

## GOOGLE ANALITYCS VS. SPSS CLEMENTINE

**Grujber Zoltán**

Berzsenyi Dániel Főiskola  
9700 Szombathely, Károlyi Gáspár tér 4.  
gzoltan@bdf.hu

---

*Absztrakt: A Google Analytics egy ingyenesen hozzáférhető, meglehetősen okos és látványos webhely-látogatottság elemző alkalmazás. Az alkalmazásának lehetőségein túl szeretném a rendszer szolgáltatásait szembeállítani a Clementine WebMining képességeivel, összehasonlítani a két rendszer tudását, korlátaikat feszegetni.*

---

### 1. Bevezető

A Google Analytics [1] rendszere ingyenes regisztrációval hozzáférhető. Legnagyobb korlátja a használatának, hogy maximum 5 millió oldalletöltést kezel le havonta. Bár a legtöbb weblapot üzemeltető cég örülne ekkora látogatottságnak. Ennél nagyobb látogatottságú weblapok használatának elemzésére már komolyabb elemző eszközt kell használni, például a Clementine-t [3].

Az alábbi fejezetekben a két eszköz képességeit állítanám szembe egymással, konkrét felhasználási célok figyelembevételével.

### 2. Adatgyűjtés – A látogatás naplózása

A Google rendszerének az alapja, hogy minden egyes olyan weboldal forrásába be kell ágyazni egy kis kódrészletet, amelynek a látogatásait naplózni szeretnénk. Emellett támogatja a weboldalon történő egyéb események naplózásait is, például: egy elem kiválasztása egy listából. A technológiai háttérben a JavaScript programozási nyelv alkalmazása áll.

A Clementine WebMining az elemzést a weblapot kiszolgáló webszerver napló állományára építi. Többféle formátum szerint is naplózhat a webszerver. A legáltalánosabb ajánlás a W3C konzorcium által kidolgozott. A legtöbb webszerveren konfigurálható, hogy a naplózás milyen mélységű legyen. [1]

A Google rendszerének előnyei és hátrányai:

- Olyan adatokat is szolgáltathat a JavaScript, amelyeket a webszerver nem naplóz, például: képernyőfelbontás, nyelvi beállítás, stb.
- Weboldalon belüli események rögzítése
- Nem működik, ha a böngészőben tiltott a JavaScript futtatása, a cookie-k használata, és a képek megjelenítése
- Nagyon kényelmetlen és sok munka lehet a webhely minden egyes oldalába beépíteni a forráskód részletet

A Clementine WebMining előnyei és hátrányai:

- Nem kell az oldalak forráskódját módosítani, csak be kell kapcsolni a webszerveren a naplózást

### 2.1. A két rendszer eredményei között mérési különbségek adódhatnak

Az látogatottsági elemzések egyik fontos sarokköve a látogatók azonosítása. Ennek két alapvető módszere elterjedt. A süti alapú, vagy az IP + User Agent alapú azonosítás. A Google az internetező számítógépre elhelyezett egyedi sütivel (cookie) azonosítja a visszatérő látogatókat. A Clementine WebMining pedig a naplóállomány alapján az internetező számítógépének IP címe és a naplóban szereplő User Agent bejegyzés alapján. Ez a második fajta módszer nagyobb találati számot produkál általában a dinamikus IP-cím kiosztás miatt, valamint a keresőrobotok találatai miatt is.

További eltérést jelenthet a cach technika megléte. A Google megoldása a JavaScript rutinon keresztül direktben készít naplóbejegyzést. Míg ha a böngésző a saját cache-éből szolgálja ki ismét az oldalt, a webszerver naplóállományában nem jön létre újabb bejegyzés.

### 2.2. Külső hivatkozások használata

A weblapon lévő külső hivatkozásokra kattintva azok is generálnak naplóbejegyzést, csak hogy azok a hivatkozott oldalt kiszolgáló szerveren keletkeznek, amelyekhez valószínűleg nem férhetünk hozzá. Általában elmondható, hogy a webszerver által rögzített naplóból nem kapunk arra vonatkozó információt, hogy mikor és hányszor kattintnak külső hivatkozásokra.

A Google Analytics nyújt megoldást erre a problémára is. Pusztán csak módosítani kell a forrásban a linkre vonatkozó részt.

```
<a href="http://www.example.com"
onClick="javascript:urchinTracker
('/outgoing/example_com');">
```

1. ábra  
Külső hivatkozások lekövetése a Google Analytics-ben

## 3. Az adatok tisztítása, szűrése

A naplóállományok a legtöbb esetben rengeteg olyan bejegyzést tartalmazhatnak, amely az elemzések szempontjából felesleges. Például, a webszerver naplóállományába bejegyzés kerül a letöltött weboldalakon túl a weboldalak felépítő stílus, kép és script állományokról is. Ez azt jelenti, hogy egy kattintás akár nagyon sok felesleges naplóbejegyzést is generálhat. Az adatok kiszűrése, tisztítása elengedhetetlen a tényleges elemzés megkezdése előtt.

A Google Analytics alkalmazása esetén valójában csak azokról az oldalokról és eseményekről készül naplóbejegyzés, amelyek forrásába a naplózást végző kódrészletet elhelyeztük. Így úgymond eleve tiszta adatokat rögzítünk csak. Továbbá a program lehetővé teszi külön szűrők definiálását is, melyekkel az elemzésben résztvevő adatok körét szűkíthetjük, vagy akár transzformációkat is végezhetünk az adatokon.

Add new Filter for Profile

Filter Name:

Filter Type:

Exclude  
 Include  
 Lowercase  
 Uppercase  
 Search and Replace  
 Lookup Table  
 Advanced

Filter Field:

Filter Pattern:  [What do the special characters mean?](#)

Case Sensitive:  Yes  No

Cancel Finish »

2. ábra  
Szűrő hozzáadása a Google Analytics-ben

A Clementine WebMining az adatok tisztítását és szűrését a WebMining Node beállításai közül az EventDefinition fájl megadásával teszi lehetővé. Ebben a fájlban részletesen megadhatjuk, hogy mely mappákra vagy fájlok bejegyzéseit szeretnénk figyelembe venni, különbségeket tehetünk a GET módszerrel átadott paraméterek között, szabadon elnevezett kategóriákba és eseményekbe csoportosíthatjuk a napló bejegyzéseit. Összességében elmondhatjuk, hogy nagyon rugalmasan testre szabható lehetőségről van szó, mely szinte minden igényt kielégíthet, bár hatékony alkalmazásához elengedhetetlen a webhely szerkezetének és működésének részletes ismerete.

Event Category	Event Name	Event Definition	Event Attributes
Hirek	EgyHir	/News/View.aspx	ID[Hir ID]
Hirek	MindenHir	/News/List.aspx	
Hirek	HirArchivum	/News/Archiv.aspx	
Események	MindenEsemeny	/Events/Calendar.aspx	CalendarPeriod[Periodus]
Események	EgyEsemeny	/Events/View.aspx	ID[Esemeny ID]
Események	UjEsemeny	/Events/NewForm.aspx	

3. ábra  
Egy példa részlet az Clementine WebMining Node EventDefinition fájlból

#### 4. Általános statisztikák

Természetesen a mindkét rendszerben megtalálhatók a leggyakrabban hivatkozott statisztikák, elemzések, melyek a látogatások gyakoriságára, hosszára, eloszlására, hálózati és platform specifikus jellemzőkre, webhely tartalomra, belépési, kilépési, és bejárési utakra, stb. vonatkoznak.

Mindkét rendszer támogatja az elemzés tárgyát képező időszak megadását, módosítását. Az eredményeket lementhetjük külön fájlként, de megtekinthetjük akár grafikonok, diagramok formájában is.

Amit kiemelnék, mint előnyét a Google Analytics rendszernek az a következő néhány hasznos funkció:

- A nagyon kényelmes időintervallum állítás, valamint az, hogy közvetlen módon két tetszőleges időintervallum is összehasonlítható.
- A Map Overlay funkció, amely földrajzi térképen ábrázolva mutatja a látogatók elhelyezkedését.
- Languages menüpont, amely nyelvi környezetek szerint csoportosítja a látogatókat.
- Connection Speeds menüpont, a látogatók sávszélessége alapján történő megoszlás.
- Traffic Sources menüpont, mely alatt az oldalunk elérésének módjairól kapunk részletes adatokat. Direkt elérés vagy keresőprogramokból történő elérés, ha kereső program, akkor milyen kulcsszavakra kerestek, stb.
- Szegmentálási lehetőség: bármely lekért statisztika részadatait tovább bonthatjuk egy tetszőlegesen megadott szempont szerint. Például a szombathely.hu-ról érkező látogatókat további csoportokra bonthatjuk a használt böngészőprogram szerint.

A Clementine WebMining beépített streamjei noha talán kicsit több tanulást és szakértelmet kívánnak, rendelkeznek azzal a nagy előnnyel, hogy átalakíthatók és bővíthetők igény szerint.

A két rendszer közti különbségek forrása alapvetően a rendelkezésre álló adatok másságából ered. Például: a webszerver ugyan naplózhatja a látogatók IP címeit, de ha nincs ismeretünk az IP címosztályok földrajzi elhelyezkedéséről, vagy az IP címekhez rendelt domain nevekről akkor abból földrajzi elhelyezkedést számolni nagyon nehéz.

Ha a webhelyünk, vagy annak csak egy része felhasználói hitelesítés után érhető csak el, akkor a webszerver naplóállományaiból a Clementine WebMining képes visszakövetni a kattintást végző felhasználó azonosítóját. Ez lehetőséget ad arra, hogy további – a felhasználóra vonatkozó – adatokat vonjunk be az elemzésekbe. Felhasználóinkról számos előismeretünk lehet, akár a regisztrációkor megadott demográfiai adatok formájában, de rendelkezésünkre állhat természetesen a felhasználó eddigi ténykedésével kapcsolatos adathalmaz is. Milyen sűrűn látogatja oldalainkat, mit vásárolt eddig, milyen hozzászólásokat írt, stb. Ezen adatoknak az elemzésbe történő bevonására a Google rendszere alkalmatlan, míg az SPSS rendszerének használata esetén csak itt kezd igazán izgalmassá válni az adatbányászat.

## **5. Összetettebb elemzések**

Ide azokat az elemzéseket sorolnám, amelyek túlmutatnak a sima oldalletöltések különböző paramétereinek statisztikai elemzésén. Például olyan kérdések merülhetnek fel, hogy milyen a látogatók lojalitása, vagy miért és hol szakad meg egy a látogatóktól elvárt „megtérés”, cél. Például ilyen cél lehet, hogy a web áruházban történő tallózás után tényleges vásárlás is történjen.

A Google Analytics-ben is találunk olyan szolgáltatást, amely ezekre a kérdésekre próbál fényt deríteni, de közel sem olyan szinten paraméterezhető és sokoldalú, mint amit az SPSS rendszere nyújt. A WebMining beépített streamjei nagyon gazdag tárházát kínálják ezen típusú szolgáltatásoknak.

## **Összefoglaló**

Összességében láthatjuk, hogy a Google megoldása nagyon kényelmes, felhasználóbarát és látványos, valamint a beépített eszközei jól megtervezettek és valóban hasznosak. Legnagyobb hátránya is pont ez. Tehát arra jó, amit tud, és nem többre. Ha olyan statisztikára van szükségünk, amely nem szerepel a beépített lehetőségeiben, akkor használhatatlan.

A weblapok látogatottságának elemzése során a leghasznosabbak és legfontosabbak pont az olyan statisztikák lehetnek, amelyek speciálisak, függenek a webhely felépítésétől, valamint a szolgáltatásaitól. A Google Analytics általános statisztikáin túl egyéni igényeket kiszolgáló statisztikák nem kérdezhetők le. Ezek elkészítését viszont nagyszerűen meg lehet oldani a Clementine alkalmazásával.

Végző soron a Clementine-ra azoknak van szüksége, akiknél a lapletöltések száma meghaladja a Google Analytics beépített korlátját, vagy egyéni igényeket kiszolgáló statisztikákra van szükségük, és erre tudnak áldozni időt, pénzt, és szakértelmet.

## **Irodalomjegyzék**

- [1] Google Analytics: <http://www.google.com/analytics/>
- [2] W3C - Extended Log File Format: <http://www.w3.org/TR/WD-logfile.html>
- [3] SPSS Webminig for Clementine: [http://www.spss.com/web\\_mining\\_for\\_clementine/](http://www.spss.com/web_mining_for_clementine/)