



This material is partially supported by the Romanian National University Research Council under award PN-II IDEI 2412/2009.

Why Google Analytics Cannot Be Used For Educational Web Content

Author: Sanda-Maria Dragoş

Web Analytics

- ▶ the **measurement, collection, analysis** and **reporting** of Internet data for the purposes of understanding and optimizing web usage [WAA]
- ▶ Main bodies defining web analytics metrics:
 - ▶ **Jicwebs** (Industry Committee for Web Standards)/**ABCe** (Auditing Bureau of Circulations electronic, UK and Europe)
 - ▶ **WAA** (Web Analytics Association, US)
 - ▶ **IAB** (Interactive Advertising Bureau).

Web Analytics metrics

▶ Building block terms

- ▶ Page view / Page impression / Page request
- ▶ Visit / Session
- ▶ (Unique) Visitor / Unique Browser / User

▶ Visit characterization terms

- ▶ Bounce / Single page Visits
- ▶ Visit Duration
- ▶ Page Views per Visit / Visit Depth

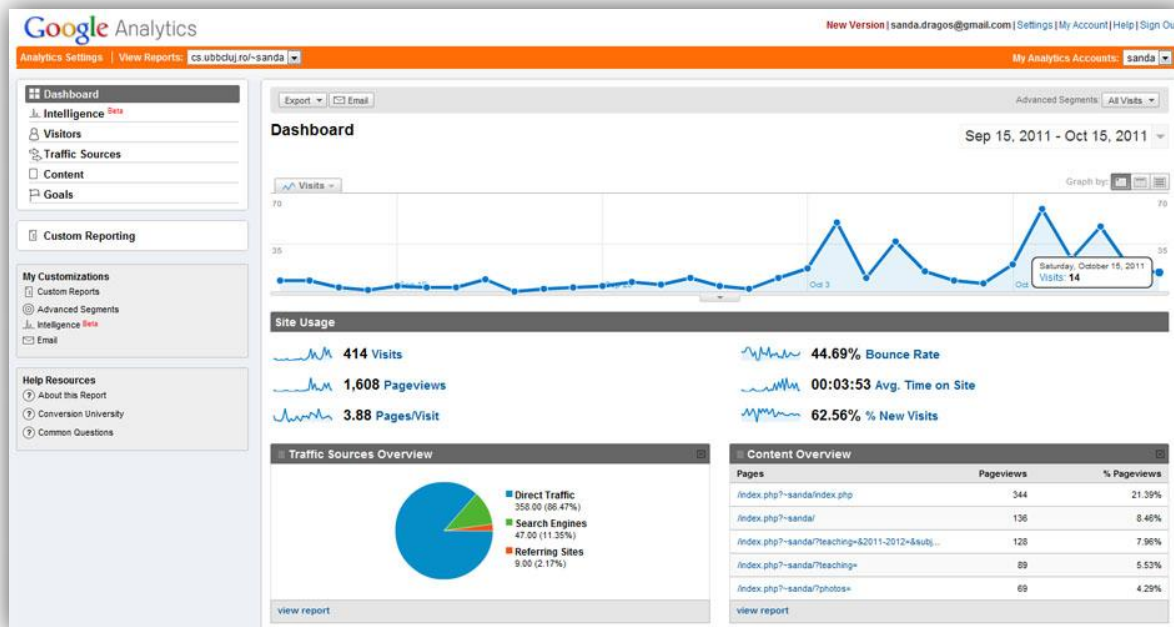
▶ Visitor characterization terms

- ▶ Frequency / Visits per Unique Visitors
- ▶ Recency
- ▶ Repeat Visitor / Repeat Unique Browser

Google Analytics

- ▶ is a free web analytics instrument offered by Google
- ▶ the most widely used web analytics instrument
- **W3Techs - World Wide Web Technology Surveys:**

“Google Analytics is used by 81.6% of all the websites whose traffic analysis tool we know. This is 53.5% of all websites.”



Limitations of page tagging:

- Blocking JavaScript code
- Deleting or blocking cookies

Web Analytics. Main Challenges

▶ **Unique user identification**

- ▶ Based on IP + UA
 - Multiple IP addresses - Single Visitor
 - Multiple User Agents - Single Visitor
- ▶ Based on cookies
- ▶ Based on registered user account information

▶ **Visit/Session identification**

- ▶ HTTP protocol is stateless and connectionless
- ▶ 3 main heuristics to determine the visit termination
 - restricts the duration of the entire visit to a predefined upper bound (temporal)
 - limits the time spent on any page to a threshold (temporal)
 - all pages have to be linked directly or indirectly (navigational pattern)

Integrated versus third party analysis

- ▶ **WATEC** (Web Analytics Tool for Educational Content)
 - ▶ integrated with an e-learning system called **PULSE** (Php Utility used in Laboratories for Student Evaluation)



The screenshot shows the PULSE login page. It has a header with the PULSE logo and 'made by Dr. Sanda Maria Dragoș'. Below the header is a 'Parent page' link. The main content area says 'Use your UBBCLUJ.RO login account and password.' and contains a login form with 'Login' and 'Password' fields and a 'Log in' button. At the bottom, it says 'Last updated : January 01, 1970' and has links for 'Feedback' and 'Contact'.

- ▶ **Google Analytics-Like (GAL)** instrument
 - ▶ uses cookies in order to identify unique visitors

WATEC versus GAL

- ▶ **Data collection**

- ▶ **Unique visitors: WATEC** - login IDs, **GAL** – cookies

- ▶ **WATEC visits**

- ▶ all web pages accessed consecutively from the **same IP** and **User Agent (UA)** by a **WATEC** visitor, **before login** and **until logout** or **closed browser**.

- ▶ **GAL visits**

- ▶ all pages accessed consecutively with the **same cookie ID** that have the **time-on-page less than 30 minutes**

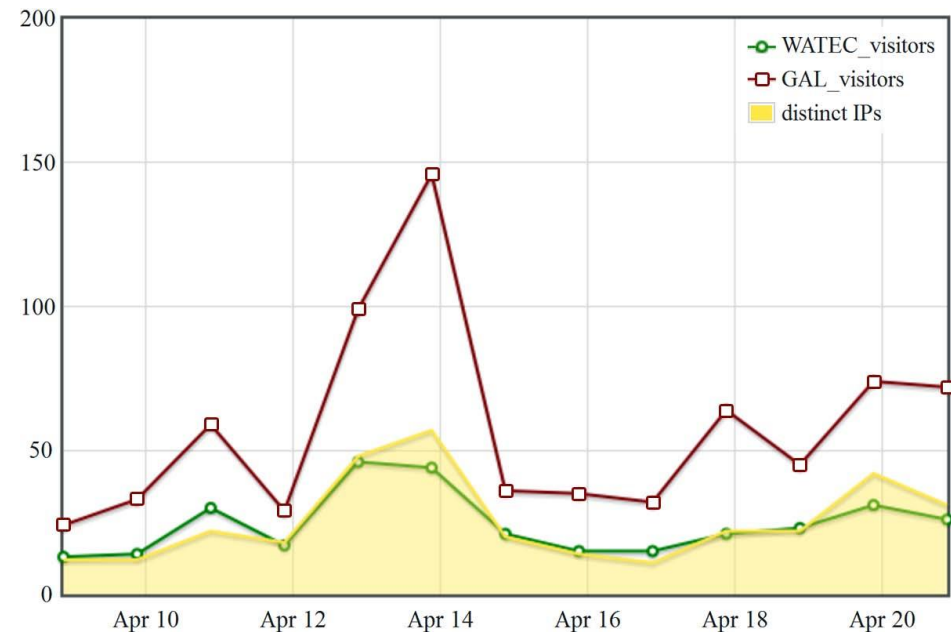
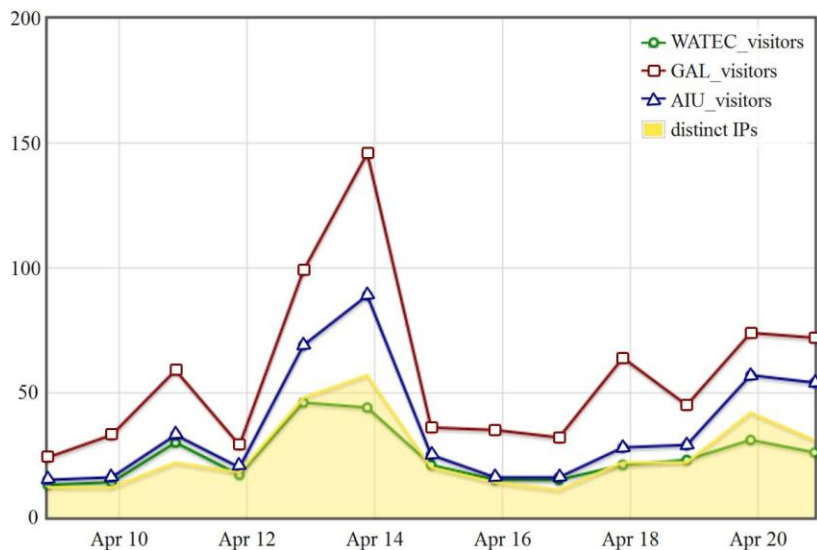
Test Results

- ▶ Most web analytics instruments consider that the visit ends after 30 minutes of inactivity

IP	UA	acc. date	time	time-on-page	referral page URL	access page URL	login	cookie
108	1385	2011-04-14	19:21:01	2"	PULSE.php?2010-2011/SO1&cursuri&SO_curs4.other	PULSE.php?2010-2011/SO1&cursuri	bpri0034	1951
108	1385	2011-04-14	19:21:03	1' 47"	PULSE.php?2010-2011/SO1&cursuri	PULSE.php?2010-2011/SO1&cursuri&SO_curs3.other	bpri0034	1951
108	1385	2011-04-14	19:22:50	36' 23"	PULSE.php?2010-2011/SO1&cursuri&SO_curs4.other	PULSE.php?2010-2011/SO1&cursuri	bpri0034	1951
				3' 5" - GAL Time				
108	1385	2011-04-14	19:59:13	'	PULSE.php?2010-2011/SO1&cursuri	PULSE.php?2010-2011/SO1&cursuri&SO_curs3.other	bpri0034	1951
				0" - GAL Time - BOUNCE -				
				39' 28" - WATEC Close				
108	1385	2011-04-15	00:14:38	35"	PULSE.php?2010-2011/SO1&cursuri&SO_curs4.other	PULSE.php?2010-2011/SO1&cursuri		2007
108	1385	2011-04-15	00:15:13	6"	PULSE.php?2010-2011/SO1&cursuri	PULSE.php	bpri0034	2007

WATEC's and GAL's *visitors*

- ▶ GAL's difficulty to accurately identify *visitors*:
 - ▶ one user – multiple locations
 - ▶ one user – multiple browsers
 - ▶ deleted or denied cookies



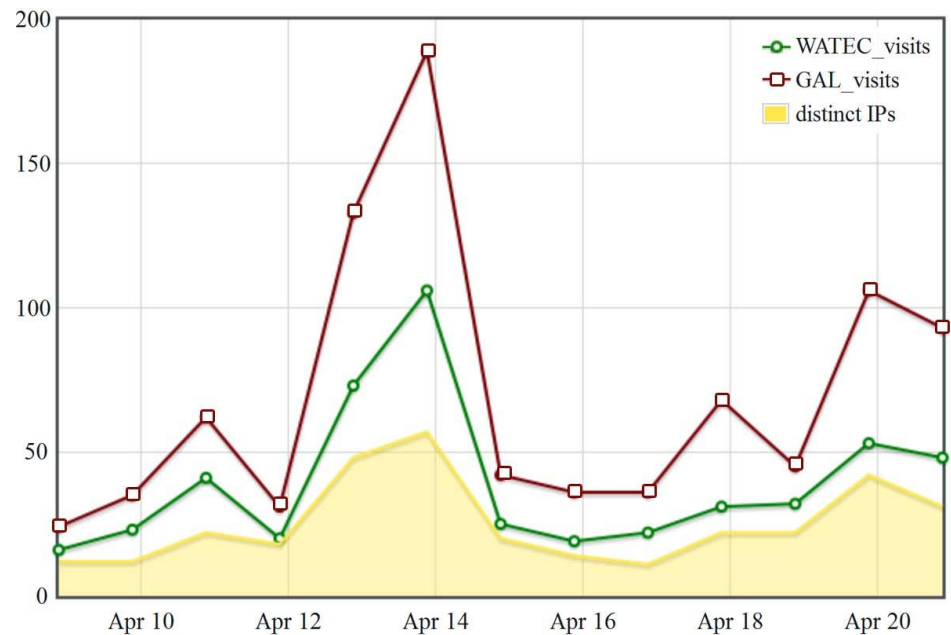
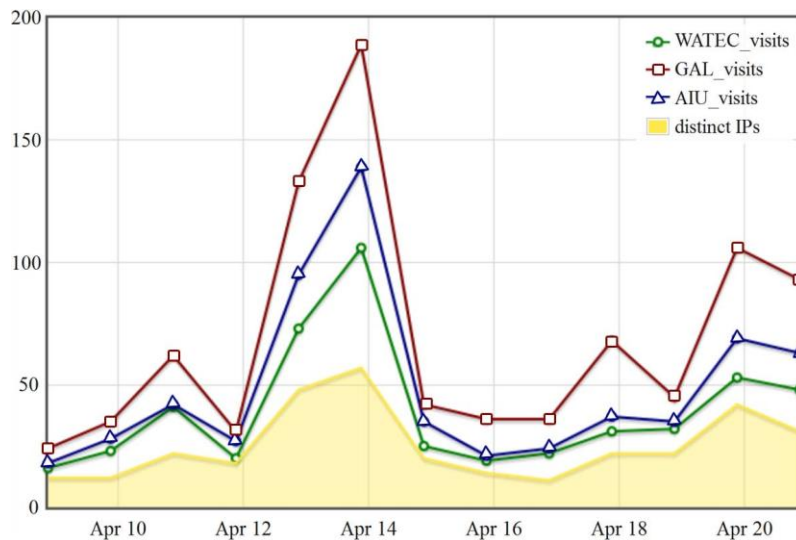
WATEC's and GAL's *visits*

▶ GAL's more *visits*

▶ time-on-page $\leq$ 30 minutes $\Rightarrow$ visit fragmentation

▶ WATEC's less *visits*

▶ *does not consider non-autheticated visits*

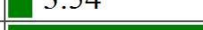
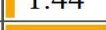


WATEC's and GAL's metric values

		WATEC	GAL
Building Block Terms	PAGE VIEWS	5663	5663
	VISITS	509	900
	VISITORS	67	740
Visit Characterization Terms	BOUNCE	0 %	25.25 %
	VISIT DEPTH = PAGE VIEWS PER VISITS	11.13	6.29
	VISIT DURATION	2h 8' 35"	6' 54"
Visitor Characterization Terms	FREQUENCY = VISITS PER VISITOR	7.6	1.22
	RECENCY	1.71 days	0.01 days
	REPEATED VISITORS	86.84 %	17.78 %

WATEC's and GAL's visit characterization

► Visit depth

Pageviews in the visit	Visits with this many pageviews		Percentage of all visits	
	WATEC	GAL	WATEC	GAL
1 pageviews		304	0	 33.78
2 pageviews	4	40	 0.79	 4.44
3 pageviews	37	64	 7.27	 7.11
4 pageviews	42	56	 8.25	 6.22
5 pageviews	73	94	 14.34	 10.44
6 pageviews	59	64	 11.59	 7.11
7 pageviews	40	37	 7.86	 4.11
8 pageviews	28	35	 5.5	 3.89
9 pageviews	29	26	 5.7	 2.89
10 pageviews	26	25	 5.11	 2.78
11 pageviews	18	13	 3.54	 1.44
12 + pageviews	153	142	 30.06	 15.78

► Visit duration

Duration of visit	Visits with this duration		Percentage of all visits	
	WATEC	GAL	WATEC	GAL
0-10 seconds	15	344	 2.95	 38.22
11-30 seconds	72	120	 14.15	 13.33
31-60 seconds	48	59	 9.43	 6.56
61-180 seconds	89	112	 17.49	 12.44
181-600 seconds	69	102	 13.56	 11.33
601-1800 seconds	85	101	 16.7	 11.22
1801 + seconds	131	62	 25.74	 6.89

WATEC's and GAL's visitor characterization

► Frequency

Count of visits from this visitor	Visits that were the visitor's nth visit		Percentage of all visits	
	WATEC	GAL	WATEC	GAL
1 times	67	740	 13.16	 82.22
2 times	61	127	 11.98	 14.11
3 times	56	25	 11	 2.78
4 times	50	4	 9.82	 0.44
5 times	44	1	 8.64	0.11
6 times	41	1	 8.06	0.11
7 times	38	1	 7.47	0.11
8 times	28	1	 5.5	0.11
9 times	23		 4.52	0
10 + times	101		 19.84	0

► Recency

Previous visit was tracked	Total visits by period		Percentage of all visits	
	WATEC	GAL	WATEC	GAL
First visit	67	740	 13.16	 82.22
Same day	205	157	 40.28	 17.44
1 days ago	71	1	 13.95	0.11
2 days ago	30	1	 5.89	0.11
3 days ago	29		 5.7	0
4 days ago	15		 2.95	0
5 days ago	21	1	 4.13	0.11
6 days ago	43		 8.45	0
7 days ago	16		 3.14	0
8-14 days ago	12		 2.36	0

Thank you for listening!

Questions?

