

AZ ESZTERHÁZY KÁROLY FŐISKOLA HONLAPJÁNAK ELEMZÉSE WEBBÁNYÁSZATI MÓDSZEREK FELHASZNÁLÁSÁVAL

Bóta László

Eszterházy Károly Főiskola
3300 Eger
Eszterházy tér 1.
botal@ektf.hu

Absztrakt: Az Eszterházy Károly Főiskola honlapját a hallgatók, az oktatók, a felvételizők évek óta használják, azonban átfogó elemzés hiányában a változtatások, bővítések eredményessége, valamint azok minőségi megfelelése nem igazolható. A vizsgálat a webbányászati technikán belül a webnapló-bányászat alkalmazásával történt, az előadás az oktatási vonatkozású eredményekre tér ki. A kutatás az SPSS Clementine adatbányászati szoftver termékével, valamint a hozzá kapcsolódó WebMining CAT webbányászati modullal valósult meg, a BME Alkalmazott Pedagógia és Pszichológia és Intézet (APPI) Ergonómia és Pszichológia Tanszék Human Controller Interaction (HCI) kutatócsoportjában.

1. Bevezetés

Az egri Eszterházy Károly Főiskola honlapja (<http://www.ektf.hu>) az évek során többször megváltozott, jelentősen kibővült, azonban a honlapot használókról, valamint az ott elhelyezett dokumentumok határfokáról eddig még nem készült átfogó, szoftver-ergonómiai szempontokra épülő minőségvizsgálat. Az első honlap 1996. június 26-án készült el, majd 1998-ban következett az első arculatváltás, mindkét esetben a fő cél a tájékoztatás volt.

2002-ben újabb arculatváltás következett, ami voltaképpen a mai arculat. A tájékoztatás mellett az oktatást is segíti, hozzáférést biztosít a tanárok által elkészített oktatási segédletekhez, így valószínűsíthető volt, hogy a honlapot használók bizonyos viselkedési szokásai is tetten érhetők. A fejlesztések után elérhető a kereső szolgáltatás, a helyi webkamera, a saját webrádió, stb. A kutatás alapjául szolgáló honlap kiválasztását elősegítette a főiskolai honlap 2007. júliusra tervezett teljes átalakítása, így várhatóan a régi felhasználói felület ergonómiai hibáinak egy része elkerülhető.

2. A webbányászat, mint vizsgálati módszer

2006 szeptemberétől - csatlakozva az Izsó Lajos professzor úr vezetésével működő BME APPI Ergonómia és Pszichológia Tanszék HCI kutatócsoportjához -, lehetőség adódott az adatbányászati technikák elméletével, majd később gyakorlatával megismerkedni. Ez adta meg a vázolt vizsgálat alapját, a kivitelezést a kutatócsoport havonta megtartott munkaértekezlete támogatta.

A HCI kutatócsoport a költséges, de nélkülözhetetlen adatbányász szoftvert térítésmentesen használhatja az SPSS Hungary 2007. végéig szóló engedélyével, kizárólag non-profit célra. Az adatbányászati Clementine szoftver mellett a WebMining CAT modult is biztosították, ami a webbányászati tevékenységek gyakorlati megvalósítását teszi lehetővé.

A webbányászat (web mining) az adatbányászat egyik részterülete. Az adatbányászat (data mining) a tudás adatbázisokban, adattárházakban vagy egyéb információtárolókban tárolt, nagy mennyiségű adatból történő kiválasztása, kibányászása [1]. Ennek megfelelően egy új tudáshoz juthatunk az alkalmazásával, ami az eddigi adatbázis-technológián alapuló eszközökkel felfedezhetetlen. Az

adatbányászat minden olyan területen használható, ahol a tevékenység során nagy mennyiségű adat keletkezik (pl. üzleti, államigazgatási, egészségügyi, oktatási, stb.). Az adatbányászat a statisztika elemeit használja ugyan, de messze túlmutat azon, egy igen mélyreható matematikai és informatikai alapot igényel. [2]

A World Wide Web a hozzá kapcsolódó osztott információs szolgáltatásokkal ma már egyebek mellett üzleti, oktatási felület is. A vásárlók, látogatók, hallgatók igényeit kielégítő szolgáltatás csak abban az esetben valósítható meg, ha felhasználók szokásait, magatartását elemezzük a weboldalakon található linkeken át bejárt útvonalak alapján. A felhasználók hozzáférési mintáinak ilyen osztott információs környezetben történő kigyűjtését nevezik hozzáférési útvonal minta bányászatnak. A vizsgálatokat néhány említésre méltó tény nehezíti: a web túlságosan bőséges tárolt adathalmaza, a weboldal bonyolultsága, a weboldal dinamikusan változásai, a felhasználók azonosítási anomáliái. A webbányászati vizsgálatok középpontjában az alábbiak állnak: webelérési minták, webstruktúrák, webtartalmak szabályossága és dinamikája. A webbányászat feladatköre három területet fog át: webtartalom-bányászat (web content mining), webstruktúra-bányászat (web structure mining) és webhasználat-bányászat (web usage mining). Ugyanakkor a webstruktúra a webtartalom része, így a webbányászat következő két területéről beszélhetünk [1]: webtartalom-bányászat, webhasználat-bányászat.

Vizsgálódásaink a webbányászaton belül webhasználat-bányászat területére terjednek ki, ami a webnapló bányászaton alapul. A webnapló-bejegyzések vizsgálata során a weblapok hozzáférési mintáit próbáljuk feltárni, így javíthatunk a felhasználók számára, az interneten keresztül nyújtott szolgáltatások minőségén, vagy javítható akár a webszerverek teljesítménye is. Egy webszerver a weboldalak elérése esetén bejegyzést készít, ezen adathalmazt nevezzük webnaplónak, amit egy naplófájlban, ismertebb nevén logfájlban tárol. Egy adott webnapló-bejegyzés mezői eltérőek lehetnek, de mindegyik webszerver tárolja a következőket:

- a kérés kiindulási helyének IP címét,
- a kérés pontos idejét,
- a kért URL címet,
- és még néhány adatot.

A webnapló adatait nem lehet azonnal felhasználni, a tudás érvényességének és megbízhatóságának érdekében többnyire szükséges az előfeldolgozás, amelynek módszerei: tisztítás, tömörítés, transzformálás, amelyhez a WebMining CAT modulja jelentékeny segítséget ad. Az előfeldolgozást követi az elemzés, amelynek eredményeként a felhasználók böngészési szokásai tárulnak fel.

3. A vizsgálat előkészületei

A jelenleg is folytatott, komplex kutatásból jelen előadásban két részeredmény szerepel: a kezdőoldal navigációjából az egyik menüpont, az Oktatás menü elemzése, illetve a saját készítésű tanári weboldalakhoz kapcsolódó kérdés megválaszolása a saját honlapom esetén.

A vizsgálathoz a rendszergazda átadta az egri főiskola webszervere által összegyűjtött, ún. access log fájlokat (elérési naplóállományokat), amelyek a kutatás alatt eltérő időpontokban lehetnek szükségesek. Kiindulópontként a WebMining program CAT moduljával a legnépszerűbb eseményeket, illetve az események számát tekintettük meg.

EKF_Events.txt - Jegyzettömb				
Fájl	Szerkesztés	Formátum	Nézet	Súgó
Event Category	Event Name	Event Definition	Event Attributes	
Homepage	Homepage	/		
Homepage	Kezdőoldal	/index.php		
Keresések	Keresések	/kereso.php/*		
Telefonkönyv	Telefonkönyv	/telefonkonyv/*		
Tanulmányi_ugy	Tanulmányi_ugy	/tik/*		
Sajatsite	Sajatsite	/~botal/*		
Mas_kepzes	Mas_kepzes	/oktatas/*		
EKF_egysegek	EKF_egysegek	/szervezet/*		

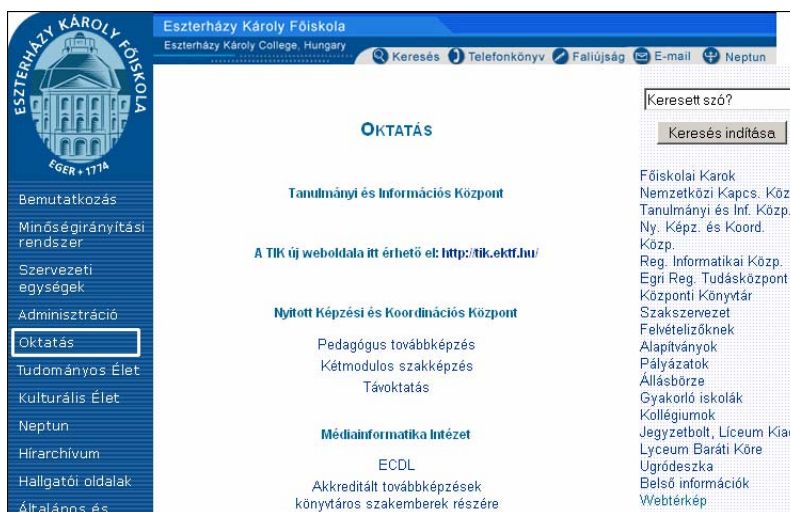
1. ábra
 A vizsgálathoz készült eseményfájl

Esemény alatt itt azt értjük, ha a honlap használója egy letöltésbe kezd, amit többnyire a böngészőben megjelenő felületen linkre, menüre, képre vagy gombokra (pl. elküld, jelentkezés) kattintást jelent. Az események száma jelentős, hiszen számottevő a vizsgált honlap napi látogatottsága. Az események közül célszerű kiszűrni azokat, amelyek a vizsgálat középpontjában vannak, így jelentősen csökkenthető az adatok feldolgozási ideje. Az említett eseményfájl voltaképpen egy szöveges állomány, ami látható az 1. ábrán.

Az eseményfájlban több esemény látható, a terjedelmi korlátok miatt itt az eredmények egy részét mutatjuk be. Az első két oszlopot nekünk kell kitölteni, ami a vizsgált weblapok alapos ismerete nélkül nem lehetséges. A harmadik oszlopba pedig a logfájlban megtalálható hivatkozások pontos útvonalát kell begépelni. Az Oktatás menü a „MAS_kepzes”, míg a saját készítésű oktatói weboldal a „Sajatsite” eseménynevet kapta, az előbbi elnevezés szerepel a 3. ábrán.

4. Az első vizsgálat

Az első vizsgálat fókuszja a honlap bal oldali menüsorának Oktatás menüje. A főiskola honlapjának viszonylag rendszeres látogatója vagyok, ennek ellenére az Oktatás menüt több alkalommal választottam ki helytelenül, ugyanis a menüben csak a látogatók egy szűk csoportjának igényeit kielégítő, kiegészítő képzések, tanfolyamok találhatók.

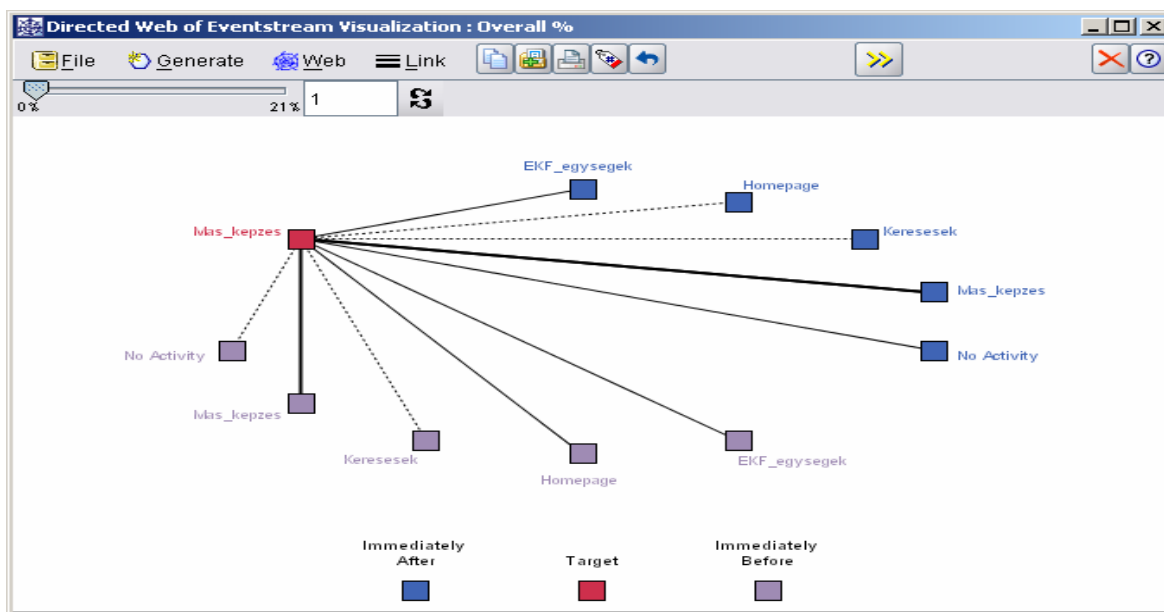


2. ábra
 Az Oktatás menü kiválasztása utáni felület

Az Oktatás menüt - a honlap hozzávetőlegesen 200 ezres kiválasztásához (hit) viszonyítva is -, elég sokszor választják ki, az almenüivel együtt januárban több, mint 11 ezerszer, februárban 8 ezer felett.

Az Oktatás menü előtti esemény (választás) a januári időszakban 500-nál több esetben olyan weboldal volt, amit az oktatással összefüggésbe lehet hozni, de ez még nem bizonyítja azt, hogy a menü félrevezető. Az Oktatás menüvel kiválasztott weboldal utáni oldalak azonban már érzékeltetik a problémát, hiszen a közel 5000 esemény, azaz oldal választása során 650-nél több utal arra, hogy a képzéssel kapcsolatban lépett tovább a látogató. Jelentős a Szervezeti egységek menü kiválasztása, ahol többek között a főiskola képzéseiről is olvashatunk.

A kiválasztott kételemű útvonalak jobban kiemelik a problémát. Az Oktatás menü után 4,21%-ban választották a látogatók a nyitóoldalt (visszaléptek), 1,99 %-ban az Szervezeti egységek menüt, és további 0,63%-ban fordultak a honlap kereső funkciójához. Ez nem tűnhet soknak az összes kételemű út között, de ez együtt 5343 kattintást jelent. Az összes kiválasztott 3 elemű útvonal 3,5%-nál választották a látogatók az Oktatás menü kétszeri választása után a nyitóoldalt (visszaléptek), illetve 1,36%-nál az Szervezeti egységek menüt, ami összesen 3353 kattintást jelent. Az összes kiválasztott 4 elemű útvonal 1,2%-nál fordult elő az Oktatás menü és az Szervezeti egységek menü egymás utáni kétszeri kiválasztása, ami 1086 kattintást jelent. Tehát az azonos kattintás sorozatok számát vizsgálva leszögezhető, hogy igen jelentős az Oktatás, Szervezeti egységek menü és a kezdőoldal közötti kattintások száma.



3. ábra

Az Oktatás menü kiválasztása előtti és utáni weboldalak (események)

Az Oktatás menü utáni események vizsgálatánál is kitűnik, hogy sokan „eltévednek”, mivel az egyik legjelentősebb kattintás szám a Szervezeti egységek (EKF_egysegek) menühöz tartozik, amit a 3. ábrán a viszonylag vastag, még nem szaggatott vonal jelöl. A menütől a tanszékeket, tanegységeket, a tanulmányi ügyekkel foglalkozó weboldalakat várják, de mégsem kapják meg.

A honlap zsákutcái alatt azt értjük, hogy a felhasználó pár kattintással visszajut valamely weboldalra (ennél a vizsgálatnál a kezdőoldalra), és itt befejezi a látogatást. Annak ellenére, hogy a

honlapon állandóan megjelennek a menük, mégis az Oktatás menüből 135-ször visszaléptek a kezdőoldalra, és befejezték a böngészést. Ez önmagában nem bizonyítaná azt, hogy az Oktatás menü félrevezető, de a fentebb soroltakkal együtt igazolható, hogy sokak számára megtévesztő a menü elnevezése, és tévútra kerülnek.

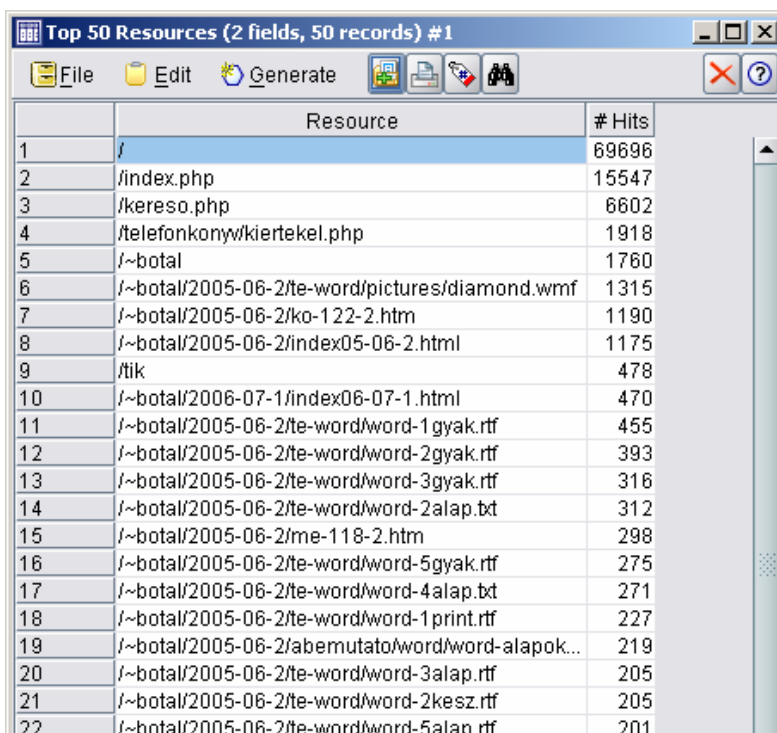
A minőségi problémát a honlap fejlesztői felé továbbítottam, akik az új honlap kialakításánál figyelembe veszik a tapasztaltakat. A vizsgálatokat mindenképpen szükséges folytatni a régi, és az új honlap adataival egyaránt.

5. A második vizsgálat

A második vizsgálat alapja egy ismert oktatási probléma. A digitális szövegfeldolgozás tanegységet felvett hallgatók első zárthelyi dolgozatának tanulmányi átlaga igen alacsony, bár a tananyag nem túl nehéz, és mindössze 3 hét ismeretiből kell beszámolni a kiadott feladatokból.

A tantárgyat 2006/07-es tanév tavaszi szemeszterében 118 hallgató vette fel, ebből az első zárthelyi dolgozatot 50 fő írta meg februárban, a tanulmányi átlag: 2,64. Az oldal látogatásához a honlapon belül a „/~botal” kezdetű útvonal valamelyikét kell kiválasztani.

A fő problémák a következők okozhatják: rossz magyarázat és feladatok, a gyakorlás hiánya, tanulás módszertani elégtelenség. A második pontot, a gyakorlás hiányát kívántam kiemelni, feltételezve, hogy ennek jelentős szerepe van.



	Resource	# Hits
1	/	69696
2	/index.php	15547
3	/kereso.php	6602
4	/telefonkonywkiertekel.php	1918
5	/~botal	1760
6	/~botal/2005-06-2/te-word/pictures/diamond.wmf	1315
7	/~botal/2005-06-2/ko-122-2.htm	1190
8	/~botal/2005-06-2/index05-06-2.html	1175
9	/tik	478
10	/~botal/2006-07-1/index06-07-1.html	470
11	/~botal/2005-06-2/te-word/word-1gyak.rtf	455
12	/~botal/2005-06-2/te-word/word-2gyak.rtf	393
13	/~botal/2005-06-2/te-word/word-3gyak.rtf	316
14	/~botal/2005-06-2/te-word/word-2alap.txt	312
15	/~botal/2005-06-2/me-118-2.htm	298
16	/~botal/2005-06-2/te-word/word-5gyak.rtf	275
17	/~botal/2005-06-2/te-word/word-4alap.txt	271
18	/~botal/2005-06-2/te-word/word-1print.rtf	227
19	/~botal/2005-06-2/abemutato/word/word-alapok...	219
20	/~botal/2005-06-2/te-word/word-3alap.rtf	205
21	/~botal/2005-06-2/te-word/word-2kesz.rtf	205
22	/~botal/2005-06-2/te-word/word-5alap.rtf	201

4. ábra

A saját honlapom, és az oktatási anyagok letöltésének vizsgálati eredményei („/~botal”-ból nyíló mappák)

A vizsgálat alapján (4. ábra) kiderült, hogy januárban - a szemeszter kezdete előtt - több mint 750-szer, ezzel szemben februárban - a szemeszter kezdetével – 6000-nél is többször voltak a hallgatók az oldalakon, ami azt érzékelteti, hogy a 118 hallgató jelentős része legalább egyszer megtekintette

a gyakorló feladatokat. A gyakorló állományok bármelyikének a letöltése is legalább 200-szor történt meg, ami bizonyossá teszi, hogy nem a gyakorlat hiánya okozza a gyenge teljesítményt.

A vizsgálat folytatása mindenképpen szükséges, a jövő évi csoportnál kérdőív segítségével kell fényt deríteni arra, hogy a hallgatók mely tulajdonsága, szokása gátolja a tudásuk alatti eredményt.

Összegzés

A webbányászati szoftver legalapvetőbb eszközeivel közelítettük meg a problémákat, de már így is pontos, megbízható vizsgálati eredmények segíthetik a további pedagógiai munkát. A lehetőség az oktatásban nagyon megbízható, gyorsan megválaszolható, az oktatásban szereplők mindegyikére vonatkozó vizsgálatot ad a pedagógus kezébe. Az általános használata egykoron vélhetően egy ingyenes ún. Web 2.0-ás technikát használó oldal segítségével valósulhat majd meg, mivel a szükséges költségek az oktatásban még nem vállalhatók.

Irodalomjegyzék

- [1] Jiawei Han, Micheline Kamber: Adatbányászat. Konceptiók és technikák, Budapest, 2004, Panem
- [2] Bodon Ferenc: Adatbányászati algoritmusok. Elektronikus dokumentum. 2006. november 26. URL: <http://www.cs.bme.hu/~bodon/magyar/adatbanyaszat/tanulmany/index.html>