, példamegoldás, tantárgytesztek kitöltése interaktív visszajelzéssel, stb.). Az internet-alapú tananyagok erőteljes terjedését kísérő gyakori negatív tapasztalat ugyanakkor az, hogy ezzel a mennyiségi növekedéssel az anyagok (szakmai, didaktikai, használhatósági/szoftverergonómiai) minősége nem tart lépést. Ennek egyenes következménye az a növekvő igény, hogy ezen tananyagok minőségét elméleti és gyakorlati szempontból egyaránt megalapozott módon tudjuk megítélni, jellemezni. Számos próbálkozást láthatunk ezért újabban ilyen tananyagok minőségének a – lehetőség szerinti objektív - mérésére (a legtöbb ilyen megközelítés különböző kérdőívekre, skálákra, interjúkra vagy strukturált megfigyelésekre épül). Mindezek mellett megállapítható, hogy a probléma még távolról sincs megoldva.

Ez az írás - a két másik egyidejűleg megjelent azonos témájú, konkrét empirikus eredményekről beszámoló cikkel egységet alkotva – bevezetésül kíván szolgálni a web-bányászati módszerek oktatási célú alkalmazásának problematikájába.

2. Web-bányászat: az üzleti szféra Internet-alapú szegmensében kidolgozott erőteljes speciális adatbányászati megközelítés

Web-bányászat: az üzleti szféra Internet-alapú szegmensében kidolgozott igen hatékony speciális adatbányászati megközelítés, lényegében az adatbányászat [1] [2] [3] [4] [8] [9] alkalmazása a webhelyek forgalom-elemzésének speciális területére. A módszer a webhelyek üzemeltetőjének a szerverén képződő, a forgalomra jellemző naplófájlok (az ún. logfájlok) kifinomult elemzésére épül és az adott webhelyek hatékonyságának a fokozását tűzi célul.

A különböző típusú naplófájlok (ún. weblog-ok) elvben a felhasználók tevékenységére vonatkozó igen gazdag információ-források [5][6][7][9]. Itt a probléma a szokásosnak éppen a fordítottja: itt nem kevés, hanem túlságosan is sok információ áll rendelkezésünkre. A naplófájlok mérete ugyanis az üzleti szférában – a látogatók számától és elemzett időszak hosszától függően - gyakran a több száz MB vagy a GB nagyságrendjébe esik. Ezt a nagy mennyiségű - és nagy mértékben pusztán technikai jellegű – információt kell célszerű módon megszűrni, átalakítani, feldolgozni úgy, hogy a kezdeti technikai információból a felhasználó viselkedését és motivációját megalapozottan jellemző információkat nyerjünk.

A gyakorlati web-bányászati célokra korábbi tapasztalataink alapján a *Clementine WebMining* eszközöket javasoljuk alkalmazni [10][11][12].

A web-bányászatnak a *Clementine*-ban követett ún. „felhasználó-középpontú” filozófiája ugyanis teljes mértékben megfelel a korszerű marketing, ergonómia és pedagógia felfogásának. Ez az újszerű felfogás – szemben a hagyományos technikai ihletésű „lap-középpontú” megközelítéssel – a felhasználó céljait, szándékait helyezi az előtérbe és ennek alárendelve tervezi meg a rendszer szolgáltatásait, amint azt már volt alkalmunk tapasztalni, igen sikeresen.

Példaként következzen néhány hasznosnak bizonyult elemzési megközelítés, amelyeket kifinomult – gyakran különböző tanulóalgoritmusokat is tartalmazó – célra irányult szoftver modulok (a *Clementine* esetében ún. *stream*-ek) – támogatnak.

- A látogatások és a látogatók szegmentációja: *Visit and User Segmentation (E-ChannelUser RFM Classifications, User Mode Determination, Visit Branding)*.

- A látogatók viselkedésének elemzése: *Web Site Activity and User Behavior (Visit Activity Variances, Identifying Undesirable Behavior, Lifetime Conversion Tracking, Points of Abandonment, User Activity Focus, Visit Activity Funnels, Navigational Usage)*.
- A látogatók leggyakoribb aktivitási sorozatainak azonosítása: *Activity Sequence Analysis (Most Common Activity Sequences, Eventstream Visualisation)*.
- A látogatók vásárlási, döntési hajlamának meghatározása tanuló algoritmusok segítségével: *Propensity Analysis*.
- A látogatók magasabb szintű szegmentációja: *Advanced User Segmentation*.
- Online hirdetési tevékenység hatásvizsgálata: *Targeting Online Promotional Activity, illetve Campaign Performance Measurement*.

Ezen *stream*-ek egy része haszonnal adaptálható az internet-alapú tananyagok és könyvtári anyagok világára.

3. A web-bányászat alkalmazása internet-alapú tananyagok használatának minőségvizsgálati céllal történő vizsgálatára

Az oktatás világa önmagában nem lett volna képes a web-bányászati eszközök kifejlesztéséhez szükséges erőforrások biztosítására. Miután azonban ezeket az eszközöket a sokkal fizetőképesebb üzleti szféra igényeinek a kielégítésére korábban már kifejlesztették, adott annak a lehetősége, hogy ezeket az oktatás világára adaptáljuk. Mindkét jelenlegi világpiacvezető adatbányász programcsomag (az *SPSS Clementine* és a *SAS Enterprise Miner*) rendelkezik kiegészítő web-bányász eszközökkel.

A BME Alkalmazott Pedagógia és Pszichológia Intézetében (APPI) - a Budapesti Műszaki Főiskola Tanárképző és Mérnökpedagógiai Központja aktív részvételével – létrejött egy kutatócsoport, amely a Clementine eszközökkel megvalósítható oktatási célú gyakorlati web-bányászati alkalmazások lehetőségeit vizsgálja. A kutatócsoportnak jelenleg kb. 14-16 aktív tagja van, akik zömmel a BME APPI és a BMF TMPK munkatársai mellett más pedagógusképző társintézmények (BDF, EKF, SZE) szakemberei, illetve diplomázói és PhD hallgatói közül kerülnek ki.

Az interakció professzionális vizsgálata lehetővé teszi a tananyag tényleges használati módjainak az azonosítását, amiből megalapozottan következtetni lehet a tananyag hatékonyságára, illetve a hatékonyság konkrét akadályaira. Ezek az azonosított konkrét akadályok megfelelő áttervezéssel megszüntethetők.

A web-bányászati technikák ezért a tananyag fejlesztésének a folyamatában – mintegy az ipari gyártásközi minőségellenőrzés analógiájára - az egzakt közbenső „mérések” lehetőségét nyújtják: a fejlesztés minden „mintavételezés” után attól függő irányokat vehet, hogy a tanulók a megelőző lépésben hogyan fogadták a tananyag számukra rendelkezésre bocsátott verzióját. A menet közben azonosított szakmai, didaktikai, használhatósági, szoftver-ergonómiai (leggyakrabban a navigációval vagy az információmegjelenítéssel kapcsolatos) stb. problémákra ilyen módon gyors áttervezési, javítási javaslat adható és az így kialakított verzió egy újabb ciklusban tesztelhető.

Ez a megközelítés annyiban radikálisan más a korábbiakhoz képest, hogy itt a tanulók és a tananyag interakciója vonatkozásában nem egyfajta mintavételezésről van szó, hanem valamennyi tanuló valamennyi interakcióját elemezhetjük a teljes tananyagban a billentyűleütések és egér-kattintások szintjének megfelelő finomságú felbontásban. Az

elemzés tehát nem a célpopulációból vett többé-kevésbé reprezentatív mintákra épül, hanem a célpopulációra nézve teljes körű.

A sikeres web-bányászati elemzés feltétele, hogy a különböző modellekkel nyert eredményeket az elemző képes legyen a tanulói/felhasználói szintű élményekkel, illetve tapasztalatokkal összekapcsolni. Ez azt jelenti, hogy csak az lehet valóban sikeres, aki mind a web-bányászati elveket és eszközöket, mind az elemzett konkrét tananyagot/tartalmat, mind pedig a tananyagot használó tanulókat eléggé mélyen ismeri.

Végül, a kétségtelenül jelentős előnyök felsorolása után szólni kell ennek a megközelítésnek két jelenleg még fennálló akadályáról. Az egyik az adatbányász, illetve web-bányász programcsomagok – üzleti szférához szabott - igen magas ára. A másik pedig az a tény, hogy a rendszer használatának elsajátítása komoly szellemi erőfeszítéseket igényel. Hosszabb távon azonban mindkét akadály vonatkozásában optimisták vagyunk. Egyfelől kifejezetten oktatási – nem profitszerző - célokra általában lehetséges a web-bányász termékek forgalmazóitól jelentős kedvezményeket, illetve engedményeket kapni, másfelől az informatikai és matematikai kultúra terjedésével – ami napjaink információs társadalmában ha lassan is, de kétségtelenül végbemegy - a web-bányászat elvi alapjainak és gyakorlati technikáinak az elsajátítása az újabb tanár nemzedékek számára már egyre könnyebb lesz.

4. Összefoglalás

Az Internet-alapú tananyagok – bár ellentmondásosan és időnkénti visszaesésekkel - erőteljesen terjednek. Ez az erőteljes terjedés gyakran azzal jár, hogy az intenzív mennyiségi növekedéssel az anyagok (szakmai, didaktikai, használhatósági/szoftver-ergonómiai) minősége nem tart lépést. Növekvő igény tapasztalható ezért ezen tananyagok megfelelő minőségére. Ennek következményeként számos hagyományosnak tekinthető - különböző kérdőívekre, skálákra, interjúkra épülő - próbálkozást láthatunk Internet-alapú tananyagok minőségének mérésére. Elismerjük ezen hagyományos megközelítések értékeit és szükségességét, úgy véljük azonban, hogy az Internet-alapú interaktív tananyagok minőségét ezeknél radikálisan jobban meg lehet ragadni alkalmas web-bányászati technikákkal nyert adatok segítségével. A BME Alkalmazott Pedagógia és Pszichológia Intézete - a Budapesti Műszaki Főiskola Tanárképző és Mérnökpedagógiai Központja aktív részvételével – egy kutatócsoportot hozott létre azzal a céllal, hogy az ún. „felhasználó-középpontú” filozófiára

épülő Clementine adatbányász programcsomag és annak WebMining CAT modulja segítségével – a tanuló és a tananyag interakciója során képződő naplófájlok kifinomult elemzése útján - kiválasztott konkrét Internet-alapú interaktív tananyagok minőségét az eddigiéknél megalapozottabb megközelítésekkel, módszerekkel jellemezze.

Irodalomjegyzék:

- [1] Abonyi János (2006). (szerk.) *Adatbányászat a hatékonyság eszköze*. COMPUTERBOOKS, Budapest.
- [2] Adrians, Pieter; Zantige, Dolf (2002). *Adatbányászat*. Panem, Budapest.
- [3] Bodon Ferenc (2007). *Adatbányászati algoritmusok*. (Szabadon letölthető és felhasználható anyag)
<http://www.cs.bme.hu/~bodon/magyar/adatbanyaszat>
- [4] Fajszi Bulcsú, Cser László (2004). *Üzleti tudás az adatok mélyén*, BME GTK, Budapest.
- [5] Izsó Lajos, Hercegfői Károly (2004). Website-ok minőségbiztosításának ígéretes új módszerei: automatikus site-értékelő eljárások. Kutatási Jelentés 27-28. 2004. október. 10-26 old. Kiadja az INFONIA Alapítvány, az INFINIT Műhely és a BME Információs Társadalom- és Trendkutató Központ szakmai támogatásával.
- [6] Izsó Lajos (2007). Web-bányászati módszerek alkalmazása internet-alapú tananyagok minőségvizsgálatára. Plenáris előadás. 13. „MultiMédia az oktatásban” konferencia, Budapesti Műszaki Főiskola. 2007. augusztus 23-24.
- [7] Izsó Lajos, Hercegfői Károly (2003). A látogatók által bejárt útvonalak rögzítésén alapuló elemző módszerek az Alkalmazott Pszichológia Alapítvány honlapjának minőségbiztosításában. *Alkalmazott Pszichológia*, V. évf. 1. sz. 79-89.
- [8] JIWAEL, Han - KAMBER, Micheline: *Adatbányászat*. Panem, Budapest, 2004.
- [9] Ketskeméty László, Izsó Lajos (2005). *Bevezetés az SPSS programrendszerbe. Módszertani útmutató és feladatgyűjtemény statisztikai elemzésekhez*. 459 oldal. Eötvös Kiadó. ISBN 963 463 823 6.
- [10] SPSS Inc. (2000). *Clementine® 10.1. In-Database Mining Guide*.
- [11] SPSS Inc. (2005). *Web Mining for Clementine® 1.5. Application Template*.
- [12] SPSS Inc. (2005). *Web Mining for Clementine® 1.5. User's Guide powered by NetGenesis*.