

BDF WEBSITE ELEMZÉSE AZ SPSS CLEMENTINE WEB MINING SEGÍTSÉGÉVEL MINŐSÉGBIZTOSÍTÁS CÉLJÁBÓL

Zsiros Péter

BDF Alkalmazott Informatika és Információmenedzsment Tanszék
9700 Szombathely, Károlyi Gáspár tér 4.
zsirp@bdf.hu

Absztrakt: A BDF honlapja saját fejlesztésű, felépítésének alakulását a készítése idején több független tényező befolyásolta. Vajon ez a struktúra megfelel-e a hallgatók és a dolgozók igényeinek? Milyen a hallgatók és az azonosíthatóan bejelentkezett belső munkatársak felhasználói viselkedése? A honlapot látogatók milyen viselkedésmintákat lehet elkülöníteni? Saját oktatói portálon látogatói milyen gyakorisággal, milyen céllal érkeznek? Állíthatunk-e valamit a látogatók viselkedéséről? Mindezek alapján milyen esetleges változtatások lennének szükségesek? – ezekre a problémákra keresem a választ vizsgálataimmal. A cél a jobb minőség elérése, a site-tal a hallgatók és a belső dolgozók hatékonyabb kiszolgálása. Előadásom több ponton kapcsolódik dr. Izsó Lajos plenáris előadásához, továbbá dr. Balogh Imre és Grujber Zoltán e szekción belüli előadásához.

1. Bevezetés

Grujber Zoltán kollégám bemutatta, hogy a Google Analytics segítségével meddig lehet eljutni website-ok elemzésében. A BDF website-ra lefuttatott Google elemzés a minőségbiztosítás és az absztraktban feltett kérdések szempontjából is több érdekes eredményt ad, ám ezek közül jó néhány további alapos elemzést igényel. A Google eredmények közt felmerülnek olyan problémák is, amelyekre tipikusan csak a Clementine Web Mining adhat választ.

Egyik Google által alátámasztott eredmény, hogy erős heti periodicitás jelentkezik a site látogatottságában, illetve hogy a hosszabb időszakokra vonatkozó adatok is mutatnak lassú növekvő-csökkenő tendenciákat. A további elemzések deríthetnek fényt ennek okára, további részleteire.



A célpopulációt tehát első lépésben részekre bontottam. A belső látogatók egy része olyan hallgató, aki a főiskola gépén keresztül lép a website-ra, másik része olyan intézményi alkalmazott, aki a belső hálózatba bejelentkezve usernév szerint azonosíthatóan nézi valamelyik lapot. A Clementine szoftver lehetőséget ad arra is, hogy a látogatók IP-címe alapján elkülönítsük a belső és külső látogatókat, illetve hogy azonosítható usernév alapján megállapíthassuk bejelentkezett dolgozó esetén akár a látogató személyét is. E látogatókat „bennfenteseknek” neveztem el vizsgálataim során.

2. A BDF website felépítésével kapcsolatos kérdések

A nyitólapon az oldallécben és fejlécben is linkek találhatók. Ezen keresztül érhetik el a külső látogatók az intézményről szóló fontosabb információkat, a hallgatók a tanulmányaikhoz szükséges információkat, az oktatók portáljaira feltett többségben letölthető tananyagokat, illetve a főiskola alkalmazottai a munkájukhoz szükséges dokumentumokat. A szervezeti egységek lapon belül az egyes szervezeti egységekre kattintva egységes felépítésű lapokat találunk, ha van az adott egységnek saját honlapja, akkor az e lapról újabb kattintással érhető el.



Ezzel kapcsolatos minőségbiztosítási megfigyelés, hogy a website hatékonysága jelentősen növelhető, ha a leggyakrabban keresett honlapokat ezek, az a Clementine Web Mining

közvetlenül a nyitólapról el lehetne érni. Hogy melyek segítségével igazolható, ennek alapján tehetünk a további fejlesztésekre vonatkozó javaslatokat. Arról készítettem kimutatást, mely oldalakon kezdődnek a felhasználói visíték, az oldalak elsőégi statisztikáját mutatja a kép.

Például az *Informatika* mint kezdőoldal hatodik e rangsorban. Ha az *Alkalmazott Informatika és Információmenedzsment Tanszék* honlapjára akar eljutni egy látogató, akkor a főlapról egy kattintás után egy 92 elemű listában kell kikeresnie a megfelelőt, aztán még kettő kattintás után éri el a honlapot. Látható, hogy a látogatók nagy része az *informatika.bdf.hu* cím bebillentyűzését jóval egyszerűbbnek találta, ha el akarta érni a tanszéki weblapot. És ha egyszer bebillentyűzte a címet, utána az első betűk beírása után már a böngésző felkínálja a címsáv egysoros listájában, így a későbbiekben visítjét ezen az oldalon kezdi.

A belső, külső látogatók és bennfentések megoszlásának vizsgálata újabb minőségbiztosítási javaslatokat adhat az oktatói portálok átgondolt újabb egységbe, egységekbe szervezésére.

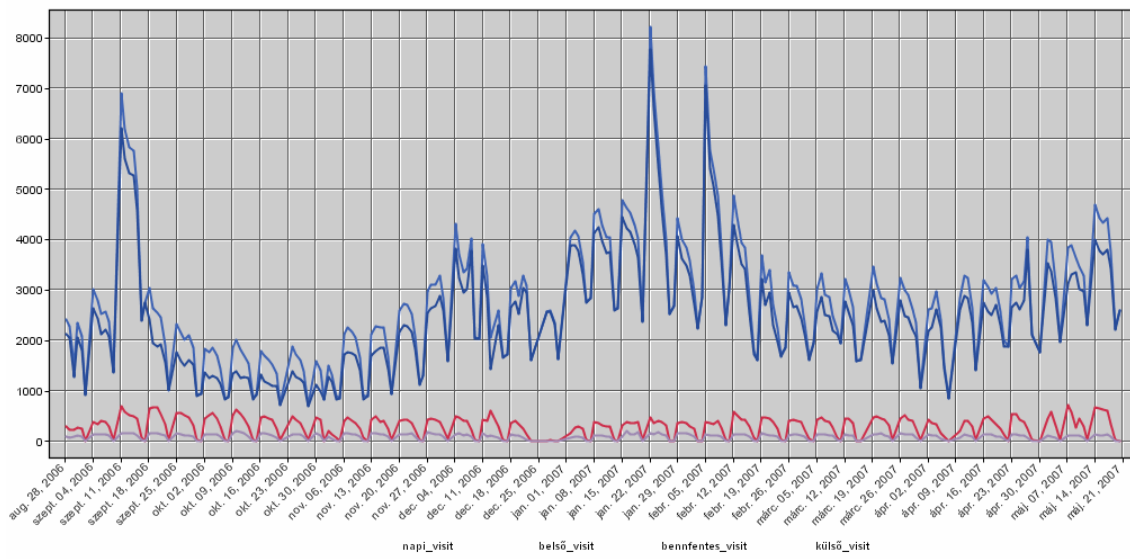
	Event Name	lttkezd_belső	lttkezd_külső	lttkezd_bennf
1	HomePage	57022	331858	17557
2	Neptun	1802	110532	41
3	HOK	1722	37088	311
4	Felveteli	221	21646	47
5	EgySzervezetiEgység	281	15825	64
6	Informatika	5709	11797	353
7	PSZK	74	9184	202
8	Hírek	934	7554	243
9	TTK	148	6839	56
10	MSZI	136	6660	59
11	Események	256	6429	71
12	Hallgatóknak	1251	6322	80
13	FKK	47	5137	170
14	Karrier	136	4540	67
15	HallgatóknakPalyaz...	59	3766	29
16	Szkl	128	3648	74
17	EN	21	2977	7
18	IanTemekEsFormat...	15	2866	9
19	TMK	25	2813	14
20	BTK	164	2230	57
21	SzervezetiEgységLi...	576	1784	79
22	MuszakiMenedzser	22	1767	7
23	Rekreacio	1	1179	0
24	Andragogia	3	1120	0
25	Erasmus	40	1113	17
26	Tanito	1	1073	0
27	HallgatóknakDoku...	63	1033	54
28	Elerhetosegek	49	1000	2
29	TermektervezőMusz...	5	986	4
30	Romanisztika	2	983	0
31	MuszakiInformatikai...	2	944	0
32	Kornyezettan	1	943	0
33	Foldrajz	2	937	0
34	InformatikusKonyrta...	3	927	1
35	GazdalkodásiMene...	5	925	1
36	IntezmenyiKommun...	0	917	0
37	Matematika	0	914	0
38	GazdalkodásiMene...	5	898	1
39	IdőnapfornalmiSzak	18	886	9

A most feltett kérdésekre adott válaszokat a prezentációmban részletezem.

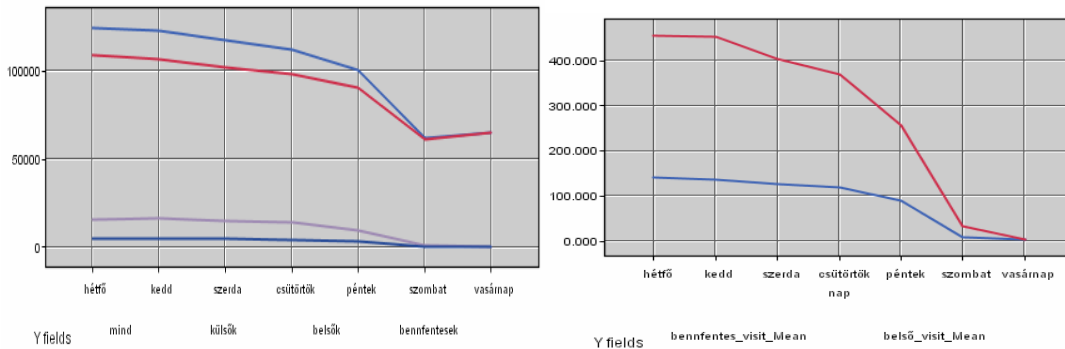
3. Belső és külső látogatók

Vizsgálatom másik területe, hogy a website belső hálózati és külső látogatóinak viselkedése közt milyen releváns különbségeket lehet megfigyelni. Egyik hipotézisem az, hogy a Neptun oldalt a belső, a képzések oldalait elsősorban a külső látogatók látogatják. További vizsgálat tárgya, hogy a userek heti periodicitásában milyen arányban vesznek részt a külső és a belső látogatók.

Kollégám a Google Analytics segítségével erős heti periodicitást mutatott ki a userek aktivitásában, ennek további vizsgálatait végzem a Clementine Web mining segítségével. A periodicitás mértékét és a szezonalitástól mentesített adatok tendenciáit vizsgálom. További hipotézis, hogy a periodicitás csak bizonyos típusú userek miatt mutatkozik. Ennek eldöntését is Clementine streamek segítségével végzem el. Kiderült, hogy mindhárom usertípusnál megjelenik a periodicitás, a belsőknél a legerősebb.



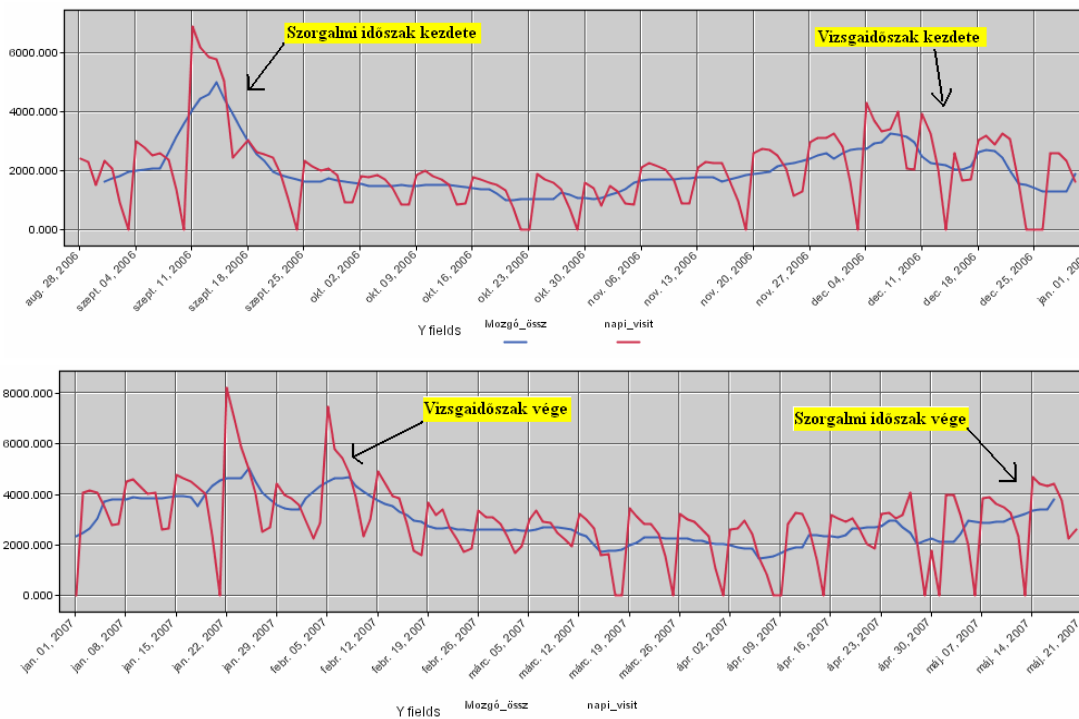
A heti periodicitást napra vonatkozó átlagolással is vizsgáltam mindhárom usertípus esetére.



Fontos, hogy milyen a kapcsolat, milyen erős a kapcsolat a userek periodikus tulajdonsága, illetve a „külső-belső látogató” tulajdonság között. A grafikonok mutatják az eredményeket.

Probléma, hogy olyan erős a heti periodicitás, hogy a grafikonokon nem lehet trendeket megfigyelni. A Clementine belső függvényeinek felhasználásával olyan streamet építettem, amely statisztikai mozgóátlagolást végez, így jó látható, elemezhető a látogatások trendje, ennek időbeli

alakulását összevettem a főiskola munkarendjével, a vizsgaidőszakok és szorgalmi időszakok váltakozásával. Az alábbi grafikonokon ez látható 2006 augusztusától év végéig, illetve 2007. elejétől május végéig.



4. Saját oktatói portálom látogatásainak elemzése

A tanszéki honlapokról el lehet érni az egyes oktatói portálokat is, ezeknek kihasználtságát is a Clementine-nal lehet vizsgálni, elsősorban a saját oktatói portálomat elemeztem. Fontos szempont a látogatók száma, a látogatók lapon töltött idejének megfigyelése, illetve az, hogy befejezik-e a lapon a látogatást vagy esetleg csak kíváncsiságból vagy tévedésből kerültek ide.

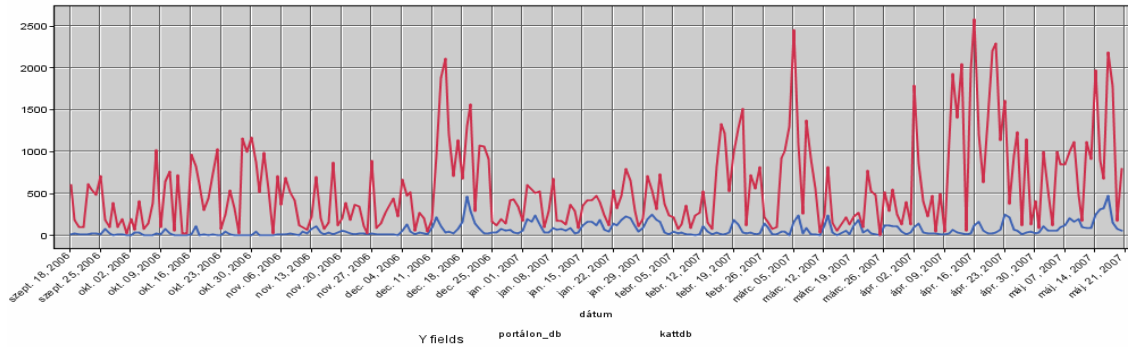
A portálon tantárgyak szerinti mappákba csoportosítva található az oktatási segédletek. A segédletek egy része elméleti jellegű tananyag, egy másik része gyakorlás célját szolgáló feladatsor. Vannak elméleti vizsgákra szóló tételSORok és elért pontszámokat, osztályzatokat tartalmazó táblázatok is. Emellett néhány fájl időpontokat, vizsga menetének leírását vagy egyéb technikai jellegű oktatással és számonkéréssel kapcsolatos információkat tartalmaz.

Azokat a visiteket figyeltem, amelyek tartalmaztak a portálon való kattintást. Az ilyen visiteketben kiszámítottam a portálon való kattintások számát és az összes kattintások számát, továbbá ezek arányát, valamint e visitek sessionhosszát.

Katt%	
Statistics	
Mean	0.534
Min	0.001
Max	1.000
Variance	0.083
Standard Deviation	0.288
Pearson Correlations	
portálon_katt	0.086 Strong
portálon_katt	
Statistics	
Mean	3.192
Min	1.000
Max	124.000
Variance	12.624
Standard Deviation	3.553
összkatt_db	
Statistics	
Mean	27.478
Min	1
Max	1741
Variance	10739.535
Standard Deviation	103.632
Pearson Correlations	
portálon_katt	0.159 Strong
session_hossza	
Statistics	
Mean	39.781
Min	0.000
Max	719.983
Variance	19584.903
Standard Deviation	139.946
Pearson Correlations	
portálon_katt	0.143 Strong

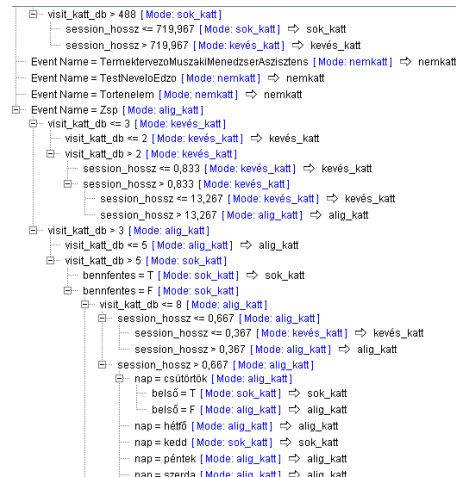
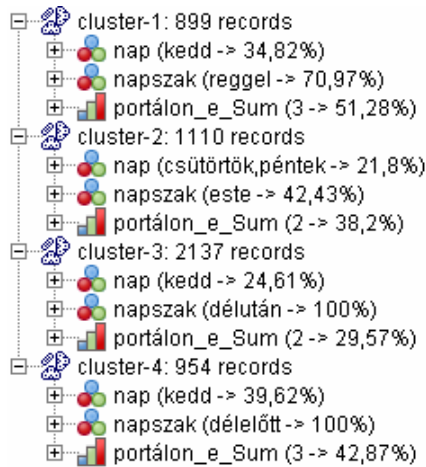
Számításokat végeztem a Clementine-nal, hogy mennyi e statisztikai változók átlaga, szórása, minimuma és maximuma, illetve korrelációt számítottam e változók és a visitek portálon való kattintásainak száma között. Az eredményeket az ábra mutatja.

A portálon való kattintást tartalmazó visitek esetében ábrázoltam a portálon való napi kattintások és az összes napi kattintások számát a dátum függvényében.



Kétféle statisztikai Clementine modellt építettem a látogatók típusokba sorolása céljából:

- Visitek Twostep klaszterezése időpont és kattintási mennyiség szerint:
- C5.0 döntési fa, a cél, hogy egy visitben mennyi kattintás a portálon. A „mennyi kattintás” fogalmát is Twostep modellel alakítottam ki.



A jövőben más oktatói portálok hasonló elemzését is érdemes lenne elvégezni, összehasonlításokat végezni. Az összehasonlításból kapott eredmények az oktatói hatékonyság növelését céloznák.

5. A kutatás folytatása

A főiskolai webserver naplófájljai 2006. augusztus 28-tól állnak rendelkezésünkre, ennyi technikai jellegű információ alapján két teljes félévre szóló elemzést tudtam végezni. Ezen adathalmaz segítségével kimutatott törvényszerűségeket a jövő félévben is ellenőrizni lehet, a megépített stream-jeimet az újabb adathalmazon is futtathatom majd. Ha az elemzések nyomán született javaslatok megvalósulnak, akkor újabb munkafázis következik, a minőségbiztosítás szabályai szerint visszacsatolásként ellenőrizni kell, hogy a most kitűzött célok megvalósultak-e.