

# Περιγραφή Ύλης για το Μάθημα Αξιολόγηση Διαδικτυακών Συστημάτων

## Στόχοι μαθήματος:

- Να γίνουν κατανοητές οι έννοιες που χρησιμοποιούνται στο Web Analytics.
- Να παρουσιαστούν τα σημεία ενός διαδικτυακού συστήματος που δημιουργούν αξιοποιήσιμες πληροφορίες για τον διαχειριστή.
- Να παρουσιαστούν τα κατάλληλα εργαλεία για την κάθε περίπτωση μετρήσεων.

## Τι μας παρέχουν τα Συστήματα Αξιολόγησης:

1. Εικόνα για τον τρόπο που επισκέπτονται χρήστες ένα διαδικτυακό σύστημα (ΔΣ).
2. Στοιχεία για τον φόρτο και τους χρόνους απόκρισης
3. Γεωγραφικές πληροφορίες επισκεπτών
4. Πληροφορίες για το περιβάλλον των υπολογιστικών τους συστημάτων
4. Παρακολούθηση αποτελεσμάτων από καμπάνιες
5. Αναγνώριση referrers (sites που οδηγούν στο ΔΣ μας)
6. Δυνατότητα ομαδοποίησης επισκεπτών και ανάλυσης των συμπεριφορών τους
7. Μέτρηση επιρροής των ΜΚΔ στην επισκεψιμότητα ενός ΔΣ
8. Εύρεση λαθών στο εξεταζόμενο ΔΣ
9. Δυνατότητα παρουσίασης μετρήσεων κίνησης στον χρόνο: ανά ώρα/ημέρα/μήνα
10. Παρουσίαση μετρήσεων σε σχεδόν πραγματικό χρόνο (near real time).

*Τα συστήματα Web Analytics πρέπει να θεωρούνται τμήμα ή λειτουργικό μέρος κάθε συστήματος e-commerce*

## Βασικά χαρακτηριστικά των διαδικτυακών συστημάτων:

Περιβάλλον κατασκευής:

- γλώσσες προγραμματισμού,
- τεχνικές κατασκευής,
- χρησιμοποιούμενος Web Server,
- βάσεις δεδομένων για persistence της πληροφορίας.

## Τέσσερα μοντέλα συστημάτων Analytics:

- Log File Analysis Systems,
- Tagged Systems,
- Network Data Collection Devices και
- Hybrid Systems.

Table 1

Page Tagging	v	Logfile Analysis
<b>Advantages</b> <ul style="list-style-type: none"><li>• Breaks through proxy and caching servers<ul style="list-style-type: none"><li>- provides more accurate session tracking</li></ul></li><li>• Track client side events<ul style="list-style-type: none"><li>- JavaScript, Flash, web v2.0</li></ul></li><li>• Client-side capture of e-commerce data<ul style="list-style-type: none"><li>- server-side access can be problematic</li></ul></li><li>• Visitor data can be collected/processed in near real-time</li><li>• Program updates performed for you</li><li>• Data storage and archiving performed for you</li></ul>		<b>Advantages</b> <ul style="list-style-type: none"><li>• Historical data can be reprocessed easily</li><li>• No Firewall issues to worry about</li><li>• Can track bandwidth and completed downloads<ul style="list-style-type: none"><li>- also differentiate completed and partial downloads</li></ul></li><li>• Track search engine spiders/robots by default</li><li>• Track mobile visitors by default</li></ul>
<b>Disadvantages</b> <ul style="list-style-type: none"><li>• Setup errors lead to data loss<ul style="list-style-type: none"><li>- If you make a mistake with your tags, data is lost and you can not go back and re-analyse</li></ul></li><li>• Firewalls<ul style="list-style-type: none"><li>- can mangle or restrict tags</li></ul></li><li>• Cannot track bandwidth or completed downloads<ul style="list-style-type: none"><li>- Tags are set when the page/file is requested <u>not</u> when the download is complete</li></ul></li><li>• Cannot track search engine spiders<ul style="list-style-type: none"><li>- robots ignore page tags</li></ul></li></ul>		<b>Disadvantages</b> <ul style="list-style-type: none"><li>• Proxy/caching inaccuracies<ul style="list-style-type: none"><li>- If a page is cached, no record is logged on your web server</li></ul></li><li>• No event tracking (javascript, Flash, web v2.0)<ul style="list-style-type: none"><li>- no Javascript, flash, web v2.0 tracking</li></ul></li><li>• Program updates performed by your own team</li><li>• Data storage and archiving performed by your own team</li></ul>

### Σύγκριση Συστημάτων Page Tagging και Logfile Analysis (Brian Clifton)

## Βασικές έννοιες Analytics:

(<http://www.analyticsmarket.com/blog/google-analytics-definitions>)

### Hit

Με το Google Analytics, σαν hit θεωρείται η κάθε αίτηση που αποστέλλεται στο σύστημα συλλογής δεδομένων GA. Αυτό περιλαμβάνει προβολές, εκδηλώσεις, προσαρμοσμένες μεταβλητές κλπ.

Ο χρήστης μπορεί να κρατήσει αντίγραφο των επισκέψεων που αποστέλλονται στην Google και να τις επεξεργαστεί με ειδικό λογισμικό ανάλυσης. Έτσι μπορεί να δει τα clickpaths κάθε επισκέπτη, ονόματα χρηστών, και άλλα στοιχεία από τα δεδομένα στο Google Analytics.

### Pageview

Μια προβολή σελίδας καταγράφεται κάθε φορά που μια σελίδα προβάλλεται. Ή, πιο τεχνικά, μια προβολή σελίδας καταγράφεται κάθε φορά που η μέθοδος παρακολούθησης προβολής του Google Analytics εκτελείται. Όταν ένας επισκέπτης πατήσει το κουμπί back, μια προβολή σελίδας καταγράφεται. Όταν ένας επισκέπτης ζητήσει ανανέωση (refresh), μια προβολή σελίδας καταγράφεται. Κάθε φορά που μια σελίδα ανοίγει στο πρόγραμμα περιήγησης, ανεξάρτητα από το αν έχει γίνει προσωρινή αποθήκευση (cache), μια προβολή σελίδας καταγράφεται.

### Visit

Μια επίσκεψη αποτελείται από σειρά σελίδων που ένας απλός επισκέπτης ζητάει κατά τη διάρκεια μιας περιόδου δραστηριότητας. Μια επίσκεψη τελειώνει αφού ο επισκέπτης είτε κλείνει το πρόγραμμα περιήγησης, καθαρίζει τα cookies, ή είναι ανενεργός για 30 λεπτά. (Το μήκος χρονικού ορίου προσαρμόζεται ανάλογα με τις ρυθμίσεις παρακολούθησης)

### Visitors

Οι επισκέπτες ορίζονται από ένα μοναδικό αναγνωριστικό το οποίο συνήθως αποθηκεύεται στα cookies του επισκέπτη. Κάθε φορά που εκτελείται ο κώδικας παρακολούθησης, ψάχνει για τα cookies στον browser που τέθηκαν από τον τρέχοντα τομέα. Εάν δεν μπορούν να βρεθούν, τίθενται νέα cookies με νέο ID. Το Google Analytics δίνει μεγαλύτερη έμφαση στις επισκέψεις παρά στους σε επισκέπτες, λόγω των πολλών εγγενών ανακριβειών εντοπισμού μεμονωμένων χρηστών. Για παράδειγμα, ένας επισκέπτης που διαγράφει τα cookies του, χρησιμοποιεί πολλαπλά προγράμματα περιήγησης ή μοιράζεται τον υπολογιστή τους θα εμφανιστεί λανθασμένα.

## **Bounce**

Μια επίσκεψη με μία προβολή ιστοσελίδας. Δεν έχει σημασία πόσο καιρό ο επισκέπτης ήταν στην πρώτη σελίδα ή πώς έφυγε. Τεχνικά, θεωρείται σαν μια επίσκεψη με μία μόνο αλληλεπίδραση.

## **Time on Page**

Ο χρόνος στη σελίδα μετριέται με την αφαίρεση του χρόνου που ένας επισκέπτης χτύπησε μια σελίδα από τη στιγμή που θα χτυπήσει την επόμενη σελίδα. (Π.χ. αν χτυπήσει την σελίδα 1 στις 12:00 και χτύπησε την σελίδα 2 στις 12:03, ο χρόνος στην σελίδα 1 θεωρείται ότι είναι τρία λεπτά.) Αυτό σημαίνει ότι ο χρόνος στη σελίδα για την τελευταία σελίδα επίσκεψης είναι πάντα μηδέν επειδή το Google Analytics δεν παρακολουθεί το κλείσιμο σελίδας.

## **Time on Site**

Χρόνος στον ιστότοπο θεωρείται το άθροισμα των χρόνων όλων των σελίδων σε μια επίσκεψη. Ακριβέστερα, είναι η διαφορά μεταξύ του χρόνου που φόρτωσης της πρώτης σελίδας και της τελευταίας σελίδα σε μια επίσκεψη. Να σημειωθεί ότι την προβολή σελίδων σε διαφορετικά tabs του browser δεν επηρεάζει τον χρόνο. Το Google Analytics βλέπει απλώς μια σειρά από σελίδες που ζητήθηκαν σε χρονολογική σειρά, χωρίς καμία αναφορά σε πολλαπλές καρτέλες ή παράθυρα.

## **New Visitor**

Νέος θεωρείται ο επισκέπτης που δεν έχει cookies του Google Analytics, όταν επισκέφτηκε την πρώτη σελίδα σε αυτήν την επίσκεψη. Εάν κάποιος επισκέπτης διαγράφει τα cookies του και ζητήσει ξανά μία ιστοσελίδα, ο επισκέπτης θα υπολογίζεται ως νέος επισκέπτης.

## **Returning Visitor**

Ο επισκέπτης με υπάρχοντα cookies του Google Analytics από μια προηγούμενη επίσκεψη.

## **Dollar Index**

Μέτρηση που δηλώνει την πόσο πολύτιμη είναι μία ιστοσελίδα του ιστότοπου. Όσο μεγαλύτερη η μέτρηση τόσο συχνότερα έχει ζητηθεί από τους επισκέπτες.

## **Pages/Visit**

Σελίδες ανά επίσκεψη: Προβολές σελίδας δια επισκέψεις. Αυτή η μέτρηση δείχνει το μέσο αριθμό των σελίδων που προβάλλονται ανά επίσκεψη.

## **Direct Traffic**

Ιδανικά, αυτή είναι η κίνηση που ήρθε σε ένα ιστότοπο απευθείας μέσω σελιδοδείκτη ή πληκτρολογώντας τη διεύθυνση URL. Στην πραγματικότητα, είναι η κίνηση για την οποία ο κώδικας δεν μπόρεσε να καθορίσει την πηγή προέλευσης. Ανάλογα με την περιοχή και το πρόγραμμα περιήγησης, κάποια links μπορεί να μην παρουσιάζουν παραπομπές και έτσι θα μπορούσαν να κατηγοριοποιηθούν ως άμεσες.

## **Referring Sites**

Εδώ έχουμε αναγνώριση του referrer και ο referrer δεν είναι μηχανή αναζήτησης. Επίσης δεν υπάρχουν μεταβλητές καμπάνιας στην επίσκεψη.

## **Search Engine Traffic**

Το Google Analytics κατηγοριοποιεί αυτόματα την κυκλοφορία ως προερχόμενη από μηχανή αναζήτησης, εάν η διεύθυνση URL ανήκει στον κατάλογο των γνωστών μηχανών αναζήτησης και υπάρχει ένας όρος αναζήτησης που προσδιορίζεται σε αυτήν τη διεύθυνση URL. Οι οργανικές αναζητήσεις και οι επισκέψεις μέσω μηχανών αναζήτησης ανήκουν σε αυτήν την ομάδα.

## **Event Tracking**

Ένα εργαλείο που επιτρέπει να παρακολουθεί κανείς τις δραστηριότητες των επισκεπτών ξεχωριστά από προβολές σελίδων. Χρησιμοποιείται συνήθως για την παρακολούθηση της αλληλεπίδρασης με AJAX ή Flash περιεχόμενο.

## **Google Analytics API**

Το API αντλεί δεδομένα από τους λογαριασμούς του Google Analytics. Δίνει στους πελάτες προγραμματιστικά πρόσβαση για εξαγωγή δεδομένων του Google Analytics έτσι ώστε να το ενσωματώσουν σε δικές τους εφαρμογές ή/και βάσεις δεδομένων.

## Το σύστημα AWStats

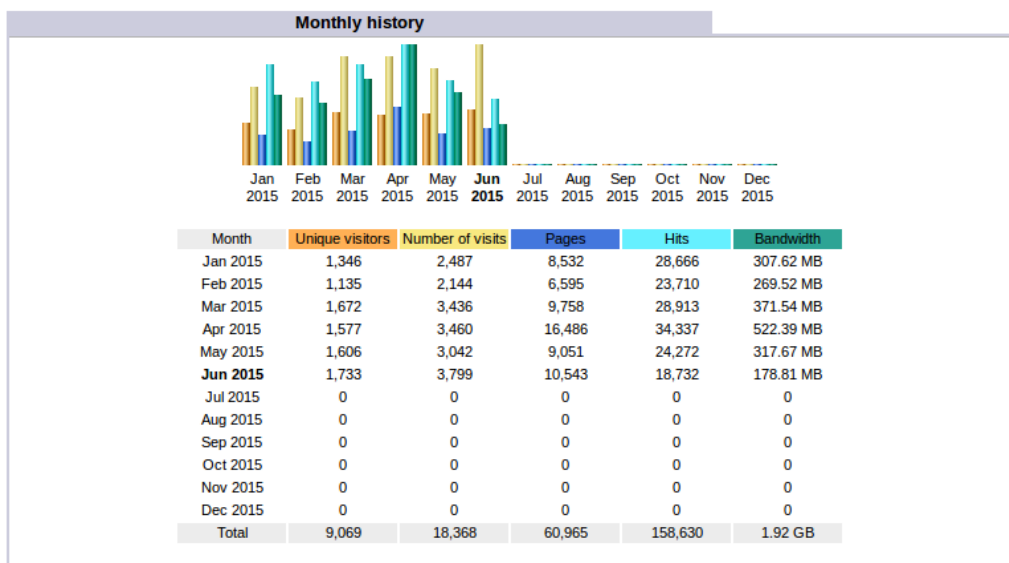
Το AWStats είναι ένα εργαλείο, γραμμένο στην γλώσσα PERL, σύστημα ελεύθερου λογισμικού, με πολλά ενδιαφέροντα χαρακτηριστικά. Δημιουργεί ιστοσελίδες για μέτρηση ιστοσελίδων, streaming, ftp, διακομιστή αλληλογραφίας και δίνει στατιστικά στοιχεία και γραφικά.

Λειτουργεί Common Gateway Interface (CGI) ή από την γραμμή εντολών και σας δίνει πολλές πληροφορίες καταγραφής.

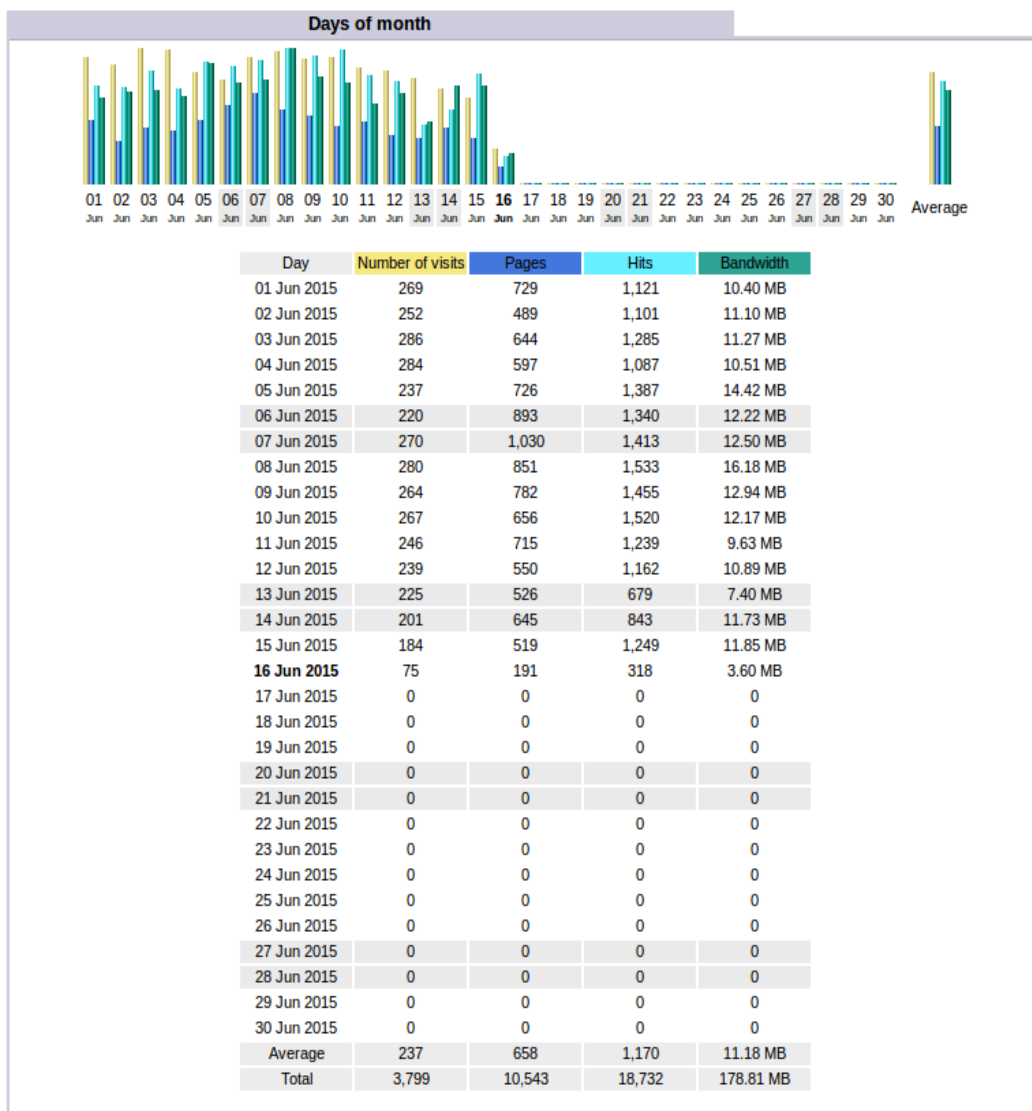
Ενδεικτικό δείγμα των μετρήσεων που παρέχει το AWStats:



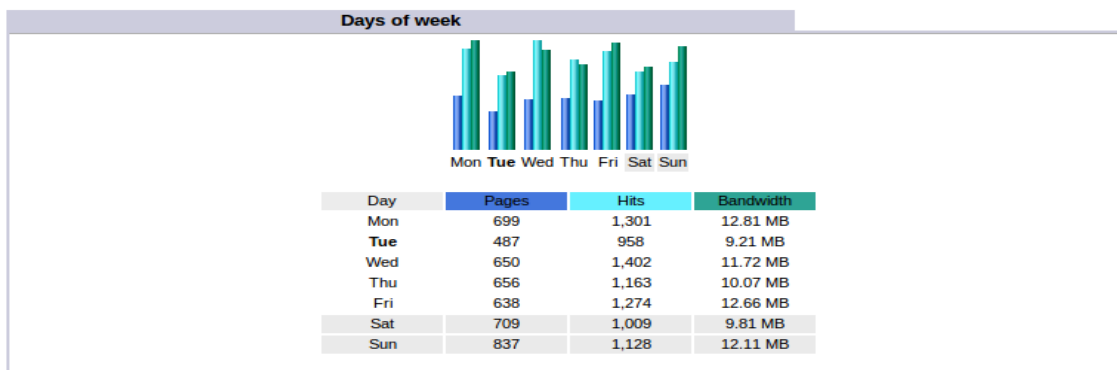
### Περίληψη



### Μηνιαίο Ιστορικό

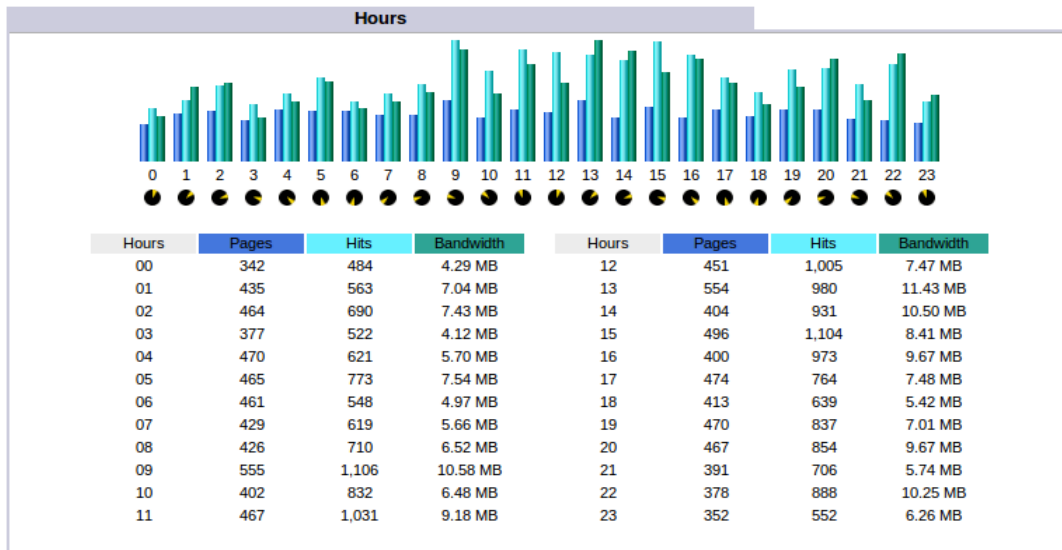


### Περίληψη ανά ημέρα του μήνα



### Περίληψη ανά ημέρα της εβδομάδας





*Περίληψη ανά ώρα του 24ώρου*

**Countries (Top 10) - Full list**

Countries	Pages	Hits	Bandwidth
Unknown	4,959	6,188	44.88 MB
United States	2,076	2,985	34.03 MB
France	1,156	4,656	37.62 MB
China	433	724	9.25 MB
Netherlands	264	305	6.41 MB
Ukraine	256	262	6.06 MB
Turkey	189	189	7.73 MB
Romania	176	217	1.19 MB
Great Britain	142	143	1.04 MB
Poland	141	276	1.69 MB
Others	751	2787	28.90 MB

*Στοιχεία περιόδου ανά χώρα επισκέπτη*

**Hosts (Top 10) - Full list - Last visit - Unresolved IP Address**

Hosts : 0 Known, 1,803 Unknown (unresolved ip) 1,733 Unique visitors	GeoIP Country	GeoIP City	GeoIP Region	Info.	Pages	Hits	Bandwidth	Last visit
5.135.199.82	Unknown	Unknown	Unknown	?	1,446	1,446	6.22 MB	15 Jun 2015 - 08:04
54.164.192.241	United States	Unknown	New Jersey	?	1,123	1,124	20.96 MB	16 Jun 2015 - 06:24
80.246.188.132	Netherlands	Catanzaro	Unknown	?	247	247	6.09 MB	09 Jun 2015 - 22:45
94.73.159.122	Turkey	Unknown	Unknown	?	184	184	7.64 MB	16 Jun 2015 - 06:05
62.210.83.225	France	Paris		?	156	156	622.00 KB	13 Jun 2015 - 14:52
195.154.182.101	France	Montreuil		?	147	147	689.31 KB	11 Jun 2015 - 11:35
46.105.100.183	Unknown	Unknown	Unknown	?	131	131	2.40 MB	14 Jun 2015 - 13:18
52.5.121.15	United States	Wilmington	Delaware	?	105	105	1.89 MB	02 Jun 2015 - 04:35
198.50.159.225	Unknown	Unknown	Unknown	?	86	86	342.91 KB	14 Jun 2015 - 10:32
80.82.65.82	Switzerland	Herisau		?	85	85	1.68 MB	15 Jun 2015 - 13:43
Others					6,833	15,021	130.32 MB	

*Οι πιο ενεργοί επισκέπτες*

Το σύστημα AWS παρέχει ακόμα περισσότερα στατιστικά. Τα στατιστικά μπορούν να φορτώνονται από ιστοσελίδα ή να εμφανίζονται τοπικά από εφαρμογή μέσω command line.

Παρεμφερή προγράμματα είναι:

το **Webalyzer**, γραμμένο σε C (<https://webalizer.net/>) το οποίο είναι ελεύθερο λογισμικό

και

το **Sawmill Analytics**, επίσης σε C, (<https://www.sawmill.net/>) το οποίο δεν είναι μεν ελεύθερο λογισμικό, έχει όμως σχετικά χαμηλή τιμή.

## Προγραμματιστικές μεθοδολογίες μέτρησης χρόνου.

Java κώδικας:

```
long time;
...
time = System.nanoTime();
methodToTime();
System.out.println(System.nanoTime() - time);
```

Η χρήση της βιβλιοθήκης Joda-Time (<https://www.joda.org/joda-time/>) διευκολύνει τους υπολογισμούς και τις συγκρίσεις μεταξύ ημερομηνιών, ωρών, λεπτών κλπ. Η βιβλιοθήκη Joda Time βρίσκεται πλέον εν μέρει ενσωματωμένη στην Java 8 SE.

```
public boolean isAfterPayDay(DateTime datetime) {
    if (datetime.getMonthOfYear() == 2) { // February is month 2!!
        return datetime.getDayOfMonth() > 26;
    }
    return datetime.getDayOfMonth() > 28;
}

public Days daysToNewYear(LocalDate fromDate) {
    LocalDate newYear = fromDate.plusYears(1).withDayOfYear(1);
    return Days.daysBetween(fromDate, newYear);
}

public boolean isRentalOverdue(DateTime datetimeRented) {
    Period rentalPeriod = new Period().withDays(2).withHours(12);
    return datetimeRented.plus(rentalPeriod).isBeforeNow();
}

public String getBirthMonthText(LocalDate dateOfBirth) {
    return dateOfBirth.monthOfYear().getAsText(Locale.ENGLISH);
}
```

Η βιβλιοθήκη Joda-Time θεωρείται μεν ξεπερασμένη επειδή έχει σταματήσει να εξελίσσεται αλλά και επειδή είναι εν μέρει ενσωματωμένη στις νέες εκδόσεις της Java. Περιέχει όμως μεθόδους και κλάσεις οι οποίες δεν έχουν ενσωματωθεί στις νέες εκδόσεις της Java.

Υπάρχει η βιβλιοθήκη επεξεργασίας χρόνου και ημερομηνιών Time4J. Πρόκειται για ελεύθερο λογισμικό (<https://github.com/MenoData/Time4J>) που καλύπτει τα κενά του JSR-310.

## Εξωτερικά συστήματα μέτρησης

### JMeter

Η εφαρμογή Apache JMeter είναι λογισμικό ανοικτού κώδικα και 100% καθαρά εφαρμογή Java. Είναι σχεδιασμένη για να δημιουργεί δοκιμαστικά τεχνητό φόρτο έτσι ώστε να μετρά την απόδοση εφαρμογών λογισμικού. Αρχικά είχε σχεδιαστεί για τη δοκιμή διαδικτυακών εφαρμογών, αλλά έχει επεκταθεί και σε άλλες λειτουργίες ελέγχου.

Μπορεί να χρησιμοποιηθεί για τον έλεγχο των επιδόσεων τόσο σε στατικές και όσο και σε δυναμικές εφαρμογές (Webservices (SOAP / REST), δυναμικές γλώσσες - PHP, Java, ASP.NET, αρχεία, κλπ. Java αντικείμενα, Βάσεις Δεδομένων και ερωτήματα, FTP Servers και πολλές άλλες εφαρμογές.

Μπορεί να χρησιμοποιηθεί για την προσομοίωση βαρέως φόρτου σε διακομιστή, ή ομάδες από servers, το δίκτυο ή το αντικείμενο να δοκιμάσει την αντοχή του ή να αναλύσει τη συνολική απόδοση κάτω από διάφορους τύπους φορτίου.

Δημιουργεί και γραφικές αναλύσεις των επιδόσεων των διακομιστών.

Τα χαρακτηριστικά του:

Δυνατότητα φόρτωσης και δοκιμή επίδοσης πολλούς διαφορετικούς τύπους server / πρωτοκόλλου:

- Web - HTTP, HTTPS
- SOAP/REST
- FTP
- Βάση δεδομένων μέσω JDBC
- LDAP
- Message-oriented middleware (MOM) μέσω JMS
- Mail - SMTP (S), POP3 (S) και IMAP (S)
- MongoDB (NoSQL)
- εντολές ή εντολές shell
- TCP