

Asynchronous educational platform usage analysis

Stavros Valsamidis, Theodosios Theodosiou, Ioannis Kazanidis,
TEI of Kavala
svalsam@teikav.edu.gr, theodosios.theodosiou@gmail.gr,
kazanidis@teikav.edu.gr

Sotirios Kontogiannis, Alexandros Karakos
Democritus University of Thrace
skontog@ee.duth.gr, karakos@ee.duth.gr

Abstract

Learning Management Systems (LMS) are being used more and more supplementing traditional educational courses. LMS store large amounts of course data, in web logs. In Greece, the Greek University Network (GUNet) uses the Open eClass (GUNet, 2010) platform, which is an evolution of Claroline (Claroline, 2010).

Earlier studies have suggested that higher education institutions could exploit the potential of Learning Management Systems (LMSs) by generating reporting tools that identify student's tracking data. Timely assessment of course usage allows educational interventions to the course material. We focus on LMS course assessment using metrics and data mining methods.

Although data mining methods in the World Wide Web have been systematically used in a lot of e-commercial applications, their usage is limited at asynchronous e-learning systems.

This paper aims at the measurement of the LMS characteristics by using quantification indexes and metrics for qualitative attributes of the courses.

It proposes the proper use of Sessions, Pages, Unique pages, UniquePpages per Course per Session, and combines them into three new indexes, namely interest, homogeneity and enrichment. It also proposes a new metric called activity for analyzing course usage. The evaluation of courses usage is achieved with a metric called final score which properly combines the above indexes and metric. The methodology has been successfully applied at data originated from LMS log files of the Information Management Department of TEI Kavala. The experimental results show that if these indexes and metrics are properly combined can quantify quality course characteristics.

Keywords: e-Learning, indexes, metrics, course evaluation, web mining, clustering.

JEL classification: I21, I23, I29

Ανάλυση χρήσης πλατφόρμας ασύγχρονης διδασκαλίας

Βαλσαμίδης Σταύρος, Θεοδοσίου Θεοδόσιος, Καζανίδης Ιωάννης,
Τεχνολογικό Εκπαιδευτικό Ίδρυμα Καβάλας
svalsam@teikav.edu.gr, theodosios.theodosiou@gmail.gr,
kazanidis@teikav.edu.gr

Κοντογιάννης Σωτήριος, Καράκος Αλέξανδρος
Δημοκρίτειο Πανεπιστήμιο Θράκης
skontog@ee.duth.gr, karakos@ee.duth.gr

Περίληψη

Τα συστήματα ασύγχρονης ηλεκτρονικής διδασκαλίας έχουν καθιερωθεί εδώ και αρκετά χρόνια. Μάλιστα στην Ελλάδα, το δίκτυο των Ελληνικών Πανεπιστημίων (GUNET) χρησιμοποιεί το Open eClass, το οποίο είναι μετεξέλιξη της γνωστής ελεύθερης πλατφόρμας Claroline (2010).

Νεότερες μελέτες έδειξαν ότι οι εκπαιδευτικοί οργανισμοί μπορούν να αξιοποιήσουν την ισχύ που παρέχουν τα συστήματα αυτά με την δημιουργία εργαλείων που αναγνωρίζουν τα δεδομένα που περιέχουν τις κινήσεις των φοιτητών. Η έγκαιρη αξιολόγηση της χρήσης των μαθημάτων επιτρέπει εκπαιδευτικές παρεμβάσεις στο υλικό τους.

Από την άλλη πλευρά ενώ οι τεχνικές εξόρυξης δεδομένων στον παγκόσμιο ιστό έχουν χρησιμοποιηθεί συστηματικά σε ποικίλες εφαρμογές, κυρίως ηλεκτρονικού εμπορίου, δεν υπήρξε ανάλογη εφαρμογή τους στα συστήματα ηλεκτρονικής διδασκαλίας.

Στόχος της εργασίας είναι η μέτρηση της χρήσης συστημάτων ηλεκτρονικής διδασκαλίας με την ποσοτικοποίηση ποιοτικών χαρακτηριστικών.

Η συγκεκριμένη εργασία προτείνει την συνδυασμένη χρήση των Sessions, Pages, Unique pages, Unique pages per course per session, τριών νέων δεικτών με τα ονόματα interest, homogeneity και enrichment και μιας μετρικής με το όνομα activity για την ανάλυση των δραστηριοτήτων των φοιτητών κατά την χρήση της πλατφόρμας ασύγχρονης διδασκαλίας. Η μέτρηση της επίδοσης της χρήσης των μαθημάτων γίνεται με την μετρική Final. Αυτοί οι δείκτες και οι μετρικές εφαρμόστηκαν σε δεδομένα που προήλθαν από τα αρχεία καταγραφής της πλατφόρμας ασύγχρονης εκπαίδευσης του Τμήματος Διαχείρισης Πληροφοριών του ΤΕΙ Καβάλας. Τα αποτελέσματα έδειξαν ότι εφόσον συνδυαστούν κατάλληλα μπορούν να ποσοτικοποιήσουν τα ποιοτικά χαρακτηριστικά των μαθημάτων.

Λέξεις Κλειδιά: e-Learning, indexes, metrics, course evaluation, web mining, clustering.

Τυποποίηση JEL: I21, I23, I29

1. Εισαγωγή

Τα συστήματα διαχείρισης εκπαιδευτικού περιεχομένου (Course Management Systems-CMS), προσφέρουν πλήθος μεθόδων εργασίας για καταμερισμό της πληροφορίας και επικοινωνία ανάμεσα στους συμμετέχοντες ανά μάθημα. Επιτρέπουν στους διδάσκοντες να κατανέμουν εργασίες σε φοιτητές, να παράγουν-τοποθετούν υλικό περιεχομένου του μαθήματος, να προετοιμάζουν εργασίες και τεστ, να δημιουργούν συζητήσεις, να διδάσκουν τάξεις από απόσταση και να ενεργοποιούν την συνεργατική μάθηση με forums, wiki, εικονική αλληλεπίδραση μαθητή-σπουδαστή με χρήση οπτικοακουστικών μέσων, χώρους αποθήκευσης αρχείων, υπηρεσίες νέων κ.α. (Romero et al, 2008). Παραδείγματα εμπορικών συστημάτων αποτελούν τα Blackboard (2010), Virtual-U (2010), WebCT (2008) και TopClass (2010) ενώ ελεύθερα συστήματα είναι τα Moodle (2010), Ilias (2010), Claroline (2010), aTutor (2010) και Secondlife (2010).

Στην Ελλάδα, το δίκτυο των Ελληνικών Πανεπιστημίων (GUNET) χρησιμοποιεί την πλατφόρμα OPEN eCLASS (2010), η οποία είναι μετεξέλιξη του Claroline. Το σύστημα αυτό είναι μια πλατφόρμα ασύγχρονης τηλεεκπαίδευσης υλοποιημένη σε PHP, που χρησιμοποιεί το σύστημα διαχείρισης βάσεων δεδομένων Mysql και τον εξυπηρετητή ιστοσελίδων Apache. Το λογισμικό είναι κάτω από την άδεια χρήσης GPL (Άδεια ελεύθερου λογισμικού - Λογισμικού ανοιχτού κώδικα) (GNU, 2010). Κάθε μάθημα στην πλατφόρμα του e-class αποτελείται από τα κάτωθι υποσυστήματα, τα οποία προσδιορίζουν και τις δυνατότητες της CMS πλατφόρμας:

1. Η Ατζέντα, όπου παρουσιάζονται χρονικά τα γεγονότα-σταθμοί του μαθήματος,
2. Τα έγγραφα, όπου αποθηκεύεται το εκπαιδευτικό υλικό του μαθήματος,
3. Οι ανακοινώσεις που αφορούν το μάθημα,
4. Οι περιοχές συζητήσεων για την ανταλλαγή απόψεων και ιδεών σε θέματα σχετικά με το μάθημα,
5. Η Δημιουργία ομάδων εργασίας (ανοικτές ή κλειστές),
6. Οι Σύνδεσμοι από το Διαδίκτυο που αφορούν το αντικείμενο του μαθήματος,
7. Οι εργασίες εκπαιδευόμενων, δηλαδή ο χρονοπρογραμματισμός, αποθήκευση και ηλεκτρονική διαχείριση των εργασιών του μαθήματος,
8. Οι ασκήσεις αυτοαξιολόγησης, που δημιουργεί ο εκπαιδευτής με στόχο την εξάσκηση των εκπαιδευόμενων στην ύλη του μαθήματος,
9. Η περιγραφή Μαθήματος,
10. Τα βίντεο μαθήματος, όπου αποθηκεύονται και παρουσιάζονται αρχεία βίντεο ή σύνδεσμοι σε αρχεία βίντεο αποθηκευμένα σε έναν Video on Demand Server που αφορούν το μάθημα,
11. Η γραμμή Μάθησης, που παρέχει τη δυνατότητα στους εκπαιδευτές να οργανώσουν το εκπαιδευτικό τους υλικό σε ενότητες και στους εκπαιδευόμενους να ακολουθούν μια σειρά από βήματα ως δραστηριότητες μάθησης ακολουθεί το πρότυπο SCORM 2004 (SCORM, 2010),
12. Η τηλεσυνεργασία, είναι ένα υποσύστημα που επιτρέπει την ανταλλαγή σε πραγματικό χρόνο μηνυμάτων μεταξύ των συμμετεχόντων στο μάθημα,
13. Τα ερωτηματολόγια είναι ένα υποσύστημα που παρέχει τη δυνατότητα δημιουργίας δημοσκοπήσεων,
14. Ο χώρος ανταλλαγής αρχείων, όπου υποστηρίζεται η ανταλλαγή αρχείων,
15. Το Wiki, όπου συμμετέχοντες στο μάθημα επεξεργάζονται από κοινού το περιεχόμενο διαφόρων κειμένων και
16. Τα εργαλεία διαχείρισης του μαθήματος.

Στόχος της συγκεκριμένης εργασίας είναι η μέτρηση της χρήσης συστημάτων ηλεκτρονικής διδασκαλίας με την ποσοτικοποίηση ποιοτικών χαρακτηριστικών.

Η εργασία έχει την παρακάτω οργάνωση. Η δεύτερη ενότητα περιέχει το θεωρητικό υπόβαθρο. Η τρίτη ενότητα αναφέρεται στις σχετικές εργασίες. Η τέταρτη ενότητα παρουσιάζει την ερευνητική μεθοδολογία. Η πέμπτη ενότητα αναδεικνύει τα αποτελέσματα της έρευνας. Τέλος στην έκτη ενότητα γίνεται σχολιασμός και στην έβδομη παρατίθενται τα συμπεράσματα.

2. Θεωρητικό υπόβαθρο

Τα κυριότερο πρόβλημα των CMS συστημάτων όπως το Claroline, Moodle, κ.α., είναι ότι γίνεται δύσκολη η διαχείριση και αξιοποίηση αυτής της πληροφορίας που περιέχουν λόγω του όγκου της. Συνήθως τέτοια συστήματα παράγουν κάποιες αναφορές με συγκεντρωτικά στοιχεία που δεν βοηθούν όμως στην εξαγωγή χρήσιμων συμπερασμάτων για το μάθημα ή τους συμμετέχοντες σε αυτό και είναι χρήσιμες μόνο για διαχειριστικούς σκοπούς της εκάστοτε πλατφόρμας. Επίσης, οι υπάρχουσες πλατφόρμες ηλεκτρονικής μάθησης, δεν προσφέρουν συγκεκριμένα εργαλεία που να επιτρέπουν αξιολόγηση των ενεργειών των χρηστών και του μαθησιακού περιεχομένου τους.

Η εξόρυξη γνώσης από δεδομένα (data mining) επιτυγχάνεται με τη χρήση αλγορίθμων για την εξαγωγή πληροφοριών και προτύπων μέσω KDD διαδικασιών (Hand et al., 2001). KDD (Knowledge Data Discovery) είναι το σύνολο των διαδικασιών ανακάλυψης γνώσης σε βάσεις δεδομένων με την εύρεση έγκυρων πληροφοριών και αναγνώρισης χρήσιμων και κατανοητών μοντέλων στα δεδομένα (Kantardzic, 2003. Η Διαδικασία KDD περιλαμβάνει:

1. Την ανάπτυξη και κατανόηση των στόχων του περιεχομένου της εφαρμογής και των στόχων του τελικού χρήστη.
2. Την επιλογή των δεδομένων.
3. Τη δημιουργία του στόχου-συνόλου δεδομένων.
4. Τον καθαρισμό και την προεπεξεργασία δεδομένων.
5. Τον μετασχηματισμό των δεδομένων.
6. Την επιλογή των στόχων και των αλγορίθμων εξόρυξης δεδομένων.
7. Την εξόρυξη δεδομένων.
8. Την αξιολόγηση των προτύπων.
9. Την σταθεροποίηση των παραγόμενων προτύπων και την παρουσίαση της γνώσης.

Οι κυριότερες μέθοδοι που χρησιμοποιούνται στην Εξόρυξη Γνώσης από δεδομένα είναι οι: (1) Συσταδοποίηση, (2) Κατηγοριοποίηση, (3) Παλινδρόμηση, (4) Πρόβλεψη, (5) Κανόνες Συσχέτισης και (6) Ανακάλυψη Ακολουθιών (Witten and Eibe, 2000). Κάποια από τα εργαλεία εξόρυξης δεδομένων είναι τα DB Miner, Speed Tracer, Commerce Trends, Clementine, Sawmill 5, Funnel Web Pro, Knowledge Studio, Net Analysis κ.α. (Galeas, 2009).

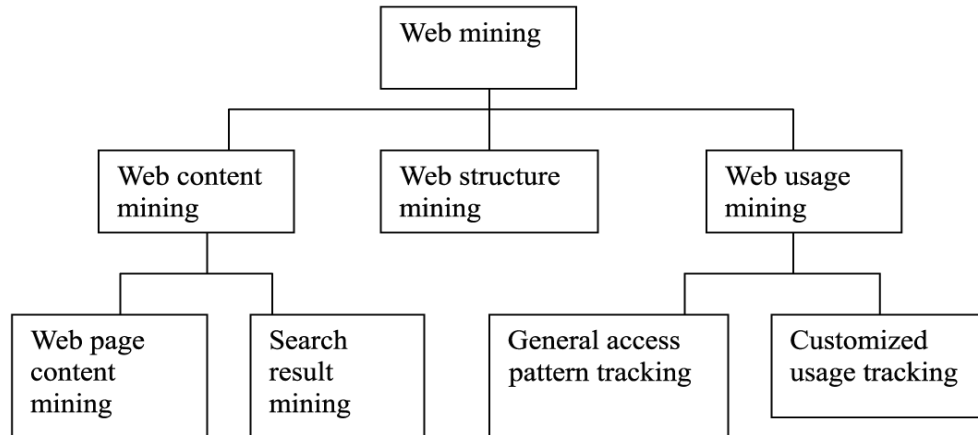
Η εξόρυξη γνώσης από δεδομένα του παγκόσμιου ιστού (web mining) αποτελεί μια μερική υποπερίπτωση εφαρμογής τεχνικών εξόρυξης δεδομένων για αποκάλυψη προτύπων (patterns) στο Διαδίκτυο. Διακρίνεται σε: Web usage mining, Web content mining και Web structure mining (Spiliopoulou, 1999; Kosala and Blockeel, 2000; Bing, 2007).

Στο Web usage mining-WUM (Srivastava et al., 2000) περιλαμβάνονται, εργαλεία παρακολούθησης των αιτημάτων των χρηστών όπως καταγράφεται από τους εξυπηρετητές ιστοσελίδων, με τη χρήση τεχνικών αποκάλυψης και ανάλυσης προτύπων (discovery and analysis of patterns) (Srivastava et al., 2000; Mobasher et al, 1997; Mobasher et al, 1996; Cooley, 2000; Cooley et al., 1997; Srivastava et al., 2000; Zaiane et al., 1998) καθώς και συνεχής παρακολούθηση των προσβάσεων σε σελίδες μιας διαδικτυακής εφαρμογής που αντιστοιχούν και αποδίδουν προσωποποιημένες σελίδες στη βάση παρόμοιων προτύπων επισκέψεων (Pierrakos et al., 2003; Nasraoui, 2000).

Web content mining είναι η διαδικασία αποκάλυψης χρήσιμης πληροφορίας από δεδομένα κειμένου, εικόνας, ηχητικών ή οπτικοακουστικών εγγραφών, τα οποία είναι τοποθετημένα σε ιστοσελίδες του διαδικτύου. Για παράδειγμα το: NLP (Natural language processing) και IR (Information retrieval) (Kosala and Blockeel, 2000; Bing, 2007; Feldman and Sanger, 2006).

Στο Web structure mining χρησιμοποιείται η θεωρία γράφων για ανάλυση κόμβου (ιστοσελίδα), της δομής σύνδεσης των κόμβων μεταξύ των (ιστοσελίδες), ενός web site για εξαγωγή προτύπων από υπερσυνδέσμους στο Διαδίκτυο (Feldman and Sanger, 2006; Thelwall, 2004). Γίνεται επίσης χρήση δομής δένδρων για ανάλυση και περιγραφή του περιεχομένου HTML (Hyper Text Markup Language) ιστοσελίδων στο διαδίκτυο ή XML (eXtensible

Markup Language) ετικετών που περιέχονται μέσα στις ιστοσελίδες (Kosala and Blockeel, 2000; Bing, 2007; Yang et al., 2002; Ghani et al., 2001). Η ιεραρχική ταξινόμηση του Web mining φαίνεται στο Σχήμα 1 (Kosala and Blockeel, 2000; Mobasher et al., 1997).



Σχήμα 1. Η δένδροειδής ιεραρχική δομή των τεχνικών εξόρυξης δεδομένων από το Διαδίκτυο (web mining).

Για το Web Usage Mining υπάρχουν εργαλεία όπως τα εμπορικά DB Miner, Speed Tracer, Commerce Trends, Clementine, WUM, Sawmill 5, Funnel Web Pro, Knowledge Studio, Net Analysis και τα ελεύθερα Net Analysis weblog_parse και WebLog (Galeas, 2009). Άλλα εργαλεία για Web Usage Mining που έχουν προταθεί είναι τα (Perkowitz and Etzioni, 1999; Yan et al., 1996; Baraglia and Palmerini, 2002).

Οι μέθοδοι εξόρυξης δεδομένων στον παγκόσμιο ιστό έχουν χρησιμοποιηθεί συστηματικά σε πολλές εφαρμογές ηλεκτρονικού εμπορίου. Δεν υπήρξε όμως ανάλογη χρησιμοποίηση και στα συστήματα ασύγχρονης ηλεκτρονικής διδασκαλίας (Zaiane, 2001).

Όσον αφορά όμως την εφαρμογή της εξόρυξης δεδομένων σε πλατφόρμες ηλεκτρονικής διδασκαλίας, υπάρχουν αρκετές διαφορές σε σχέση τις αντίστοιχες πλατφόρμες ηλεκτρονικού εμπορίου, πεδίο που έχει διερευνηθεί συστηματικά (Romero and Ventura, 2007). Οι διαφορές συνίστανται στα εξής:

1. Περιοχή: Ο σκοπός του ηλεκτρονικού εμπορίου είναι να καθοδηγεί τους πελάτες στην αγορά ενώ ο σκοπός της ηλεκτρονικής μάθησης είναι να καθοδηγηθούν οι φοιτητές στην μάθηση.
2. Δεδομένα: Στο ηλεκτρονικό εμπόριο τα χρησιμοποιημένα δεδομένα είναι κανονικά, απλές καταγραφές πρόσβασης σε εξυπηρετητές web, αλλά στην ηλεκτρονική μάθηση υπάρχει περισσότερη πληροφορία σε σχέση με την αλληλεπίδραση του φοιτητή. Το πρότυπο χρήστη είναι επίσης διαφορετικό στα δύο συστήματα.
3. Στόχος: Ο στόχος της εξόρυξης δεδομένων στο ηλεκτρονικό εμπόριο είναι η αύξηση του κέρδους, το οποίο είναι από και μπορεί να μετρηθεί από την άποψη των χρηματικών ποσών, του αριθμού πελατών και της πίστης πελατών. Ο στόχος της εξόρυξης δεδομένων στην ηλεκτρονική μάθηση είναι η βελτίωση της εκμάθησης. Αυτός ο στόχος είναι υποκειμενικός και δύσκολο να ποσοτικοποιηθεί.
4. Τεχνικές: Τα εκπαιδευτικά συστήματα έχουν ειδικά χαρακτηριστικά που απαιτούν διαφορετική μεταχείριση του προβλήματος της εξόρυξης, άρα απαιτούνται ειδικές τεχνικές εξόρυξης δεδομένων για να αναλύσουν ειδικότερα τη διαδικασία μάθησης.

3. Σχετικές εργασίες

Στο Web mining υπάρχει το εργαλείο Analog που αποτελείται από δύο μέρη, ώστε να πραγματοποιεί εξόρυξη σε δύο επίπεδα, online και offline, ανάλογα με την δραστηριότητα του εξυπηρετητή (Perkowitz and Etzioni, 1999). Η δραστηριότητα των χρηστών που είναι καταγεγραμμένη σε log

αρχεία του εξυπηρετητή, προεπεξεργάζεται έτσι ώστε να σχηματιστούν συστάδες (clusters) συνόδων (sessions) του χρήστη. Το online τμήμα δημιουργεί ενεργές συνόδους χρήστη που στην συνέχεια ταξινομούνται (classified) ανάλογα με τις συστάδες που υπάρχουν στο offline τμήμα.

Το Page Gather είναι ένα σύστημα Web Usage Mining που δημιουργεί δείκτες (index) σελίδων που περιέχουν συνδέσμους σε σελίδες παρόμοιες μεταξύ τους (Yan et al., 1996). Το Page Gather βρίσκει συστάδες σελίδων αντί για συστάδες συνόδων (sessions). Ξεκινώντας από τις συνόδους δραστηριότητας χρήστη, δημιουργείται ο πίνακας συνάφειας (co-occurrence matrix). Μια σημαντική υπόθεση που αναφέρεται στο (Perkowitz and Etzioni, 1999) είναι ότι οι χρήστες συμπεριφέρονται συνεκτικά (coherently) κατά την διάρκεια της πλοήγησης, δηλαδή σελίδες μέσα στην ίδια σύνοδο σχετίζονται εννοιολογικά. Αυτή η παραδοχή ονομάζεται συνεκτικότητα (coherence) (Yan et al., 1996).

Το SUGGEST είναι ένα WUM σύστημα που σχεδιάστηκε για να παράγει συνδέσμους σε σελίδες (προτάσεις) του εν δυνάμει ενδιαφέροντος για ένα χρήστη (Baraglia and Palmerini, 2002). Υλοποιήθηκε σαν επέκταση του Apache web server.

Οι προηγούμενες εργασίες αποτέλεσαν το κινητήριο έναυσμα στην εφαρμογή της εξόρυξης δεδομένων στην ηλεκτρονική διδασκαλία (Romero and Ventura, 2007), δηλαδή στα παραδοσιακά εκπαιδευτικά συστήματα, στις βασισμένες στο Διαδίκτυο σειρές μαθημάτων, στα γνωστά συστήματα διαχείρισης μάθησης (Open eClass, Moodle) και στα προσαρμοστικά και ευφυή, βασισμένα στο Διαδίκτυο εκπαιδευτικά συστήματα.

Οι λειτουργίες της εξόρυξης δεδομένων σε μια πλατφόρμα ηλεκτρονικής μάθησης, μπορούν να επιτρέπουν (Romero and Ventura, 2007; Ueno, 2004):

1. Περιληπτική παρουσίαση της κατάστασης γνώσης των εκπαιδευομένων.
2. Περιληπτική παρουσίαση των διαδικασιών μάθησης των εκπαιδευομένων.
3. Περιληπτική παρουσίαση των διαδικασιών συζητήσεων των εκπαιδευομένων.
4. Πρόβλεψη της κατάστασης γνώσης των εκπαιδευομένων στο μέλλον.
5. Ανίχνευση των εκπαιδευομένων που χρειάζονται την βοήθεια του διδάσκοντα.
6. Αναλύσεις των περιεχομένων ηλεκτρονικής μάθησης.
7. Ανάλυση των χαρακτηριστικών καθενός εκπαιδευομένου στη συζήτηση.

Επιπρόσθετα, είναι και τα πλεονεκτήματα που αναφέρονται στο σύστημα IIMS "Samurai" (Ueno, 2004) που έχει υλοποιηθεί:

8. Ο διδάσκων μπορεί να γνωρίζει τις διαδικασίες εκμάθησης των εκπαιδευομένων και τις καταστάσεις γνώσης τους, προκειμένου να δοθούν μερικές αποτελεσματικές εκπαιδευτικές συστάσεις στους εκπαιδευόμενους.
9. Οι διδάσκοντες μπορούν να αναλύσουν τα περιεχόμενα τους προκειμένου να τα βελτιώσουν.
10. Οι διδάσκοντες μπορούν να κατανοήσουν το περιεχόμενο της συζήτησης των εκπαιδευομένων συνολικά.
11. Οι εκπαιδευόμενοι μπορούν να αξιολογήσουν τις καταστάσεις γνώσης τους για να αντιστοιχίσουν μόνοι, τους τρόπους εκμάθησής τους.
12. Οι εκπαιδευόμενοι μπορούν να αξιολογήσουν τις καταστάσεις εκμάθησής τους μόνοι τους.
13. Οι εκπαιδευόμενοι μπορούν να αξιολογήσουν τα δικά τους χαρακτηριστικά στη συζήτηση που αντανακλούν την συζήτηση τους.
14. Οι εκπαιδευόμενοι μπορούν να αξιολογήσουν τον βαθμό υποστήριξης των δικών τους απόψεων στον πίνακα συζήτησης και να παρακινηθούν για να συμμετέχουν στη συζήτηση.

Ένα άλλο ενδιαφέρον σύστημα είναι το CelGrid (Myszkowski et al., 2008). Το σύστημα αυτό βελτιώνει την ηλεκτρονική μάθηση μετά από άμεση ανάλυση δεδομένων. Επιπρόσθετα, κάνει προτάσεις στους εκπαιδευόμενους για το εκπαιδευτικό υλικό που τους ενδιαφέρει, προσφέρει ικανοποιητική επανατροφοδότηση στους καθηγητές ώστε να το προσαρμόσουν κατάλληλα για τους εκπαιδευόμενους και τέλος προσφέρει ένα εργαλείο στους διαχειριστές συστήματος για την άμεση αναγνώριση τυχόν παράνομων εισόδων στο σύστημα.

Ανάλογη εργασία έχει γίνει με το σύστημα MATEP, το οποίο παρέχει ένα ικανοποιητικό εργαλείο για ανάλυση δεδομένων προερχόμενων από διάφορες πηγές ηλεκτρονικής διδασκαλίας και δρα σε δύο επίπεδα (Zorrilla and Alvarez, 2008). Πρώτα, κάνει μια μίξη δεδομένων από διαφορετικές πηγές κατάλληλα επεξεργασμένες και ολοκληρωμένες. Τα δεδομένα προέρχονται από log αρχεία πλατφόρμας e-learning, εικονικά μαθήματα, ακαδημαϊκά και δημογραφικά δεδομένα. Δεύτερο, τα τροφοδοτεί σε ένα web house δεδομένων που παρέχει στατικές και δυναμικές αναφορές. Το MATEP αξιοποίησε προηγούμενες εργασίες στα CourseVis και GISMO.

Το CourseVis (Mazza and Dimitrova, 2007) είναι ένα εργαλείο απεικόνισης που παρακολουθεί τα web log δεδομένα από ένα LMS. Με το μετασχηματισμό αυτών των δεδομένων, παράγει γραφικές απεικονίσεις που πληροφορεί τους εκπαιδευτικούς καλά για αυτό που συμβαίνει ακριβώς σε κατηγορίες της απόσταση εκμάθησης.

Το GISMO (2010) είναι ένα εργαλείο παρόμοιο με CourseVis, αλλά παρέχει στους εκπαιδευτικούς διαφορετικές πληροφορίες, όπως λεπτομέρειες του σπουδαστή στην χρησιμοποίηση του υλικού των μαθημάτων (Mazza and Milani, 2004; GISMO, 2010).

Το Sinergo/ColAT (Avouris et al, 2005) είναι ένα εργαλείο που ενεργεί ως διερμηνέας της δραστηριότητας των σπουδαστών LMS. Ο Mostow περιγράφει ένα εργαλείο που χρησιμοποιεί τα log αρχεία προκειμένου να αναπαραστήσει την αλληλεπίδραση εκπαιδευτικού-σπουδαστή σε ιεραρχική δομή (Mostow et al., 2007).

Ανάλυση δεδομένων που αφορούν την χρήση του Moodle από τους φοιτητές, έχει γίνει στις εργασίες (Baruque et al., 2007; Nagi and Suesawaluk, 2008), με τη χρήση του συστήματος Reports το οποίο ενσωματώνεται άμεσα στο Moodle. Με το σύστημα αυτό εξασφαλίζεται σε μεγάλο βαθμό η διαδραστικότητα ανάμεσα σε καθηγητές και φοιτητές.

4. Μεθοδολογία

Όπως προαναφέρθηκε στην ενότητα 1 ο στόχος της εργασίας είναι η μέτρηση της χρήσης συστημάτων ηλεκτρονικής διδασκαλίας.

Η προτεινόμενη μεθοδολογία αποτελείται από τέσσερα βήματα, συγκεκριμένα (1) την καταγραφή των δεδομένων με τις σχετικές διαδικασίες, (2) την προ-επεξεργασία των δεδομένων και τον υπολογισμό των μετρικών τιμών ανά μάθημα (3) την επεξεργασία των δεδομένων και τον υπολογισμό των δεικτών ανά μάθημα και τέλος (4) την αξιολόγηση από την αποτίμηση της χρήσης των μαθημάτων.

Τα δύο πρώτα βήματα περιγράφησαν με λεπτομέρεια στο framework (Kazanidis et al., 2009) και διευκολύνουν την εξαγωγή χρήσιμων πληροφοριών από δεδομένα που καταγράφηκαν σε εξυπηρετητή στον οποίο τρέχει ένα LMS.

4.1 Καταγραφή των δεδομένων με τις σχετικές διαδικασίες (Logging data and procedures)

Το βήμα αυτό περιλαμβάνει την καταγραφή των δεδομένων και τις διαδικασίες σε πλατφόρμες e-learning.

Ως γνωστό, οι εξυπηρετητές παγκόσμιου ιστού έχουν την δυνατότητα να καταγράφουν όλες τις αιτήσεις. Πιο συγκεκριμένα, ο Apache εξυπηρετητής ιστοσελίδων χρησιμοποιεί τις ακόλουθες διαμορφώσεις στα log αρχεία του:

1) Common Log Format (CLF). Μία τυπική διαμόρφωση που χρησιμοποιείται από πολλούς εξυπηρετητές ιστοσελίδων. Η διαμόρφωση αυτή χρησιμοποιεί ένα συγκεκριμένο κορδόνι διαμόρφωσης (format string), το οποίο περιγράφει τον τύπο και τον τρόπο που θα αποθηκευτούν τα δεδομένα πρόσβασης του χρήστη στο log αρχείο. Π.χ.:

```
LogFormat "%h %l %u %t \"%r\" %>s %b" common
CustomLog logs/access_log common
```

Μερικά από τα στοιχεία του χρήστη που μπορούν να αποθηκευτούν από την διαμόρφωση common είναι τα εξής:

- Remote host (%h): hostname φυλλομετρητή ή IP.
- Remote log name (%l) χρήστη (σχεδόν πάντοτε "-" που σημαίνει "unknown").

- Authuser (%u): αυθεντικοποιημένο όνομα χρήστη.
- Date (%t): Ημερομηνία και ώρα του αιτήματος του πελάτη στον εξυπηρετητή (σε μορφή ημερομηνίας ή timestamp).
- "request" (%r): Η ακριβής αίτηση του πελάτη στον εξυπηρετητή. Περιλαμβάνει και τη χρησιμοποιούμενη μέθοδο : GET, POST, HEAD, κ.α.
- Status (%s): Ο κωδικός κατάστασης HTTP που επιστρέφει ο εξυπηρετητής. Συνήθεις κωδικοί περιλαμβάνουν:
 - 200: OK.
 - 3xx: Κάποιο είδος ανακατεύθυνσης.
 - 4xx: Κάποιο είδος λάθους πελάτη.
 - 5xx: Κάποιο είδος λάθους εξυπηρετητή.

• Bytes (%b): Το μήκος του περιεχομένου της απόκρισης του εξυπηρετητή.
 2) Extended Log Format το οποίο είναι ίδιο με το common log, άλλα μπορεί να αποθηκεύσει επιπλέον δυο στοιχεία του χρήστη: 1. Το πεδίο *Referer*, δηλαδή την ιστοσελίδα πάνω από την οποία εκτελέστηκε η αίτηση του περιηγητή ιστοσελίδων του χρήστη (base URI) και 2. Το πεδίο *User-Agent* που περιέχει στοιχεία που ο περιηγητής ιστοσελίδων του πελάτη δηλώνει για των εαυτό του (έκδοση περιηγητή, λειτουργικό σύστημα πελάτη, κ.α.). Το Extended Log Format, έχει επίσης τη δυνατότητα αποθήκευσης των αιτήσεων του πελάτη προς τον εξυπηρετητή ή των απαντήσεων του εξυπηρετητή προς τον πελάτη καθώς και το χρόνο ολοκλήρωσης μια αίτησης σε δευτερόλεπτα(sec).

3) Cookie Log Format. Το Cookie log, είναι μια δυνατότητα του mod_log_config υποσυστήματος (module) του Apache που επιτρέπει την αποθήκευση των cookies του πελάτη στο log file του εξυπηρετητή. Μπορεί να χρησιμοποιηθεί σε συνδυασμό με το CLF.

4) Forensic Log Format. Σε αυτή τη διαμόρφωση, η αποθήκευση των αιτήσεων του εξυπηρετητή γίνεται πριν την επεξεργασία τους από τον εξυπηρετητή και μετά την επεξεργασία τους από τον εξυπηρετητή (δύο εγγραφές στο log αρχείο ανά αίτηση πελάτη στον εξυπηρετητή). Για την αναγνώριση κάθε αίτησης ένα μοναδικό ID ανατίθεται σε κάθε μια αίτηση και ένα ζευγάρι πρόσημων +/- που σηματοδοτούν την πρώτη ή τη δεύτερη καταγραφή της αίτησης.

Στις περιπτώσεις του Cookie και Forensic log format, αποθηκεύονται τεράστιες ποσότητες δεδομένων.

Ενδεικτικά, από καταγραφές που έγιναν τον Απρίλιο του 2009 στην υπάρχουσα πλατφόρμα ασύγχρονης διδασκαλίας e-University του ΤΕΙ Καβάλας, δείγμα δεδομένων που συλλέχτηκαν, φαίνεται στον πίνακα 1.

Πίνακας 1: Δείγμα δεδομένων σε forensic log format

remote_host	request_uri	remote_logname	remote_user	request_method	request_time	request_protocol	status	bytes_sent	referer	agent
66.249.72.212	/component/search/smb.conf.html	-	-	GET	[03/Mar/2009:18:57:00 +0200]	HTTP/1.1	200	9805	-	Mozilla/5.0 (compatible; Googlebot/2.1; #http://ww...
66.249.72.212	/mailing-list/56.html	-	-	GET	[03/Mar/2009:18:57:31 +0200]	HTTP/1.1	200	12730	-	Mozilla/5.0 (compatible; Googlebot/2.1; #http://ww...
66.249.72.212	/newsfeeds.html	-	-	GET	[03/Mar/2009:18:58:01 +0200]	HTTP/1.1	200	7967	-	Mozilla/5.0 (compatible; Googlebot/2.1; #http://ww...

Με πεδία:

- Client IP: 66.249.72.212
- Authenticated User ID: - -
- Time/Date: [[03/Mar/2009:18:57:31 +0200]]
- Request: "GET / HTTP/1.1"
- Status: 200
- Bytes: -
- Referrer: "--"
- Agent: "Mozilla/5.0 (compatible; Googlebot/2.1

- Μέθοδο: GET. Άλλες συνήθειες μέθοδοι είναι οι POST και HEAD.
- URI: /
- Πρωτόκολλο: HTTP/1.0-1.1

Η ανάπτυξη μιας τέτοιας ενότητας (module) έχει τα ακόλουθα πλεονεκτήματα: (i) η γρήγορη αποθήκευση των πληροφοριών χρηστών, δεδομένου ότι εκτελείται κατ' ευθείαν από τον εξυπηρετητή API και όχι από την εφαρμογή e-learning και (ii) τα παραχθέντα δεδομένα είναι ανεξάρτητα από τη συγκεκριμένη τυποποίηση που χρησιμοποιείται με την πλατφόρμα e-learning.

4.2 Προ-επεξεργασία και μετρικές τιμές ανά μάθημα

Το παραγόμενο log αρχείο, από το προηγούμενο βήμα, φιλτράρεται με την χρήση της γλώσσας Perl, έτσι περιλαμβάνει μόνο τα παρακάτω πεδία: (i) courseID, που είναι το αναγνωριστικό του κάθε μαθήματος, (ii) sessionID, που είναι το αναγνωριστικό της κάθε συνόδου (session), (iii) page Uniform Resource Locator (URL), που περιέχει τα αιτήματα της κάθε σελίδας της πλατφόρμας που επισκέφθηκε ο χρήστης. Στο βήμα της προ-επεξεργασίας τα προαναφερθέντα στοιχεία χρησιμοποιούνται για τον υπολογισμό των μετρικών τιμών ανά μάθημα. Οι προτεινόμενες μετρικές τιμές που φαίνονται στον πίνακα 2 με σκοπό να διευκολύνουν στην συνέχεια στην αποτίμηση της χρήσης του μαθήματος.

Πίνακας 2. Πεδία και περιγραφή

Πεδίο	Περιγραφή
Course ID	Ο κωδικός του μαθήματος
Sessions	Ο αριθμός των συνόδων ανά μάθημα που πρόβαλε ο χρήστης
Pages	Ο αριθμός των μοναδικών σελίδων ανά μάθημα που πρόβαλε ο χρήστης
Unique pages	Ο αριθμός των μοναδικών σελίδων ανά μάθημα που πρόβαλε ο χρήστης
Unique Pages per CourseID per Session (UPCS)	Ο αριθμός των μοναδικών σελίδων ανά μάθημα που πρόβαλε ο χρήστης σε μια σύνοδο

Προκειμένου να υπολογιστεί η δραστηριότητα στην χρήση των μαθημάτων μετρήθηκαν ο αριθμός των συνόδων και ο αριθμός των σελίδων που αναφέρθηκαν σ' αυτά. Η μετρική unique pages (μοναδικές σελίδες) μετρά τον συνολικό αριθμό των μοναδικών σελίδων ανά μάθημα που προβλήθηκαν από όλους τους χρήστες. Η μετρική UPCS εκφράζει τις μοναδικές επισκέψεις χρηστών ανά μάθημα και ανά σύνοδο προκειμένου να υπολογιστεί η δραστηριότητα με έναν πιο αντικειμενικό τρόπο. Για παράδειγμα, μερικοί αρχάριοι χρήστες μπορεί, καθώς πλοηγούνται σε ένα μάθημα, να επισκεφθούν μια σελίδα περισσότερες από μία φορά. Η UPCS αφαιρεί τις διπλές επισκέψεις σελίδων, στην ίδια σύνοδο χρήστη.

4.3 Επεξεργασία και δείκτες ανά μάθημα

Με βάση τις προηγούμενες μετρικές υπολογίζουμε τους δείκτες του πίνακα 3.

Πίνακας 3. Δείκτες και περιγραφή

Δείκτες	Περιγραφή
Course ID	Ο κωδικός του μαθήματος
Enrichment	Ο εμπλουτισμός των σελίδων (Unique pages/Pages)
Disappointment	Η απογοήτευση των χρηστών όταν προβάλουν σελίδες των μαθημάτων (Sessions/Pages)
Interest	Είναι το συμπλήρωμα ως προς ένα της απογοήτευσης (1-disappointment)
Homogeneity	Η ομοιογένεια των μοναδικών σελίδων ανά σύνοδο (Unique Pages/Sessions)

Το Enrichment είναι μία μετρική τιμή η οποία εκφράζει τον εμπλουτισμό των σελίδων ενός μαθήματος. Ορίσαμε τον δείκτη Enrichment, ως το πηλίκο του αριθμού των μοναδικών σελίδων ενός μαθήματος (Unique Pages) προς τον αριθμό των συνολικών σελίδων που επισκέφτηκαν οι χρήστες της πλατφόρμας (Total Pages), όπου ο αριθμός των μοναδικών σελίδων είναι πάντα μικρότερος του αριθμού των συνολικών σελίδων ενός μαθήματος.

$$\text{Enrichment} = \frac{\text{Unique Pages}}{\text{Total Pages}} \quad (1)$$

Οι μοναδικές σελίδες είναι ένα σύνολο σελίδων μοναδικά καθοριζόμενων από τις συνεδρίες του χρήστη ανά μάθημα. Οι τιμές του Enrichment βρίσκονται ανάμεσα στο (0, 1]. Σε ένα μάθημα με ελάχιστες μοναδικές σελίδες το Enrichment τείνει στο 0 ενώ σε μαθήματα που οι εκπαιδευόμενοι ακολουθούν σχεδόν μοναδικές διαδρομές πλοήγησης ο δείκτης πλησιάζει το 1.

Το Enrichment δείχνει το ποσό πληροφορίας που περιλαμβάνεται σε κάθε μάθημα και χρησιμοποιείται από τους εκπαιδευόμενους. Παρέχει ένα μέτρο του αριθμού των μοναδικών σελίδων τις οποίες επισκέπτονται οι εκπαιδευόμενοι και αντιπροσωπεύει πόσο πλούσιο είναι το εκπαιδευτικό υλικό ενός μαθήματος. Αν ο δείκτης Enrichment είναι χαμηλότερος από τον μέσο όρο όλων των μαθημάτων, συνίσταται ο εμπλουτισμός των υφιστάμενων ιστοσελίδων με περισσότερο υλικό. Ωστόσο ο δείκτης εμπλουτισμού του μαθήματος ίσως οδηγήσει σε αναξιόπιστα αποτελέσματα στην περίπτωση που οι χρήστες επισκεφτούν μόνο λίγες σελίδες από κάθε μάθημα σε κάθε σύνοδο.

Για τον λόγο αυτό προτείνουμε ένα νέο δείκτη που ονομάσαμε Disappointment. Ο δείκτης Disappointment αποτυπώνει την απογοήτευση του χρήστη από ένα μάθημα. Υπολογίζει πόσο αρνητικό είναι ένα μάθημα στην λογική ότι όταν ο χρήστης επισκέπτεται σε μία σύνοδο λίγες σελίδες ενός μαθήματος και δεν ικανοποιηθεί από το περιεχόμενό του τότε αποσυνδέεται από το μάθημα. Ο δείκτης της απογοήτευσης ορίζεται ως το πηλίκο του αριθμού των συνόδων προς τον αριθμό των συνολικών ιστοσελίδων ενός διαδικτυακού μαθήματος.

$$\text{Disappointment} = \frac{\text{Sessions}}{\text{Total Pages}} \quad (2)$$

Ο Binali κ.α. (2009) έχουν χρησιμοποιήσει τεχνικές εξόρυξης απόψεων (opinion mining) για τον εντοπισμό των αντιλαμβανόμενων συναισθημάτων των χρηστών και την ανάλυση συγκεκριμένων συστημάτων ηλεκτρονικής εκπαίδευσης. Η απογοήτευση του χρήστη από ένα μάθημα, έχει παρατηρηθεί ότι μπορεί να επηρεάσει την συμπεριφορά του κατά την διαδικασία της εξ' αποστάσεως εκπαίδευσης. Ο δείκτης της απογοήτευσης που παρουσιάσαμε παραπάνω, αποτυπώνει πόσο γρήγορα οι χρήστες σταματούν να επισκέπτονται σελίδες ενός μαθήματος και αποσυνδέονται από αυτό. Δεν χρησιμοποιήθηκε ο συγκεκριμένος δείκτης λόγω της αρνητικής του φύσης. Προτείνεται λοιπόν ο ορισμός ενός νέου δείκτη με θετική σημασία ο οποίος να εκφράζει το ενδιαφέρον του εκπαιδευόμενου για ένα μάθημα. Ο δείκτης αυτός ονομάστηκε Interest και ορίστηκε ως το συμπλήρωμα ως προς την μονάδα της τιμής του δείκτη Disappointment.

$$\text{Interest} = 1 - \text{Disappointment} \quad (3)$$

Θεωρείται δηλαδή ότι το άθροισμα της απογοήτευσης και του ενδιαφέροντος του χρήστη για κάθε μάθημα θα πρέπει να ισούται πάντα με την μονάδα. Έτσι όσο αυξάνεται η απογοήτευση μειώνεται το ενδιαφέρον και αντίστροφα. Η μελέτη της εξόρυξης του ενδιαφέροντος (Interest mining) σε εικονικά μαθησιακά περιβάλλοντα έχει μελετηθεί από τον Rong κ.α. (2008), αλλά με διαφορετικό τρόπο, δεδομένου ότι μελετήθηκε η εξόρυξη ενδιαφέροντος μέσω της ανάλυσης της συμπεριφοράς.

Ένας ακόμη δείκτης που ορίζουμε είναι η ομοιογένεια των μοναδικών σελίδων σε μία σύνοδο ενός μαθήματος. Ο δείκτης αυτός ονομάστηκε ως Homogeneity και ορίζεται ως ο λόγος των μοναδικών σελίδων ενός μαθήματος προς τον αριθμό των συνόδων των χρηστών που επισκέφτηκαν το συγκεκριμένο μάθημα.

$$\text{Homogeneity} = \frac{\text{Unique pages}}{\text{Sessions}} \quad (4)$$

Όπου ο αριθμός των μοναδικών σελίδων είναι πάντα πολύ μικρότερος του αριθμού των συνόδων στο μάθημα (Unique Pages<<Sessions). Για τον ορθότερο υπολογισμό των μετρικών τιμών προσπαθούμε να κρατούμε ένα ικανοποιητικό και σταθερό χρονικό διάστημα των συνόδων του χρήστη σε κάθε μάθημα. Υποθέτουμε ότι μία ορθή τιμή του δείκτη ομοιογένειας κυμαίνεται μεταξύ [0,1), όπου όταν ο δείκτης έχει τιμή κοντά στο 0 σημαίνει ότι οι χρήστες δεν ακολούθησαν κάποια μοναδική διαδρομή σε ένα μάθημα, ενώ όταν ο δείκτης πλησιάζει στο 1 ότι κάθε χρήστης έχει εξακολουθήσει μία μοναδική διαδρομή.

Ο συνυπολογισμός των προαναφερθέντων δεικτών θα επιτρέψει την αξιολόγηση της δραστηριότητας σε κάθε μάθημα.

4.4 Αξιολόγηση χρήσης των μαθημάτων

Η αξιολόγηση της χρήσης των μαθημάτων βασίστηκε σε δύο μετρικές. Τη μετρική Activity και τη μετρική Final. Η μετρική Activity είναι ο μέσος όρος των Enrichment, Interest και Homogeneity, ενώ η μετρική Final είναι το γινόμενο του Activity metric επί το UPCS. Οι δύο αυτές μετρικές μας επιτρέπουν να κατατάξουμε και να ομαδοποιήσουμε τα μαθήματα ανάλογα με τη χρήση τους στην ηλεκτρονική πλατφόρμα.

Η ομαδοποίηση στηρίζεται στην ιεραρχική ταξινόμηση. Η ιεραρχική ταξινόμηση ταξινομεί ιεραρχικά τις ομάδες, οι οποίες μπορούν να παρασταθούν γραφικά με τη δομή ενός δέντρου το οποίο καλείται δενδρογράμμα (Markos et al., 2009). Η ρίζα του δενδρογράμματος αποτελεί την υπερομάδα και περικλείει όλες τις ομάδες του γραφήματος. Τα φύλλα του δενδρογράμματος αντιπροσωπεύουν την κάθε παρατήρηση (μάθημα στην παρούσα εργασία). Οι αλγόριθμοι της ιεραρχικής ταξινόμησης χωρίζονται γενικά σε δύο μεγάλες κατηγορίες. Αυτούς που ξεκινάνε από τα φύλλα του δενδρογράμματος και ομαδοποιώντας σε κάθε βήμα τα φύλλα φτάνουν στη ρίζα (agglomerative) και σε αυτούς (divisive) που ξεκινάνε από τη ρίζα και σε κάθε βήμα σπάνε τη ρίζα σε επιμέρους ομάδες μέχρι να φτάσουν στα φύλλα, δηλαδή τις παρατηρήσεις/μαθήματα. Ο αλγόριθμος που χρησιμοποιήθηκε στην παρούσα εργασία ανήκει στην πρώτη κατηγορία και είναι υλοποιημένος στο στατιστικό λογισμικό ανάλυσης δεδομένων SPSS v17.

Σημαντικό στοιχείο κάθε αλγορίθμου ιεραρχικής ταξινόμησης αποτελεί το μέτρο που χρησιμοποιείται για να ποσοτικοποιήσει τη διαφορά/ομοιότητα μεταξύ των παρατηρήσεων/μαθημάτων, ώστε να καθοριστεί ποιες παρατηρήσεις/μαθήματα θα ομαδοποιηθούν. Στην περίπτωση των μαθημάτων χρησιμοποιούμε ως μέτρο την Ευκλείδεια απόσταση μεταξύ των μετρικών Activity και Final.

5. Αποτελέσματα

Ενδεικτικά, από καταγραφές που έγιναν τον Απρίλιο του 2009, στην πλατφόρμα ασύγχρονης διδασκαλίας e-University του ΤΕΙ Καβάλας τον Απρίλιο του 2009 στο ΤΕΙ Καβάλας, συλλέχτηκαν δεδομένα (πίνακας 1), στα οποία μετά από προ-επεξεργασία προέκυψαν οι μετρικές και δείκτες των πινάκων 2 και 3.

Τα αποτελέσματα για 20 μαθήματα, τα πρώτα 10 μαθήματα, 5 μαθήματα στη μέση και τα τελευταία 5, με την κατάταξη μέσω του final, παρουσιάζονται στον πίνακα 4. Θελήσαμε να εξετάσουμε τους δείκτες και τις μετρικές μας, σε βέλτιστες, μέσες και χειρίστες περιπτώσεις από πλευράς χρήσης.

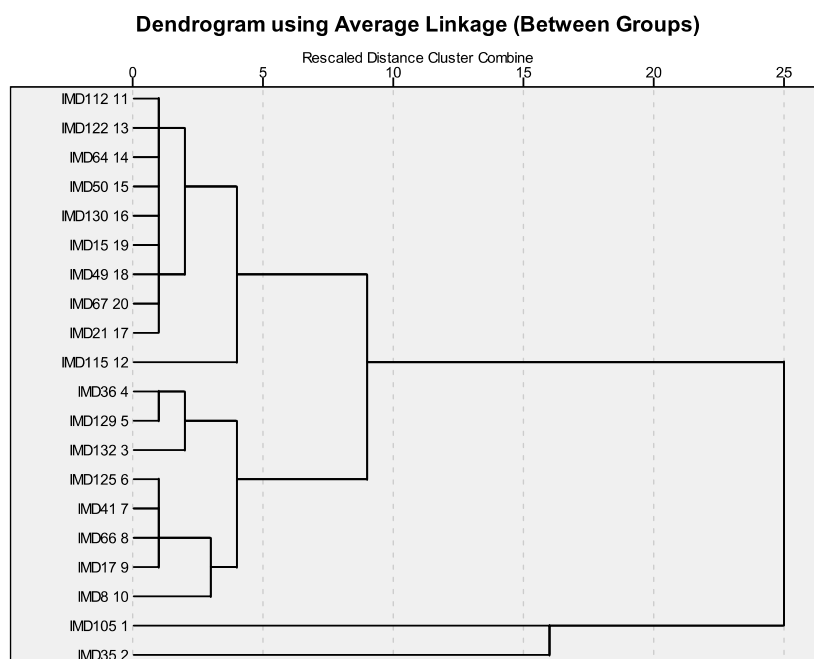
Πίνακας 4. Επεξεργασία που έγινε σε δεδομένα που καταγράφηκαν από την πλατφόρμα ηλεκτρονικής τάξης e-class του ΤΕΙ Καβάλας

Course ID	Ses-sions	Pages	Unique pages	UPCS	Enri- chment	Disappo- intment	Inte- rest	Homoge- neity	Activity metric	Final
IMD105	91	297	11	216	0,963	0,306	0,694	0,121	0,592	127,98
IMD35	87	338	8	179	0,976	0,257	0,743	0,092	0,604	108,05
IMD132	152	230	7	184	0,970	0,661	0,339	0,046	0,452	83,09
IMD36	72	217	7	134	0,968	0,332	0,668	0,097	0,578	77,41
IMD129	75	209	6	131	0,971	0,359	0,641	0,080	0,564	73,90
IMD125	93	164	6	134	0,963	0,567	0,433	0,065	0,487	65,25
IMD41	98	185	8	129	0,957	0,530	0,470	0,082	0,503	64,87
IMD66	56	144	9	107	0,938	0,389	0,611	0,161	0,570	60,97
IMD17	53	206	11	89	0,947	0,257	0,743	0,208	0,632	56,27
IMD8	45	135	8	82	0,941	0,333	0,667	0,178	0,595	48,80
IMD112	30	62	6	46	0,903	0,484	0,516	0,200	0,540	24,83
IMD115	15	73	12	42	0,836	0,205	0,795	0,800	0,810	34,02
IMD122	33	71	7	45	0,901	0,465	0,535	0,212	0,550	24,73
IMD64	22	47	7	39	0,851	0,468	0,532	0,318	0,567	22,12
IMD50	22	46	7	38	0,848	0,478	0,522	0,318	0,563	21,38
IMD130	12	30	5	22	0,833	0,400	0,600	0,417	0,617	13,57
IMD21	11	25	8	24	0,680	0,440	0,560	0,727	0,656	15,74
IMD49	14	23	5	21	0,783	0,609	0,391	0,357	0,510	10,72
IMD15	11	24	7	20	0,708	0,458	0,542	0,636	0,629	12,58
IMD67	18	23	4	22	0,826	0,783	0,217	0,222	0,422	9,28

Εξετάσαμε την αξιολόγηση των μαθημάτων αρχικά από το Unique Pages per CourseID per Session που είναι ποσοτική μετρική. Είναι ποσοτική μετρική επειδή μετρά ακριβώς τον αριθμό περιπτώσεων. Τα μαθήματα με έναν υψηλό αριθμό Unique Pages per CourseID per Session είναι αρκετά δημοφιλή μεταξύ των φοιτητών.

Δεδομένου ότι υπάρχουν μαθήματα με ίδια UPCS (Unique Pages per CourseID per Session) ή με τις τιμές πολύ κοντά η μια στην άλλη, θελήσαμε να ξεκαθαρίσουμε την κατάσταση και να προσθέσουμε μια απολύτως ποιοτική μετρική, που ονομάζεται Activity metric, η οποία συνδυάζει κατάλληλα Enrichment, Interest και Homogeneity με ίσα βάρη. Το συνολικό αποτέλεσμα προκύπτει ως το γινόμενο των UPCS με το Activity metric.

Στο Σχήμα 2 βλέπουμε το δενδρόγραμμα της ιεραρχικής ταξινόμησης.



Σχήμα 2: Δενδρόγραμμα της ιεραρχικής ταξινόμησης

Όπως φαίνεται στην αρχή τα μαθήματα χωρίζονται σε δύο μεγάλους κλάδους. Ο πρώτος περιλαμβάνει τα μαθήματα με τη μεγαλύτερη κινητικότητα (IMD105, IMD35 και Final score πάνω από 100) και ο δεύτερος όλα τα υπόλοιπα. Ο δεύτερος κλάδος χωρίζεται ένα επίπεδο πιο κάτω σε δύο επιμέρους κλάδους, δηλαδή τα μαθήματα χωρίζονται περαιτέρω σε μαθήματα μέτριας και χαμηλής κινητικότητας. Όσο συνεχίζουμε προς τα φύλλα του δέντρου τα μαθήματα χωρίζονται σε επιμέρους ομάδες ανάλογα με την κινητικότητα τους. Η ιεραρχική λοιπόν ταξινόμηση μας επιτρέπει να ομαδοποιήσουμε σε διάφορα επίπεδα τα μαθήματα ανάλογα με τις μετρικές Activity και Final. Επιπλέον, φανερώνει πως η πληροφορία που περιέχουν οι δύο μετρικές είναι σημαντική και μπορεί να χρησιμοποιηθεί για την ανάλυση των δεδομένων.

6. Σχολιασμός

Η κατάταξη των μαθημάτων βασίστηκε στις προαναφερθείσες μετρικές και δείκτες. Αρχικά υπολογίσαμε τις μετρικές: σύνοδοι, σελίδες, μοναδικές σελίδες, μοναδικές σελίδες ανά μάθημα και ανά σύνοδο.

Στην συνέχεια, υπολογίσαμε τον εμπλουτισμό (enrichment), το ενδιαφέρον (interest) και την ομοιογένεια (homogeneity) των μαθημάτων. Χρησιμοποιώντας τους τελευταίους δείκτες, υπολογίσαμε τη συνολική μετρική με το όνομα δραστηριότητα. Τέλος, υπολογίσαμε το συνολικό αποτέλεσμα που καθορίζεται από την τελική κατάταξη των μαθημάτων.

Όπως αναφέραμε στην παράγραφο 4.3, ο δείκτης εμπλουτισμός προσφέρει ένα μέτρο των πόσων μοναδικών σελίδων προβλήθηκαν από τους χρήστες, αντιπροσωπεύοντας πλούσιο εκπαιδευτικό υλικό στο μάθημα. Η πρόταση για εμπλουτισμό των υπάρχουσών σελίδων με περισσότερο υλικό γίνεται, εάν ο εμπλουτισμός είναι μικρότερος από τον μέσο όρο όλων των μαθημάτων.

Όπως αναφέραμε στην παράγραφο 4.3, το ότι δεν υπάρχουν πολλές μοναδικές προβληθείσες σελίδες ανά σύνοδο και ανά μάθημα εκφράζει το χαμηλό ενδιαφέρον για ένα μάθημα, επομένως δεν είναι αρκετά δημοφιλές μεταξύ των φοιτητών. Αυτό οφείλεται είτε ότι οι φοιτητές δεν ήταν ευχαριστημένοι με το εκπαιδευτικό υλικό ή δεν υπάρχουν εκεί πολλές σελίδες που να επισκεφθούν. Σε περίπτωση που οι μοναδικές σελίδες ήταν μικρότερες από τον υπολογισμένο μέσο όρο (π.χ. για τις 39 σειρές μαθημάτων ήταν 85.20) τότε το βασικό μειονέκτημα είναι οι "φτωχές" σελίδες που περιέχει το μάθημα. Διαφορετικά, το πρόβλημα έχει να κάνει με την κακή ποιότητα του εκπαιδευτικού υλικού. Ο δείκτης απογοήτευση είναι πιο αντιπροσωπευτικός αφού εκφράζει το πόσο γρήγορα οι χρήστες δεν συνεχίζουν την προβολή των σελίδων των μαθημάτων.

Όπως αναφέραμε στην παράγραφο 4.3, ομοιογένεια 0 σημαίνει ότι κανένας χρήστης δεν ακολούθησε μοναδικό μονοπάτι και 1 ότι κάθε χρήστης ακολούθησε μοναδικό μονοπάτι. Με βάση την τιμή του δείκτη ομοιογένεια μπορούμε να ταξινομήσουμε τα μαθήματα σε: (i) Στατικά με τιμές μεταξύ 0.05 και 0.15. Αυτές οι τιμές χαρακτηρίζουν μαθήματα με στατικό περιεχόμενο πληροφοριών που ενημερώνεται από τους εκπαιδευτικούς με την χρήση μηχανισμών μεταφόρτωσης (upload) όπως file up loaders, editors. (ii) Δυναμικά ετερογενή με τιμές μεταξύ 0.15 και 0.5. Δηλαδή οι πληροφορίες των μαθημάτων που προστίθενται από τους εκπαιδευτικούς, κατηγοριοποιούνται σε διαφορετικές ομάδες χρηστών και ενημερώνονται τακτικά και από τους εκπαιδευτικούς και από τους χρήστες. Τέτοια παραδείγματα είναι τα blogs, οι ειδήσεις (Newsletters) και εφαρμογές ηλεκτρονικού εμπορίου, όπου οι πληροφορίες διακρίνονται σε ομάδες. (iii) Πτητικά ετερογενή (Volatile heterogeneous) με τιμές μεταξύ 0.5 και 0.7, που περιέχουν πληροφορίες που τροποποιούνται συχνά. Πληροφορίες που μπήκαν πριν από μερικές εβδομάδες θεωρούνται ξεπερασμένες και δεν προβάλλονται από τους χρήστες. Σε τέτοια περιβάλλοντα οι χρήστες συνήθως έχουν δικαιώματα πρόσβασης για να προσθέτουν, διαγράφουν και ενημερώνουν πληροφορίες κατά βούληση, χωρίς κανένα μηχανισμό ελέγχου. Τυπικά παραδείγματα αποτελούν τα φόρουμ και τα wiki των μαθημάτων. Τιμές ομοιογένειας άνω των 0,7 πρέπει να εξεταστούν με σκεπτικισμό και αυτά τα μαθήματα πρέπει να ελεγχθούν για καταστάσεις λάθους.

7. Συμπεράσματα

Τα κύρια πλεονεκτήματα της προτεινόμενης μεθοδολογίας είναι ότι: (i) Χρησιμοποιεί τις τεχνικές εξόρυξης δεδομένων για την αξιολόγηση των μαθημάτων, (ii) προτείνει νέους δείκτες και μετρικές, (iii) μπορεί να προσαρμοστεί εύκολα σε οποιοδήποτε LMS.

Πάντως κάποια μελλοντική εργασία θα απαιτηθεί, προκειμένου να αυτοματοποιηθεί η διαδικασία. Για να ξεπεραστεί αυτός ο περιορισμός αναπτύσσεται ένα plug-in εργαλείο το οποίο θα αυτοματοποιεί τα προηγούμενα βήματα. Αυτό το εργαλείο θα τρέχει περιοδικά (κάθε μήνα) και θα στέλνει μήνυμα μέσω ηλεκτρονικού ταχυδρομείου με τις προτάσεις του στους εκπαιδευτικούς. Παρόμοια τακτική εφαρμόστηκε αρχικά (Feng et al, 2005), όπου μετά από μακρόχρονη θεώρηση, οι εκπαιδευτικοί ενημερώθηκαν αυτόματα μέσω ηλεκτρονικού ταχυδρομείου για την ποιότητα του περιεχομένου των μαθημάτων τους. Η εφαρμογή του προτεινόμενου εργαλείου θα βοηθήσει τους εκπαιδευτικούς να βελτιώσουν την ποιότητα των μαθημάτων τους με τη χρήση του κατάλληλου οδηγού GUI.

Επομένως ο στόχος της εργασίας ικανοποιείται, αφού οι εκπαιδευτικοί μπορούν να ωφεληθούν από την εφαρμογή της μεθοδολογίας με τις μετρικές αξιολόγησης και την κατάταξη των μαθημάτων. Αυτή η κατάταξη μπορεί να οδηγήσει τους εκπαιδευτικούς στο να βελτιώσουν το περιεχόμενο των μαθημάτων τους ως προς την αποδοτικότητα τους, μετά από τις προτάσεις που προέρχονται από τη σύγκριση με μαθήματα άλλων εκπαιδευτικών. Η αποδοτικότητα του εκπαιδευτικού περιεχομένου, θα μπορούσε να δώσει στους σπουδαστές την ευκαιρία της ασύγχρονης μελέτης μαθημάτων με βέλτιστο εκπαιδευτικό υλικό.

Ευχαριστίες

Θέλουμε να ευχαριστήσουμε το εκπαιδευτικό προσωπικό του Τμήματος Διαχείρισης Πληροφοριών του ΤΕΙ Καβάλας για την πρόσβαση στα δεδομένα της πλατφόρμας OPEN eClass και ειδικότερα τον καθηγητή Χατζή Βασίλειο. Επίσης θα θέλαμε να ευχαριστήσουμε τον καθηγητή εφαρμογών Μαρδύρη Βασίλειο για την τεχνολογική βοήθειά του στην εγκατάσταση του απαραίτητου λογισμικού στον web server του Τμήματος Διαχείρισης Πληροφοριών του ΤΕΙ Καβάλας για τις ανάγκες της εργασίας.

Αναφορές - Βιβλιογραφία

- aTutor, 2010, Available at <http://www.atutor.ca>.
- Avouris, N., V. Komis, G. Fiotakis, M. Margaritis, and G. Voyiatzaki, 2005, "Logging of fingertip actions is not enough for analysis of learning activities," Proc. Workshop Usage Analysis in learning systems (AIED'05), Amsterdam.
- Baraglia, R., and P. Palmerini, 2002, "SUGGEST : A Web Usage Mining System," In Proc. of IEEE International Conference on Information Technology: Coding and Computing.
- Baruque, C., A. Amaral. A. Barcellos, J. Freitas, and C. Longo, 2007, "Analysing Users' Access Logs in Moodle to Improve e Learning," In Proc. of the ACM Euro American conference on Telematics and information systems.
- Binali, H.H., Wu Chen, V. Potdar, 2009, "Emotion detection in E-learning using opinion mining techniques". Digital Ecosystems and Technologies, 3rd IEEE International Conference, 259 - 264.
- Bing L., 2007, "Web Data Mining: Exploring Hyperlinks, Contents and Usage Data," Springer.
- BlackBoard, 2010, Available at <http://www.blackroad.com>.
- Claroline, 2010, Available at <http://www.claroline.net>.
- Cooley R. W., 2000, "Web Usage Mining: Discovery and Application of Interesting Patterns from Web data," PhD Thesis, Dept of Computer Science, University of Minnesota.
- Cooley, R., B., Mobasher, and J. Srivastava, 1997, "Web Mining: Information and Pattern Discovery on the World Wide Web," *IEEE Computer*, 558-566.
- Feldman, R., and J. Sanger, 2006, "The Text Mining Handbook," Cambridge University Press.
- Feng, M. and N.T. Heffernan, 2005, "Informing Teachers Live about Student Learning: Reporting in the Assistent System," In Proc. of 12th Annual Conference on Artificial Intelligence in Education 2005, Amsterdam.
- Galeas P, 2009, "Web mining", Available at <http://www.galeas.de/webmining.html>.
- Ghani, R., S. Slattery, and Y. Yang, 2001, "Hypertext Categorization using Hyperlink Patterns and Meta Data," In Proc. of the Eighteenth International Conference on Machine Learning.
- GISMO, 2010, Available at <http://gismo.sourceforge.net/>.
- GNU, 2010, "The GNU General Public License," Available at <http://www.gnu.org>.
- Hand, D., H. Mannila, and P. Smyth, 2001, "Principles of Data Mining," MIT Press, Cambridge.
- Ilias, 2010, Available at <http://www.ilias.de>.
- Kantardzic, M., 2003, "Data Mining: Concepts, Models, Methods, and Algorithms," JohnWiley & Sons, ISBN 0471228524.
- Kazanidis, I., S. Valsamidis, T. Theodosiou, S. Kontogiannis, 2009, "Proposed framework for data mining in e-learning: The case of Open e-Class," Proc. IADIS Applied Computing, 254-258.
- Kosala, R., and H. Blockeel, 2000, "Web Mining Research: A Survey," *SIGKDD Explorations*, 2(1), 1-15.
- Markos, A., G. Menexes, and I. Papadimitriou, 2009, "The CHIC Analysis v1.0 Software," International Federation of Classification Societies 2009 Conference - IFCS09, Dresden, Germany.
- Mazza, R. and V. Dimitrova, 2007, "CourseVis: A graphical student monitoring tool for supporting instructors in web-based distance courses," *Journal of Human-Computer Studies*, 65(2), 125-139.
- R. Mazza and C. Milani, 2004, "GISMO: a Graphical Interactive Student Monitoring Tool for Course Management Systems". Proc. Technology Enhanced Learning International Conference (TEL 04). Milan, 18-19 Nov. 2004.

- Mobasher, B., J. Namit, S. Eui-Hong Han and J. Srivastava, 1996, "Web Mining: Pattern Discovery from World Wide Web Transactions," Tech. Report TR 96-060, University of Minnesota, Dept. of Computer Science, Minneapolis.
- Mobasher, B., R. Cooley, and J. Srivastava, 1997, "Web Mining: Information and Pattern Discovery on the World Wide Web," In Proc. of the 9th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'97).
- Moodle, 2010, Available at <http://moodle.org>.
- Mostow, J., J. Beck, H. Cen, A. Cuneo, E. Gouvea, and C. Heiner, 2007, "An Educational Data Mining Tool to Browse Tutor-Student Interactions: Time Will Tell!" Proc. Workshop on educational data mining, 2007, 15-22.
- Myszkowski, P.B., H. Kwaśnicka, and U. Markowska-Kaczmar, 2008, "Data Mining techniques in e-learning CelGrid system," In Proc. of Computer Information Systems and Industrial Management Applications CISIM '08, 315-319.
- Nagi, K. and P. Suesawaluk, 2008, "Research Analysis of Moodle Reports to gauge the level of interactivity in elearning courses at Assumption University, Thailand," In Proc. of ICCCE International Conference on Computer and Communication Engineering, 772-776.
- Nasraoui, O., 2000, "World Wide Web Personalization," Invited chapter in Encyclopedia of Data Mining and Data Warehousing.
- OPEN eCLASS, 2010, Available at <http://eclass.gunet.gr>.
- Perkowitz, M. and O. Etzioni, 1999, "Adaptive web sites: Conceptual cluster mining," In Proc of the International Joint Conference on Artificial Intelligence, 264-269.
- Pierrakos, D., G. Paliouras, C. Papatheodorou, C.D. Spyropoulos, 2003, "Web usage mining as a tool for personalization: a survey," *User modeling and user adapted interaction journal*, 13(4), 311-372.
- Romero, C., and S. Ventura, 2007, "Educational data mining: A survey from 1995 to 2005," Elsevier journal of Expert Systems with Applications.
- Romero, C., S. Ventura, and E. Garcea, 2008, "Data Mining in course management systems: Moodle case study and tutorial," *Computers & Education*, 51(1), 368-384.
- Rong, G., Z. Miaoliang, Z. Liying, Z. Ningning, 2008, "Interest mining in virtual learning environments," *Emerald Journal of Online Information Review*, 32(2), 133 - 146.
- SCORM, 2010, Available at <http://www.adlnet.gov/Technologies/scorm>
- Secondlife, 2010, Available at <http://secondlife.com>.
- Spiliopoulou M., 1999, "Data mining for the web," In Principles of Data Mining and Knowledge Discovery, Second European Symposium, PKDD '99, 588-589.
- Srivastava, J., R. Cooley, M. Deshpande and Tan Pang-Ning, 2000, "Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data," *ACM SIGKDD Explorations Newsletter*, 1(2), 12-23.
- Thelwall M., 2004, "Link Analysis: An Information Science Approach", Academic Press.
- TopClass, 2010, Available at <http://www.topclass.nl>.
- Ueno, M., 2004, "Data mining and text mining technologies for collaborative learning in an ILMS 'Samurai'," In Proc. of IEEE International Conference on Advanced Learning Technologies, 1052-1053.
- Virtual-U, 2010 Available at <http://www.vu.org>.
- WebCT, 2008, Available at <http://www.webct.com>.
- Witten I., and F. Eibe, 2000, "Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations," Morgan Kaufmann.
- Yan, T. W., M. Jacobsen, H. Garcia-Molina, and J. Umeshwar, 1996, "From user access patterns to dynamic hypertext linking," In Proc. of the Fifth International World Wide Web Conference.
- Yang, Y., S. Slattery, Ghani R., 2002, "A Study of Approaches to Hypertext Categorization," *Journal of Intelligent Information Systems*.

- Zaiane O.R., 2001, "Web Usage mining for a better web-based learning environment," In Proc. of Conference on Advanced Technology for Education.
- Zaiane, O.R., M. Xin, and J. Han, 1998, "Discovering Web access patterns and trends by applying OLAP and data mining technology on Web logs," *IEEE*, 19-29.
- Zorrilla, E., and E. Alvarez, 2008, "MATEP: Monitoring and Analysis Tool for e-Learning Platforms," In Proc. of the Eighth IEEE International Conference on Advanced Learning Technologies, 611-613.