

Bóta László

Eszterházy Károly Főiskola, Médiainformatika Intézet

botal@ektf.hu

OKTATÁSI ADATBÁNYÁSZAT

Adatbányászat az oktatásban

Az intézményi szintű döntéshozatal és a stratégiai irányítás a felső vezetés, illetve a középvezetők felső szintjének feladatkörébe tartozik, amelyek kiemelt támogatását az üzleti intelligencia (Business Intelligence, BI) rendszerek jelentik (Krauth, 2008, 549.). A BI elemző folyamatában az adatbányászat egy jelentős, erőteljes eszköz. Az adatbányászat (data mining) üzleti megközelítéssel egy tudásfeltáró folyamat, a kimenetként kapott újszerű, hasznos és érthető minták feltárásával a menedzsment egy szubjektivitástól mentes, viszonylag gyorsan és költséghatékonyan előállított információhoz juthat (Bodon 2010, 10.). Az adatbányászat a vezetők szemszögéből egy döntéstámogató folyamat, amely előzőleg nem ismert információt tár fel nagy adathalmazból (Abonyi, 2006,10.).

Az adatbányászatot az 1980-as évek végén az üzleti élet és a marketing igényei alakították önálló tudományterületté, és napjainkban is ugyanezen területek kiemelték. Az adatbányászat jelentősége az 1990-es években kezdett erősödni. Az adatbányászat fő irányzatai: távközlés, orvostudomány, valamint a gazdasági alkalmazások (Han, 2004, 447.). Az adatbányászati eszközök oktatási területen történő alkalmazása az oktatási adatbányászat, angol elnevezéssel Educational Data Mining (EDM).

Az EDM az oktatásban működő informatikai rendszerek által használt, illetve ezen rendszerek használata során keletkező digitális adatok bányászatával foglalkozik. Az oktatási rendszerek működése tartalmi és szervezési oldalról is vizsgálhatók. A tartalmi oldal vizsgálata esetén a tananyagok használhatósága (Izsó-Tóth, Nyéki), illetve azok tartalma lehet az elemzés fókuszában. Az első aspektus többnyire a webbányászat, míg a második a szövegbányászat módszereit használja, ezek az adatbányászat részterületei. A szervezési oldal vizsgálatánál elsősorban egy webes alapú Learning Management System (LMS) által rögzített adatok webbányászati elemzése gyakori. Az oktatás szervezéséhez azonban számos más informatikai rendszert is használnak az intézmények, végső soron menedzsment szemszögéből közvetve vagy közvetlenül az összes gazdasági rendszer a hallgatókat szolgálja ki (Balogh, 2010, 7).

A felsőoktatási intézmények működése szempontjából az idej és jövő év központi intézkedései meghatározóak. A hallgatói létszám csökkentését célzó törekvések miatt az intézmények még komolyabb versengést fognak folytatni a hallgatók megszerzéséért és megtartásáért. Alapvető (marketing) cél a leendő hallgatók hiteles tájékoztatása és toborzása mellett a jelenlegi hallgatók elégedettségének növelése, a hallgatói vagy egyéb fogyasztói élettartam értékének maximalizálása, a szolgáltatásminőség javítása.

A szűkülő központi erőforrások miatt a marketingkutatás kiterjesztése versenyelőnyt biztosíthat az intézmények számára. A marketingkutatás, – Kotler megfogalmazása szerint – egy konkrét marketinghelyzethez kötődő adatok és megállapítások szisztematikus tervezése, gyűjtése, elemzése és jelentése (Kotler, 2006, 158), ami napjainkban nem az

informatikai rendszerek integrációjára épül. Az integráció a kezdeti szakaszában tart a felsőoktatási intézményeknél, ami még inkább igaz az informatikai rendszer részét jelentő marketing információs rendszerre (MIR). Magyarországon a felsőoktatási intézmények folyamatait mindössze 2007 óta vizsgálják adatbányászati eszközökkel, míg a hazai tőkeerős profitorientált cégeknél ez évekkel korábban elkezdődött.

Az adatbányászat bemutatása után az egri Eszterházy Károly Főiskola informatikai rendszereire alapozva a harmadik fejezet vázolja az adatbányászatra alapozott marketingkommunikációs problémák döntéstámogató lehetőségeit, majd egy példa leírása következik az adatbányászat alkalmazhatóságára.

Az adatbányászat mint kutatási eszköz

Az adatbányászat kutatói oldalról iteratív folyamat, amelynek során intelligens műveleteket, műveletsort végrehajtása történik az adatminták kiemelése érdekében. Az intelligens műveletek különféle statisztikai alapú elemző technikákat jelentenek, pl. neurális hálózat, faktoranalízis, klaszteranalízis (Bodon, 2010, 7.).

Számos adatbányász programcsomag létezik, jelen kutatást támogató alkalmazás az IBM-SPSS Modeler 14.1, valamint egy, a program korábbi verziójával megjelent kiegészítés, a Web Mining for Clementine 1.5 Application Template (CAT), ami a webbányászatot támogatja. A termékek jelenlegi árfekvésük miatt nem érhetőek el az oktatási intézmények számára. A vizsgálatokhoz alkalmazott adatbányászati szoftvert az SPSS Hungary a Budapesti Műszaki és Gazdaságtudományi Egyetem nonprofit kutatásai számára térítésmentesen biztosította.

Az adatbányászati modellek segítségével két alapvető feladatot oldhatunk meg: feltárás és előrejelzés. A feltárás során az adatok általános jellemzői, azaz a minták keresése történik, amelyek trendeket, csoportokat és azok jellemzőit, a marketing területéhez kapcsolódva sokszor az ügyfelek szokásait írják le. Az előrejelzés esetén az ismert értékek és a feltárt tudás alapján következtet egy jövőbeni eseményre. A két alapvető feladatnál a megszerzett tudással szemben négy elvárást fogalmazhatunk meg: legyen könnyen érthető, érvényes, hasznos és újszerű (Bodon, 2010,12-15.).

A webbányászat alapfogalmai

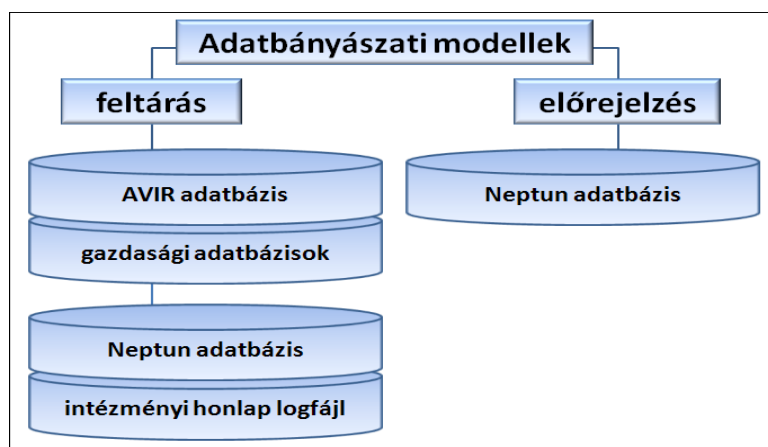
Az adatbányászat egyik részterülete a webbányászat (web mining), amihez a leíró osztályba tartozó adatbányászati feladatok tartoznak (Han 2004, 433.). A bemutatott példában a webhasználat-bányászat részterület segítségével tanulmányozzuk a fogyasztók szokásait. A webhasználat-bányászat nevezhető webnapló bányászatnak is, hiszen a webkiszolgálók (webservice) által rögzített webnapló-bejegyzéseken (weblog) alapul.

A leírásban több aspektusból vizsgálja a webet használó embert: marketing oldalról fogyasztó, webergonómiai és webbányászati szempontból látogató vagy felhasználó. Az előbb említett fogalmak a vizsgálatnál egymás szinonimáiként értendők. Az esemény (event) a felhasználó (user) egy konkrét kérését (request) (pl. weboldal, dokumentum, kép letöltése) jelenti, amely az adott online látogatás során megvalósult. A felhasználó (user) jelen esetben azonosíthatatlan, vagyis csak az IP címe (Host név mező) alapján követhetjük nyomon a tevékenységét, azaz nem kellett bejelentkeznie a honlapra felhasználói névvel és jelszóval, valamint nem kapott a számítógépére egy ún. cookie-t, így azzal

sem azonosítható. A látogatás (visit) egy felhasználóhoz (user) tartozó, időben korlátozott kérések véges sorozata. A látogatás befejeződik, ha a felhasználó egy lekérése után nincs 30 percnél belüli lekérés, 30 perc után már egy új látogatás kezdődik.

Az oktatási adatbányászat lehetséges irányai egy konkrét intézményben

A felsőoktatási intézményekben, így az Eszterházy Károly Főiskolán is az integrált informatikai rendszerek kiépítése az utóbbi években kezdődött el, az egymástól eltérő egyedi rendszerek hosszú időn át sem a döntések támogatását, sem a központi adatszolgáltatást nem támogatták. Mindez az AVIR programrendszer kapcsán már a valóságban is létező, országos lehetőség. Az elmúlt években országosan bevezetett AVIR programcsomag a felsőoktatás döntéstámogató programrendszere kíván lenni, de nem rendelkezik adatbányászati képességekkel.



1. ábra: Az adatbányászati modellekkel megoldható feladatok az Eszterházy Károly Főiskolán használt programrendszerek összefüggésében. (Forrás: saját ábra.)

A vizsgált főiskola szoftvereihez kapcsolódó adatbányászati lehetőségek állandóan változnak, de a kiemelt kutatási irányok megjelölhetők (1. ábra). Az első kutatási irányt az AVIR mutatószámok alapjául szolgáló, speciálisan tárolt adatok egyéb lekérdezései, elemzései jelentik. Jelenleg az adatok feltöltése történik, a rendszer még nem szolgáltat adatokat a vezetés számára. A gazdasági adatbázisok jelentik a kutatás másik fő területét. Az adatbányászati kutatásánál módszertan és az adathozzáférés jelenthet nehézséget. A harmadik fő kutatási irány a Neptun tanulmányi nyilvántartó rendszer adatai alapján tűzhető ki. A Neptun adataiból előre lehetne jelezni az adatbányászati modellekkel azon hallgatókat, akik jó eséllyel el fogják hagyni az intézményt, és számukra egy egyedi kommunikációt, szükség esetén szolgáltatást kínálhat a főiskola. A negyedik fő kutatási terület a honlap látogatók tevékenységét rögzítő állomány adatain alapul, ehhez kapcsolódik a következő fejezetben bemutatott kutatás. A honlaphoz köthető vizsgálatok adatbányászati részterülete a webbányászat.

Kutatási példa az oktatási adatbányászat alkalmazására

A leendő hallgatónál az online marketingkommunikáció egyre nagyobb súlyát bizonyítja, hogy a felsőoktatási intézmény kiválasztásához a diákok közel 80%-a az internet web szolgáltatását veszi igénybe, míg a hagyományos felvételi tájékoztató ebben a rangsorban csak a harmadik helyen áll (Dinya 2010, 65.). Ehhez társul az aktív hallgatók napi szintű internethasználata. Az online kommunikáció az esetek többségében igen rövid, legfeljebb pár perc terjedelmű egy szervezet honlapján. A cél az, hogy a lehető legkevesebb mentális erőfeszítéssel ériék el a látogatók a keresett információt. A fejezetben bemutatott kutatási példa egy korábbi publikáció vázlatára (Bóta, 2011).

A kutatás célja

A kutatás célja azonos az üzleti céllal. Az intézményi honlap főmenüinek kiválasztása alapján legyenek elkülöníthetők az azonos főmenüket választók felhasználói csoportjai. A kutatás eredménye a menedzsment marketingkommunikációs döntéseinek előkészítését támogatja, pontos célja az intézményről kialakult kép és attitűd javítása, a főiskola kiemelt PR eszközén, a honlapján keresztül. A kutatás eszköze az adatbányászat klaszterező modellje. Az eredményeket a webergonómiai szempontokkal összevetve ad javaslatot a honlap navigációjának javítására. A netgenerációhoz tartozó fogyasztó akkor lesz lojálisabb az intézményhez, ha a főiskola honlapja mentális terhelés nélkül szolgáltatja az általa keresett információkat.

Az alkalmazott adatbányászati módszertan és modell

A marketingkutatás lépései adottak, azok ebben az esetben is követendők: kutatási cél, terv, információgyűjtés, információ elemzése, eredmények bemutatása, döntéshozatal (Kotler 2006, 159.). Az alkalmazott adatbányászati módszertan a CRISP-DM (CRoss Industry Standard Process for Data Mining). A CRISP-DM hat lépése (Chapman et al., 2000) jelenik meg a következő alfejezetekben.

1. lépés. Az üzleti cél definiálása és értelmezése és a hipotézis

Az üzleti cél: a honlap főmenüinek kiválasztása alapján legyenek elkülöníthetők a hasonló menüket választó látogatók csoportjai. Ha a klaszterek előállnak, akkor webergonómiai szempontból áttekinthető, hogy a csoportokat leíró menüpontok elhelyezése segíti vagy gátolja az adott csoport globális navigációját. A globális navigáció a honlap eltérő oldalain állandóan elérhető menüpontok halmazát jelenti. A fentiek alapján a kutatás hipotézise: a statikus weboldalakból álló honlap esetén az azonosíthatatlan, online látogatók kiszolgálásának hatékonysága növelhető az adatbányászati klaszterező modellek eredményeinek felhasználásával.

2. lépés. A rendelkezésre álló adatok megismerése

Az Eszterházy Károly Főiskola történetiségében előző honlapját 2007-ben használták utoljára. A jelenlegi hivatalos honlap (www.ektf.hu) 2007. október 9-től már az itt bemu-

tatott honlaptól eltérő. A honlapot látogatók adatainak, viselkedésének tárolása egy napló-fájlban (weblogfile) történik. A vizsgálatban érintett, digitálisan tárolt webnapló adatok a korábbi honlap esetén 2007. január 7-től, a váltás napjáig állnak rendelkezésre.

Az adatbányász szoftver lehetőséget ad a statisztikai adatok megjelenítésére is. A szoftver 2007. január 7-e és március 18-ig tartó adatokat volt képes egyszerre feldolgozni. A honlapot a vizsgált időszakban 67 837 felhasználó (user) látogatta, amelyből a főiskola területén mindössze 88 látogató volt, akik összesen 183 283 látogatást (visit) valósítottak meg. Az időszak leggyakrabban választott menüpontja (2. ábra) a „Felvételizőknek” 919 744 választással (hit), a második helyen a „Szervezeti egységek” 103 882, míg a harmadik helyen, a bal oldalon lévő „Oktatás” menü 28 854 választással.

3. lépés. Az adatok előkészítése

A kiinduló adatok az említett napló-fájl bejegyzéseinek halmazát jelentik. Az eredeti állomány alkalmatlanok volt a vizsgálat lefolytatására, ezért azokat elő kellett készíteni.

4. lépés. Modellalkotás és az eredmény

Megjelenésétől kezdve az adatbányászat leggyakrabban alkalmazott területe a klaszterezés. A felhasználás célja sokrétű, például weboldalak, gének, betegségek, ügyfelek csoportosítása. A klaszterezésre azért van szükség, mert az ügyfelek számossága miatt a kézi kategorizálás túl nagy költséget jelentene. Marketing szempontból nem az a fontos, hogy az egyes elemeket melyik csoportba soroljuk, hanem az, hogy mi jellemző a csoportosítás után létrejött osztályokra (Bodon 2010, 147.). Az IBM SPSS Modeler szoftver három klaszterező algoritmust kínál: Kohonen, K-Means, TwoStep. A három modell közül kell kiválasztani a kutatási célnak legjobban megfelelőt.

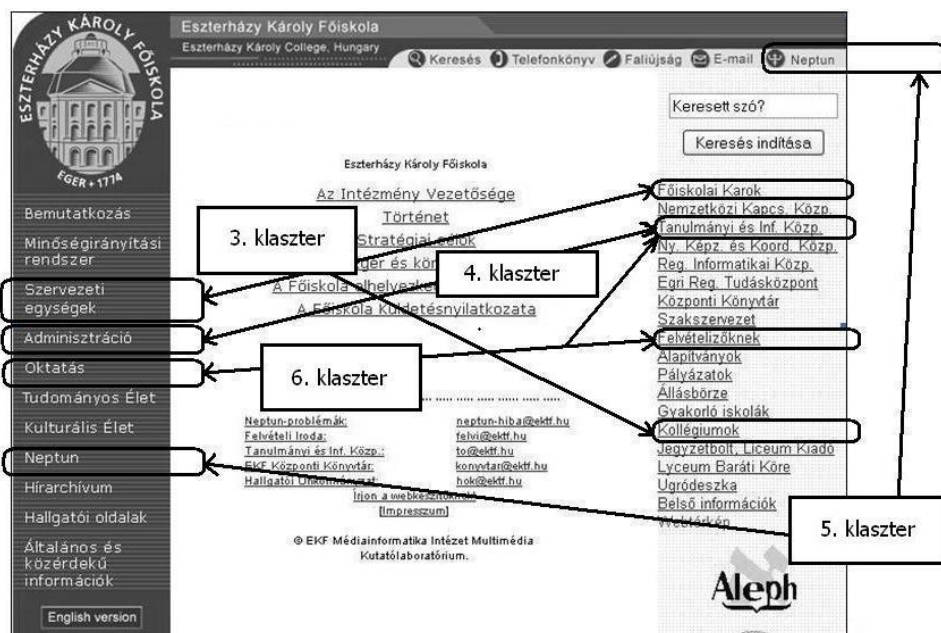
A vizsgálatra legalkalmasabb modell a 6 klasztert tartalmazó TwoStep modell, a klaszterekhez jelentést rendelve megkapjuk a látogatók csoportjait:

- 1. klaszter (12,8%): webkamerát kedvelő látogatók,
- 2. klaszter (25,1%): általános információt keresők,
- 3. klaszter (23,8%): a főiskola struktúráját látogatók,
- 4. klaszter (15,9%): tanulmányi osztályt vagy adminisztrációs információkat keresők,
- 5. klaszter (11,9%): NEPTUN-t keresők,
- 6. klaszter (10,5%): felvételi előtt álló látogatók.

A menü-rendszer szegmentumokhoz fűződő viszonyát a webergonómiai szempontok szerint a következőkben részletesen megvizsgáljuk (2. ábra). A webkamera kedvelők (cluster-1, 12,8%) a bal oldali menük alatti képre kattintva érik el a szolgáltatást, a menüpont elhelyezése elfogadható, hiszen az online szolgáltatások között jelentősége csekély. Az általános információt keresők (2. klaszter, 25,1%) a többi szegmentumba nem tartozó, tetszőleges tartalmú látogatások együttese.

A főiskola struktúráját látogatók (3. klaszter, 23,8%) a főiskolai karokat vagy a főiskola egységeit vagy a kollégiumot választják. A karokhoz tartozó menü a jobb oldali menücsoport első helyére került. A bal oldali elhelyezés a tartalmat és a kiválasztások számát tekintve sem indokolt, mivel a jelentősebb menüpontokat a bal oldalra helyezik el a honlap készítői. A „Kollégiumok” a bal oldali menücsoport alján helyezkednek el. A

„Főiskolai Karok” menüpontja a „Szervezeti egységek” és a „Kollégiumok” menüponttal (2. ábra) a menütervezés diszjunktivitás elvét is megsértik, mivel több menüpontban is elérhetjük ugyanazon egységeket. A totalitás elvét is megsérti a „Szervezeti egységek” menüpont, mivel nem tartalmazza az összes egységet. A menüpontokat egymástól eltérő menücsoportban helyezték el, ami átgondolandó, mert a látogató számára mentálisan megterhelő.



2. ábra: A 3-as, 4-es, 5-ös és 6-os klaszterekhez tartozó menüpontok a főiskola korábbi honlapján (Forrás: saját ábra.)

A Tanulmányi és Információs Központ látogatói szegmense (4. klaszter, 15,9%) jelentős látogatási arányt jelent. Ennek ellenére nem elme sem a bal, sem pedig a fenti, a felhasználók számára preferált menühalmazoknak, ugyanis a „Tanulmányi és Inf. Közp.” menüpontot jobb oldali menücsoport része. Az egység jelentősége megkívánná a könnyebben elérhető elhelyezést.

A Neptun tanulmányi rendszer információit letöltők (5. klaszter, 11,9%) számára a kapcsolódó menü a bal oldali menücsoport nyolcadik eleme, de található egy Neptun menüpont a felső menüsorban is. A két menüpont távol, egymástól eltérő menücsoportokban van és redundáns.

A felvételi előtt álló látogatók (6. klaszter, 10,5%) csoportja a felvételi információkat, a tanulmányi osztály oldalát és az oktatás menüt választják. A felvételi információkhoz a jobb oldalt 9. helyen szereplő „Felvételizőknek” menüpontot, a tanulmányi osztályhoz a jobb oldali harmadik („Tanulmányi és Inf. Közp.”) menüpontot, míg az „Oktatás” menüt a bal oldali ötödik menüpontban találja meg. A leendő hallgatók szokásaihoz

kapcsolódó menüpontok egymástól távol találhatóak. A fentebb említett statisztika alapján a felvételizők biztosan több klasztert is lefednek, ami széttagoltabbá teszi az általuk keresett információk menüpontjait.

A kutatás hipotézise igazolódott, mivel a megállapítások alapján lehet növelni az online látogatók kiszolgálásának hatékonyságát, és a megállapításokat egy adatbányászati klaszterező modell alkalmazása után lehetett megfogalmazni.

5. lépés. Az eredmények üzleti célú kiértékelése

Webergonómiai szempontból tekintve kijelenthető, hogy a felhasználók a mentálisan megterhelő használatot kerülik, ilyen esetben sokszor gyorsan elhagyják a honlapot (Krug, 2008, 21). A kutatásban a meghatározott szegmentumokhoz kapcsoltan több webergonómiai problémát tártunk fel, a felhasználók feltárt csoportjai elé számos akadályt gördít a honlap látogatás közben. A csoportok viselkedését a vizsgált honlap felépítése kevésbé támogatja. Javasolt a honlap struktúrájának jelentős átalakítása. Az oktatásmenedzsment megfogadhatja a leírtakat, és indítványozhatja a honlap menüstruktúrájának megváltoztatását, de dönthet a vizsgálat kiterjesztése és pontosítása mellett is.

6. lépés. Az eredmények üzleti célú alkalmazása

Beláttuk, hogy lehet és érdemes javítani a navigációt. A menedzsment döntése végül egy teljesen új honlap elkészítése volt, hiszen a honlap más hibákat is rejtett a feltártakon kívül, de ezeket 2007-ben az adatbányászat nélkülözésével tárták fel. A kutatási adatok birtokában akár évekkkel előbb megvalósulhatott volna a honlap korszerűsítése.

Összegzés

Az empirikus vizsgálat megmutatta, hogy az oktatási adatbányászat a vezetői döntések megalapozását képes elősegíteni az online marketingkommunikáció esetén. A döntések előkészítését vélhetően más adatokra támaszkodva is képes segíteni. A gazdasági előnyök számszerűsítve nehezen mutathatók ki, de hosszú távon az adatbányászat eszközeinek felhasználása bizonyosan versenylőnyt jelent az intézménynek.

Irodalom

- Abonyi, János Dr. (szerk.): Adatbányászat a hatékonyság eszköze. Gyakorlati útmutató kezdőknek és haladóknak, Budapest: ComputerBooks, 2006.
- Balogh, Imre (2010): Adatbányászat alkalmazása az oktatásmenedzsmentben in: Balogh, Imre-Horváth, Ádám (szerkesztők): Felhasználói viselkedés elemzése webes környezetekben. Web-analitikai módszerek alkalmazása viselkedés-elemzésre. DSGI, Bp.
- Bodon, Ferenc (2010): Adatbányászati algoritmusok 2010. február 28., Budapest: (online dokumentum)
URL: <http://www.cs.bme.hu/~bodon/magyar/adatbanyaszat/> (letöltés: 2010. május 20.)
- Bóta, László (2011): Supporting e-marketing decision making by the management of the Eszterházy Károly College via behaviour-based segmentation of the visitors of the institutional web-page, Problems of Management in the 21st Century, Siauliai (Litvánia): Scientific

- Methodical Center "Scientia Educologica", pp. 26-37.
URL: <http://journals.indexcopernicus.com/abstracted.php?level=5&icid=949988>
- Chapman, P., Clinton, J., Kerber, R. Khabaza, T., Reinartz T., Shearer C., Wirth R. (2000): CRISP-DM 1.0. Step-by-step data mining guide [online dokumentum], CRISP-DM Consortium, URL:
<ftp://ftp.software.ibm.com/software/analytics/spss/support/Modeler/Documentation/14/UseManual/CRISP-DM.pdf> (letöltve 2011. június 23.)
- Dinya, László (2010): Verseny a felsőoktatásban, a versenyhelyzet értelmezése, in: Törőcsik, M., Kuráth, G. (szerk.), Egyetemi marketing marketing a felsőoktatásban, Pécs: Pécsi Tudományegyetem, 49-67. old.
- Han, Jiawei-Kamber, Micheline (2004): Adatbányászat. Koncepciók és technikák, Budapest: Panem Kft.
- Izsó, Lajos-Tóth, Péter (2008): Applying Web-Mining Methods for Analysis of Student Behaviour in VLE Courses in Acta Polytechnica Hungarica Vol. 5. No. 4.
URL: <http://www.bmf.hu/journal/>
- Kotler, Philip-Keller, Kevin Lane (2006): Marketingmenedzsment, Budapest: Akadémiai Kiadó.
- Krauth, Péter (2008): Üzleti intelligencia in: Dömölki B., Kósa Zs., Kömlődi F., Krauth P., Rátai B.: Égen-Földön informatika. Az információs társadalom technológiai távlatai. Budapest: Typotex, 549-587. old.
- Krug, Steve (2008): Ne törd a fejem. Felhasználóbarát webdizájn. HVG, Bp.
- Nyéki, Lajos (2009): LMS usage mining in distance education. A mérnök-tanár képzés múltja, jelene és jövője – Bizterszky Elemeér Emlékkonferencia, BMF-BME, Budapest, 2009. november 3. ISBN 978-963-7154-89-8