



ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΔΥΤΙΚΗΣ ΕΛΛΑΔΑΣ

ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ ΚΑΙ ΟΙΚΟΝΟΜΙΑΣ

ΤΜΗΜΑ ΛΟΓΙΣΤΙΚΗΣ ΚΑΙ ΧΡΗΜΑΤΟΟΙΚΟΝΟΜΙΚΗΣ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

« Η εξόρυξη δεδομένων (data mining) και οι εφαρμογές της στη χρηματοοικονομική λογιστική και ελεγκτική »

ΣΠΟΥΔΑΣΤΕΣ: ΤΣΑΡΟΥΧΑΣ ΠΑΝΑΓΙΩΤΗΣ (Α.Μ.15592)

ΧΡΙΣΤΟΠΟΥΛΟΣ ΦΙΛΙΠΠΟΣ (Α.Μ.15169)

ΕΠΟΠΤΕΥΟΥΣΑ ΚΑΘΗΓΗΤΡΙΑ:ΓΡΙΒΟΚΩΣΤΟΠΟΥΛΟΥ ΦΩΤΕΙΝΗ

ΜΕΣΟΛΟΓΓΙ 2018

Ευχαριστίες

Η παρακάτω εργασία που εκπονήθηκε στο πλαίσιο των προπτυχιακών σπουδών στο τμήμα Λογιστικής και Χρηματοοικονομικής του Α.Τ.Ε.Ι. Δυτικής Ελλάδας. Θα θέλαμε να εκφράσουμε τις θερμές μας ευχαριστίες σε όλους αυτούς που συνέβαλαν στο να ολοκληρώσουμε τις υποχρεώσεις αλλά και τη παρούσα πτυχιακή μας εργασία.

Πρωτίστως, θα θέλαμε να ευχαριστήσουμε την επιβλέπουσα καθηγήτρια μας, Κ. Γριβοκωστοπούλου Φωτεινή καθηγήτρια του τμήματος Λογιστικής και Χρηματοοικονομικής του Α.Τ.Ε.Ι. Δυτικής Ελλάδας για την πολύτιμη βοήθειά της αλλά και ενθάρρυνση που συντέλεσε ακέραια στην ολοκλήρωση της εργασίας μας. Επίσης, ευχαριστούμε τα μέλη της επιτροπής για την πολύτιμη συνεισφορά τους.

Περίληψη

Τα ηλεκτρονικά συστήματα λογιστικής και ελεγκτικής τα τελευταία χρόνια έχουν δει μια αύξηση της πολυπλοκότητας λόγω του ανταγωνιστικού οικονομικού περιβάλλοντος, αλλά με τη βοήθεια λύσεων ανάλυσης δεδομένων όπως το OLAP και το Data Mining μπορεί να είναι μια πολυδιάστατη ανάλυση δεδομένων, μπορεί να ανιχνεύσει την απάτη και να ανακαλύψει τη γνώση που κρύβεται στα δεδομένα, διασφαλίζοντας ότι αυτές οι πληροφορίες είναι χρήσιμες για τη λήψη αποφάσεων στο πλαίσιο του οργανισμού. Τα εργαλεία εξόρυξης δεδομένων είναι σημαντικά στην ελεγκτική και τη λογιστική. Τα τελευταία χρόνια, ο όγκος και η πολυπλοκότητα των λογιστικών συναλλαγών σε μεγάλους οργανισμούς έχουν αυξηθεί δραματικά. Για τον έλεγχο τέτοιων οργανισμών, οι ελεγκτές πρέπει συχνά να ασχολούνται με ογκώδη δεδομένα και με αρκετά περίπλοκη δομή των δεδομένων. Συνεπώς, οι ελεγκτές δεν μπορούν πλέον να βασίζονται μόνο σε αναφορές ή συνοπτικά εργαλεία στη διαδικασία ελέγχου. Αντίθετα, επιπλέον εργαλεία όπως οι τεχνικές εξόρυξης δεδομένων που μπορούν να εξάγουν αυτόματα σημαντικές πληροφορίες από ένα τεράστιο όγκο δεδομένων μπορεί να είναι πολύ χρήσιμες για τους ελεγκτές. Παρόλο που η υιοθέτηση τεχνικών εξόρυξης δεδομένων στις διαδικασίες ελέγχου είναι ένας σχετικά νέος τομέας, η εξόρυξη δεδομένων έχει αποδειχθεί ότι είναι οικονομικά αποδοτική σε πολλές επιχειρηματικές εφαρμογές που σχετίζονται με τον έλεγχο. Η ταξινόμησή τους και οι ικανότητες πρόβλεψης τους επιτρέπουν να χρησιμοποιηθούν για τους σκοπούς της πρόβλεψης της πτώχευσης, της εξέλιξης, της ανησυχίας, της πρόβλεψης οικονομικής δυσπραγίας, της ανίχνευσης της απάτης, της εκτίμησης του πιστωτικού κινδύνου και της πρόβλεψης των επιδόσεων των εταιρειών. Οι πληροφορίες αυτές χωρίς τη χρήση εργαλείων εξόρυξης δεδομένων θα ήταν πολύ δύσκολο να επιτευχθούν από τους ελεγκτές. Οι πληροφορίες που λαμβάνονται από τη διαδικασία εξόρυξης δεδομένων έχουν το πλεονέκτημα ότι απαντούν μόνο στο ερώτημα για το τι συμβαίνει, αλλά ταυτόχρονα υποστηρίζουν και δείχνουν γιατί συμβαίνουν ορισμένα πράγματα. Οπότε σκοπός της εργασίας είναι να αναφερθούν πληροφορίες και στοιχεία για την εξόρυξη δεδομένων (data mining) και τις εφαρμογές της στη χρηματοοικονομική λογιστική και ελεγκτική.

Λέξεις κλειδιά : Εξόρυξη δεδομένων, ελεγκτική, χρηματοοικονομική λογιστική.

Summary

In recent years, electronic accounting and auditing systems have seen an increase in complexity due to a competitive economic environment, but with the help of data analysis solutions such as OLAP and Data Mining can be a multidimensional data analysis, it can detect fraud and discover the knowledge hidden in the data, ensuring that this information is useful for making decisions within the organization. Data mining tools are important in auditing and accounting. In recent years, the volume and complexity of accounting transactions in large organizations have increased dramatically. To control such organizations, controllers often have to deal with bulky data and a rather complex data structure. Consequently, auditors can no longer rely only on reports or summary tools in the audit process. Conversely, additional tools such as data mining techniques that can automatically extract information from a large volume of data can be very useful. Although the adoption of data mining techniques in control procedures is a relatively new field, data mining has proven to be cost effective in many audit-related business applications. Their classification and forecasting capabilities allow them to be used for the purposes of predicting bankruptcy, worrying and predicting financial distress, fraud detection, credit risk assessment, and company performance forecasting. Information without the help of data mining tools would be very difficult to achieve. The information extracted from the data mining process has the advantage of answering only the question of what is happening, but at the same time supporting and showing why certain things happen. So the purpose of the thesis is to provide information and data on data mining and its applications in financial accounting and auditing.

Key words: Data mining, audit, financial accounting.

Περιεχόμενα

Ευχαριστίες	2
Περίληψη	3
Abstract	Σφάλμα! Δεν έχει οριστεί σελιδοδείκτης.
Περιεχόμενα.....	6
Εισαγωγή	8
Κεφάλαιο 1 ^ο : Τα χαρακτηριστικά της γενικής λογιστικής	10
1.1 Ο ορισμός της γενικής λογιστικής	10
1.2 Η ιστορία και η ετυμολογία της γενικής λογιστικής.....	12
1.3 Τα θέματα της γενικής λογιστικής	14
1.4 Οι οργανισμοί.....	18
1.5 Η εκπαίδευση και τα προσόντα.....	20
1.6 Η λογιστική έρευνα	22
1.7 Το λογιστικό σύστημα πληροφοριών.....	24
1.8 Τα λογιστικά σκάνδαλα	25
Κεφάλαιο 2ο : Τα χαρακτηριστικά της Ελεγκτικής.....	27
2.1 Εισαγωγή στον τομέα της Ελεγκτικής.....	Σφάλμα! Δεν έχει οριστεί σελιδοδείκτης.
2.2 Η ιστορία και η ανάπτυξη της Ελεγκτικής.....	Σφάλμα! Δεν έχει οριστεί σελιδοδείκτης.
2.3 Το αντικείμενο του ελέγχου και οι στόχοι της Ελεγκτικής.....	39
2.4 Οι διακρίσεις στην Ελεγκτική	31
<u>2.5 Κλάδοι της Ελεγκτικής.....</u>	<u>34</u>
Κεφάλαιο 3ο : Η εξόρυξη δεδομένων (Data Mining).....	36
3.1 Ο ορισμός της εξόρυξης δεδομένων	36
3.2 Η ετυμολογία.....	39
3.3 Η έρευνα και τα πρότυπα	41

3.4 Η προστασία των προσωπικών δεδομένων και η ηθική	43
3.5 Ο Νόμος περί πνευματικών δικαιωμάτων.....	47
3.6 Τα λογισμικά προγράμματα για την εξόρυξη δεδομένων.....	49
Κεφάλαιο 4ο : Το λογισμικό για την εξόρυξη δεδομένων Weka	52
4.1 Εισαγωγή.....	52
4.2 Οι διεπαφές του χρήστη	55
4.3 Τα εργαλεία της παλινδρόμησης και τα πακέτα επέκτασης του λογισμικού.....	57
4.4 Τα σχετικά εργαλεία του λογισμικού.....	58
Επίλογος.....	59
Βιβλιογραφία	60

Εισαγωγή

Ο έλεγχος είναι ένα σχετικά αρχαϊκό πεδίο και οι ελεγκτές αντιμετωπίζονται συχνά ως ασήμαντα άτομα. Αυτό δεν είναι πλέον αλήθεια. Τα τελευταία χρόνια, οι ελεγκτές αναγνώρισαν τη δραματική αύξηση του όγκου των συναλλαγών και την πολυπλοκότητα των λογιστικών και μη λογιστικών στοιχείων των πελατών τους. Κατά συνέπεια, τα ηλεκτρονικά εργαλεία, όπως το γενικευμένο λογισμικό ελέγχου, χρησιμοποιούνται όλο και περισσότερο για να συμπληρώσουν την παραδοσιακή διαδικασία χειρωνακτικού ελέγχου.

Η εμφάνιση των συστημάτων ERP (Enterprise Resource Planning), με την έννοια της ενοποίησης όλων των λειτουργικών λειτουργιών, προκειμένου να αυξηθεί η κερδοφορία ενός οργανισμού στο σύνολό του, καθιστά το λογιστικό σύστημα πλέον απλό σύστημα χρεώσεων και πιστώσεων. Αντ' αυτού, είναι ο κεντρικός καταχωρητής όλων των λειτουργικών δραστηριοτήτων.

Παρόλο που μπορεί να υποστηριχθεί ποια είναι ή που δεν είναι, η λογιστική συναλλαγή, εξακολουθεί να περιέχει πολύτιμες πληροφορίες. Είναι ευθύνη των ελεγκτών να ελέγχουν επαρκή ποσότητα συναλλαγών που καταγράφονται στις βάσεις δεδομένων του πελάτη προκειμένου να αποκτήσουν επαρκή αποδεικτικά στοιχεία βάσει των οποίων μπορεί να στηριχθεί μια ελεγκτική γνώμη και να διασφαλίσουν ότι δεν υπάρχει κίνδυνος.

Το ποσό και η πολυπλοκότητα των λογιστικών συναλλαγών αυξήθηκαν εντυπωσιακά λόγω της καινοτομίας του ηλεκτρονικού εμπορίου, των ηλεκτρονικών πληρωμών και άλλων συσκευών υψηλής τεχνολογίας. Τα ηλεκτρονικά αρχεία έχουν γίνει πιο συνηθισμένα. Ως εκ τούτου, ο έλεγχος μέσω διαδικτύου είναι όλο και πιο δύσκολος, πόσο μάλλον η χειρωνακτική πρόσβαση. Παρόλο που αυτές οι πολύπλοκες λογιστικές συναλλαγές μπορούν τώρα να παρουσιαστούν με πιο ολοκληρωμένο τρόπο χρησιμοποιώντας το βελτιωμένο γενικό λογισμικό ελέγχου, απαιτούν ακόμα

τους ελεγκτές να κάνουν υποθέσεις, να πραγματοποιούν αναλύσεις και να ερμηνεύουν τα αποτελέσματα.

Μερικά μηχανογραφημένα εργαλεία που χρησιμοποιούνται σήμερα επιτρέπουν στους ελεγκτές να εξετάζουν τα δεδομένα μιας εταιρείας σε συγκεκριμένες προκαθορισμένες μορφές εκτελώντας ποικίλες εντολές ερωτήματος αλλά όχι να εξάγουν πληροφορίες από αυτά τα δεδομένα, ειδικά όταν αυτές οι πληροφορίες είναι άγνωστες και κρυφές. Οι ελεγκτές χρειάζονται κάτι περισσότερο από τα εργαλεία παρουσίασης για να ενισχύσουν τη διερεύνησή τους για γεγονότα ή απλά για ουσιώδη ζητήματα.

Από την άλλη πλευρά, οι τεχνικές εξόρυξης δεδομένων έχουν βελτιωθεί με την πρόοδο της τεχνολογίας βάσεων δεδομένων. Τις τελευταίες δύο δεκαετίες, η βάση δεδομένων έχει γίνει κοινός τόπος στις επιχειρήσεις. Ωστόσο, η ίδια η βάση δεδομένων δεν ωφελεί άμεσα την εταιρεία, προκειμένου να επωφεληθεί η ίδια, η αφθονία των δεδομένων μετατρέπεται σε χρήσιμες πληροφορίες.

Έτσι, τα εργαλεία εξόρυξης δεδομένων που διευκολύνουν την εξαγωγή δεδομένων και την ανάλυση δεδομένων έχουν λάβει μεγαλύτερη προσοχή. Φαίνεται ότι υπάρχουν ευκαιρίες για σύγκλιση του ελέγχου και της εξόρυξης δεδομένων. Ο έλεγχος χρειάζεται ένα μέσο για την αποκάλυψη ασυνήθιστων προτύπων συναλλαγών και η εξόρυξη δεδομένων μπορεί να ικανοποιήσει αυτήν την ανάγκη.

Κεφάλαιο 1^ο : Τα χαρακτηριστικά της γενικής λογιστικής

1.1 Ο ορισμός της γενικής λογιστικής

Η λογιστική είναι η μέτρηση, η επεξεργασία και η επικοινωνία οικονομικών πληροφοριών σχετικά με οικονομικές οντότητες, όπως επιχειρήσεις και εταιρείες. Ο σύγχρονος τομέας όπως τον γνωρίζουμε ιδρύθηκε από τον Ιταλό μαθηματικό Luca Pacioli το 1494 (Agoglia, et al. 2011).

Η λογιστική, που αναφέρεται και «γλώσσα της επιχείρησης», αναλύει τα αποτελέσματα των οικονομικών δραστηριοτήτων ενός οργανισμού και μεταφέρει αυτές τις πληροφορίες σε διάφορους χρήστες, συμπεριλαμβανομένων των επενδυτών, των πιστωτών, της διοίκησης και των ρυθμιστικών αρχών που ανήκει. Οι όροι "λογιστική" και "χρηματοοικονομική αναφορά" χρησιμοποιούνται συχνά ως συνώνυμα.

Η λογιστική μπορεί να χωριστεί σε διάφορους τομείς, όπως είναι η οικονομική λογιστική, η λογιστική διαχείρισης, ο εξωτερικός έλεγχος, η φορολογική λογιστική και η λογιστική του κόστους. Τα λογιστικά πληροφοριακά συστήματα έχουν σχεδιαστεί για να υποστηρίζουν διάφορες λογιστικές λειτουργίες και συναφείς δραστηριότητες (Alon & Dwyer, 2014).

Η χρηματοοικονομική λογιστική επικεντρώνεται στην αναφορά οικονομικών πληροφοριών ενός οργανισμού, συμπεριλαμβανομένης της σύνταξης οικονομικών καταστάσεων, σε εξωτερικούς χρήστες των πληροφοριών, όπως επενδυτές, ρυθμιστές και προμηθευτές και η λογιστική διαχείρισης επικεντρώνεται στη μέτρηση, ανάλυση και υποβολή στοιχείων για εσωτερική χρήση από τη διοίκηση.

Η καταγραφή των χρηματοοικονομικών συναλλαγών, έτσι ώστε οι περιλήψεις των χρηματοοικονομικών στοιχείων να μπορούν να παρουσιάζονται στις οικονομικές εκθέσεις, είναι γνωστή ως χρηματοοικονομική λογιστική, εκ των οποίων η

χρηματοοικονομική διπλής λογιστικής είναι το πιο κοινό σύστημα που χρησιμοποιείτε.

Η λογιστική διευκολύνεται από λογιστικούς οργανισμούς όπως οι πρότυποι, οι εταιρείες λογιστικής και οι επαγγελματικοί φορείς. Οι οικονομικές καταστάσεις ελέγχονται συνήθως από τις λογιστικές επιχειρήσεις και είναι ετοιμασμένες σύμφωνα με τις γενικά αποδεκτές λογιστικές αρχές (GAAP) (Armstrong, et al. 2010).

Τα GAAP καθορίζονται από διάφορους οργανισμούς τυποποίησης, όπως το Συμβούλιο Δημοσιονομικών Λογιστικών Προτύπων (FASB) στις Ηνωμένες Πολιτείες και το Συμβούλιο Δημοσιονομικής Πληροφόρησης στο Ηνωμένο Βασίλειο. Από το 2012, "όλες οι μεγάλες οικονομίες" πρόκειται να συγκλίνουν ή να υιοθετήσουν τα Διεθνή Πρότυπα Χρηματοοικονομικής Πληροφόρησης (ΔΠΧΠ).

1.2 Η ιστορία και η ετυμολογία της γενικής λογιστικής

Η ιστορία της λογιστικής είναι χιλιάδων χρόνων και μπορεί να ανιχνευθεί στους αρχαίους πολιτισμούς. Η πρώιμη εξέλιξη της λογιστικής χρονολογείται από την αρχαία Μεσοποταμία και συνδέεται στενά με τις εξελίξεις στο γράψιμο, την καταμέτρηση και το χρήμα.

Υπάρχουν επίσης στοιχεία για πρώιμες μορφές λογιστικής στο αρχαίο Ιράν, από τους αρχαίους Αιγυπτίους και τους Βαβυλώνιους. Ακόμη και μέχρι τον αυτοκράτορα Αύγουστο που η ρωμαϊκή κυβέρνηση είχε πρόσβαση σε λεπτομερείς οικονομικές πληροφορίες (Atwood, et al. 2011).

Η χρηματοοικονομική διπλής λογιστικής αναπτύχθηκε στη μεσαιωνική Ευρώπη και η λογιστική κατανεμήθηκε σε λογιστική και λογιστική διαχείρισης με την ανάπτυξη των μετοχικών εταιρειών. Η πρώτη μελέτη που έγινε σε σύστημα λογιστικής διπλής εισόδου δημοσιεύθηκε στην Ιταλία από τον Luca Pacioli.

Η λογιστική άρχισε να γίνεται ένα οργανωμένο επάγγελμα τον 19ο αιώνα, ενώ τοπικοί επαγγελματικοί φορείς στην Αγγλία ξεκίνησαν την ενοποίηση τους για να ιδρύσουν το Ινστιτούτο Ορκωτών Λογιστών στην Αγγλία και την Ουαλία το 1880.

Η λογιστική και η λογιστική διαχείρισης χρησιμοποιήθηκαν στη Μεγάλη Βρετανία από τα μέσα του 1800 και προέρχονται από τις λέξεις που συνέταξαν και λογιστικοποίησαν τον 18ο αιώνα (Atwood, et al. 2011).

Στη μέση αγγλική γλώσσα (που χρησιμοποιείται κατά προσέγγιση μεταξύ του 12ου και του τέλους του 15ου αιώνα) το ρήμα «να λογαριάσει» είχε τη μορφή της καταγγελίας, η οποία προήλθε από την παλαιά γαλλική λέξη *aconter* η οποία με τη σειρά της συνδέεται με τη λέξη λατινική λέξη *computare*, που σημαίνει "να υπολογίζω".

Η βάση του computare είναι το putare, το οποίο "προοριζόταν διαφορετικά να σταματήσει, να καθαρίσει, να διορθώσει έναν λογαριασμό, συνεπώς, να μετρήσει ή να υπολογίσει, καθώς και να σκεφτεί" (Ball, et al. 2015).

Η λέξη "λογιστής" προέρχεται από τη γαλλική λέξη compteur, η οποία προέρχεται επίσης από την ιταλική και τη λατινική λέξη computare. Η λέξη ήταν παλαιότερα γραμμένη στα αγγλικά ως "συμβιβαστική", αλλά κατά τη διάρκεια του χρόνου, η λέξη, η οποία ήταν πάντα προφέρεται από την πτώση του "p", μεταβλήθηκε σταδιακά τόσο στην προφορά όσο και στην ορθογραφία στην μορφή που έχει σήμερα.

Η λογιστική έχει οριστεί ως η διατήρηση ή η προετοιμασία των οικονομικών αρχείων της οικονομικής οντότητας, η ανάλυση, η επαλήθευση και η αναφορά τέτοιων λογιστικών εγγραφών και "οι αρχές και οι διαδικασίες λογιστικής", αναφέρονται επίσης στην εργασία του λογιστή. Η λογιστική αναφέρεται στην κατοχή ή το επάγγελμα του λογιστή, ιδίως στα βρετανικά αγγλικά (Ballas, et al. 2010).

1.3 Τα θέματα της γενικής λογιστικής

Η λογιστική έχει διάφορα υποπεδία και τομείς, συμπεριλαμβανομένων των συστημάτων οικονομικής λογιστικής, της λογιστικής διαχείρισης, του ελέγχου, της φορολογίας και της λογιστικής πληροφόρησης (Balsmeier & Vanhaverbeke, 2016).

Χρηματοοικονομική Λογιστική

Η χρηματοοικονομική λογιστική εστιάζει στην ενημέρωση οικονομικών πληροφοριών ενός οργανισμού σε εξωτερικούς χρήστες των πληροφοριών, όπως είναι οι επενδυτές, οι δυνητικοί επενδυτές και οι πιστωτές.

Υπολογίζει και καταγράφει τις επιχειρηματικές συναλλαγές και συντάσσει οικονομικές καταστάσεις για τους εξωτερικούς χρήστες σύμφωνα με τις γενικά αποδεκτές λογιστικές αρχές (GAAP).

Το GAAP, με τη σειρά του, απορρέει από την ευρεία συμφωνία μεταξύ της θεωρίας και της πρακτικής των λογιστικών βιβλίων και την αλλαγή με την πάροδο του χρόνου ώστε να ικανοποιηθούν οι ανάγκες των υπευθύνων λήψης αποφάσεων (Barth & Landsman, 2010).

Η χρηματοοικονομική λογιστική δημιουργεί παρελθούσες εκθέσεις όπως οι οικονομικές καταστάσεις που καταρτίστηκαν στις εκθέσεις του 2006 σχετικά με τις επιδόσεις το 2005, σε ετήσια ή τριμηνιαία βάση, γενικά για τον οργανισμό στο σύνολό του.

Αυτός ο κλάδος της λογιστικής μελετάται επίσης ως μέρος των εξετάσεων του διοικητικού συμβουλίου για να χαρακτηριστεί ως αναλογιστής. Είναι ενδιαφέρον να σημειώσουμε ότι αυτοί οι δύο επαγγελματίες, λογιστές και αναλογιστές, έχουν δημιουργήσει μια κουλτούρα της ύπαρξης archrivals (Beneish, et al. 2015).

Λογιστική διαχείριση

Η λογιστική διαχείρισης επικεντρώνεται στη μέτρηση, ανάλυση και αναφορά πληροφοριών που μπορούν να βοηθήσουν τους διαχειριστές στη λήψη αποφάσεων για την εκπλήρωση των στόχων ενός οργανισμού. Στη λογιστική διαχείρισης, τα εσωτερικά μέτρα και οι εκθέσεις βασίζονται στην ανάλυση κόστους-οφέλους και δεν υποχρεούνται να ακολουθούν τη γενικά αποδεκτή λογιστική αρχή (GAAP) (Bengtsson, 2011).

Το 2014, η CIMA δημιούργησε τις Αρχές Λογιστικής της Global Management (GMAP). Το αποτέλεσμα της έρευνας από 20 χώρες και πέντε ηπείρους, ήταν οι αρχές να στοχεύουν στην καθοδήγηση των βέλτιστων πρακτικών στον τομέα.

Η λογιστική διαχείρισης παράγει εκθέσεις που στρέφονται προς το μέλλον για παράδειγμα, ο προϋπολογισμός για το έτος 2006 καταρτίζεται το 2005, και η χρονική διάρκεια των εκθέσεων ποικίλλει ευρέως.

Οι εκθέσεις αυτές μπορούν να περιλαμβάνουν τόσο χρηματοοικονομικές όσο και μη οικονομικές πληροφορίες και μπορούν, για παράδειγμα, να επικεντρωθούν σε συγκεκριμένα προϊόντα και υπηρεσίες (Botzem, 2014).

Λογιστικός έλεγχος

Ο έλεγχος είναι η επαλήθευση των ισχυρισμών των άλλων σχετικά με την πληρωμή και στο κομμάτι της λογιστικής είναι η «αμερόληπτη εξέταση και αξιολόγηση των οικονομικών καταστάσεων ενός οργανισμού».

Ο έλεγχος των οικονομικών καταστάσεων έχει ως στόχο να εκφράσει ή να απορρίψει γνώμη σχετικά με τις οικονομικές καταστάσεις. Ο ελεγκτής εκφράζει τη γνώμη του σχετικά με την ορθότητα με την οποία οι οικονομικές καταστάσεις παρουσιάζουν την κατάσταση, τα αποτελέσματα των εργασιών και τις ταμειακές ροές μιας οικονομικής οντότητας, σύμφωνα με τη γενικά αποδεκτή λογιστική αρχή (GAAP) και "από κάθε ουσιώδη άποψη". Ένας ορκωτός ελεγκτής πρέπει επίσης να εντοπίσει περιστάσεις

στις οποίες δεν τηρούνται γενικά αποδεκτές λογιστικές αρχές (GAAP) (Botzem & Dobusch, 2012).

Λογιστικά συστήματα πληροφοριών

Ένα σύστημα λογιστικής πληροφόρησης αποτελεί μέρος του συστήματος πληροφοριών ενός οργανισμού που επικεντρώνεται στην επεξεργασία των λογιστικών δεδομένων (Bratton & Cunningham, 2009).

Φορολογική λογιστική

Η φορολογική λογιστική στις Ηνωμένες Πολιτείες έχει ως επίκεντρο την προετοιμασία, ανάλυση και παρουσίαση φορολογικών πληρωμών και φορολογικών δηλώσεων.

Το φορολογικό σύστημα των Ηνωμένων Πολιτειών απαιτεί τη χρήση εξειδικευμένων λογιστικών αρχών για φορολογικούς σκοπούς, οι οποίες μπορεί να διαφέρουν από τις γενικά αποδεκτές λογιστικές αρχές για τη χρηματοοικονομική πληροφόρηση.

Ο φορολογικός νόμος των Ηνωμένων Πολιτειών καλύπτει τρεις βασικές μορφές επιχειρηματικής ιδιοκτησίας (Brüggemann, et al. 2013) :

- A.E.
- O.E.
- E.Π.E.

Τα εταιρικά και προσωπικά εισοδήματα φορολογούνται με διαφορετικούς συντελεστές, οι οποίοι ποικίλλουν ανάλογα με τα επίπεδα εισοδήματος και περιλαμβάνουν ποικίλους οριακούς συντελεστές (φορολογούμενοι για κάθε πρόσθετο

δολάριο εισοδήματος) και μέσους συντελεστές (που ορίζονται ως ποσοστό του συνολικού εισοδήματος) (Byard, et al. 2011).

1.4 Οι οργανισμοί

Επαγγελματικά σώματα

Τα επαγγελματικά όργανα λογιστικής περιλαμβάνουν το Αμερικανικό Ινστιτούτο Πιστοποιημένων Λογιστών (AICPA) και τα άλλα 179 μέλη της Διεθνούς Ομοσπονδίας Λογιστών (IFAC), συμπεριλαμβανομένης της CPA Australia, της Ένωσης Πιστοποιημένων Λογιστών (ACCA) και του Ινστιτούτου Ορκωτών Λογιστών Αγγλία και Ουαλία (ICAEW) (Chalmers, et al. 2011).

Υπάρχουν επίσης πολλοί επαγγελματικοί οργανισμοί για τα υποπεδία των λογιστικών επαγγελμάτων, όπως το Chartered Institute of Accountants (CIMA). Πολλοί από αυτούς τους επαγγελματικούς φορείς προσφέρουν εκπαίδευση και κατάρτιση, συμπεριλαμβανομένων των προσόντων και της διοίκησης για διάφορους λογιστικούς χαρακτηρισμούς, όπως ο πιστοποιημένος δημόσιος λογιστής και ο εξουσιοδοτημένος λογιστής.

Λογιστικές εταιρείες

Ανάλογα με το μέγεθός της, μια εταιρεία μπορεί νομικά να υποχρεούται να ελέγχει τις οικονομικές της καταστάσεις από έναν εξειδικευμένο ελεγκτή και οι λογιστικοί έλεγχοι συνήθως διεξάγονται από λογιστικές επιχειρήσεις (Chen & Sami, 2008).

Οι λογιστικές επιχειρήσεις αυξήθηκαν στις Ηνωμένες Πολιτείες και την Ευρώπη στα τέλη του 19ου και στις αρχές του 20ου αιώνα και μέσω αρκετών συγχωνεύσεων υπήρχαν μεγάλα διεθνή λογιστικά γραφεία από τα μέσα του 20ου αιώνα.

Περαιτέρω μεγάλες συγχωνεύσεις στα τέλη του 20ου αιώνα οδήγησαν στην κυριαρχία της ελεγκτικής αγοράς από τις λογιστικές εταιρίες "Big Five": Arthur Andersen, Deloitte, Ernst & Young, KPMG και PricewaterhouseCoopers. Η κατάρρευση του Arthur Andersen μετά το σκάνδαλο της Enron μείωσε το Big Five στα Big Four (Chen, et al. 2011).

Πρότυποι ρυθμιστές

Οι γενικά αποδεκτές λογιστικές αρχές (GAAP) είναι λογιστικά πρότυπα που εκδίδονται από εθνικούς ρυθμιστικούς φορείς. Επιπλέον, το Συμβούλιο Διεθνών Λογιστικών Προτύπων (IASB) εκδίδει τα Διεθνή Πρότυπα Χρηματοοικονομικής Πληροφόρησης (Δ.Π.Χ.Π.) που εφαρμόζονται σε 147 χώρες (Chen, et al. 2013)

Ενώ τα πρότυπα για τον διεθνή έλεγχο και τη διασφάλιση, τη δεοντολογία, την εκπαίδευση και την λογιστική του δημόσιου τομέα καθορίζονται από ανεξάρτητους πίνακες ρυθμίσεων που υποστηρίζονται από το IFAC.

Το Διεθνές Συμβούλιο Προτύπων Ελέγχου και Διασφάλισης ορίζει τα διεθνή πρότυπα για τον έλεγχο, τη διασφάλιση και τον ποιοτικό έλεγχο. Ο IESBA ορίζει τον διεθνή καταλληλότερο Κώδικα Δεοντολογίας Επαγγελματικών Λογαριασμών, ο Διεθνής Οργανισμός Λογιστικών Εκπαιδευτικών Προτύπων (IAESB) θέτει τα επαγγελματικά πρότυπα λογιστικής εκπαίδευσης και το Διεθνές Συμβούλιο Λογιστικών Προτύπων για τον Δημόσιο Τομέα (IPSASB) ορίζει τα διεθνή λογιστικά πρότυπα για τον δημόσιο τομέα.

Οι οργανισμοί σε μεμονωμένες χώρες μπορούν να εκδίδουν λογιστικά πρότυπα μοναδικά για τις χώρες. Για παράδειγμα, στις Ηνωμένες Πολιτείες, το Συμβούλιο Δημοσιονομικών Λογιστικών Προτύπων (FASB) εκδίδει τις Δηλώσεις Χρηματοοικονομικών Λογιστικών Προτύπων, οι οποίες αποτελούν τη βάση των GAAP των ΗΠΑ, ενώ στο Ηνωμένο Βασίλειο το Συμβούλιο Δημοσιονομικής Αναφοράς (FRC) ορίζει τα λογιστικά πρότυπα. Ωστόσο, από το 2012 "όλες οι μεγάλες οικονομίες" σχεδιάζουν να συγκλίνουν ή να υιοθετήσουν τα ΔΠΧΠ (Chen, et al. 2015).

1.5 Η εκπαίδευση και τα προσόντα

Λογιστικοί βαθμοί

Τουλάχιστον ένα πτυχίο στη λογιστική ή σε σχετικό τομέα κρίνονται απαραίτητα για τις περισσότερες θέσεις λογιστών και ελεγκτών, και ορισμένοι εργοδότες προτιμούν τους υποψήφιους οι οποίοι κατέχουν μεταπτυχιακό τίτλο. Ο βαθμός λογιστικής μπορεί επίσης να απαιτηθεί ή να χρησιμοποιηθεί για την εκπλήρωση των προϋποθέσεων συμμετοχής σε επαγγελματικούς λογιστικούς φορείς (Chen & Khurana, 2016).

Για παράδειγμα, η εκπαίδευση κατά τη διάρκεια ενός λογιστικού πτυχίου μπορεί να χρησιμοποιηθεί για να εκπληρωθεί η απαίτηση που έχει το Αμερικανικό Ινστιτούτο της CPA (AICPA) 150 ωρών, και η συνδεδεμένη συμμετοχή με τον Σύνδεσμο Πιστοποιημένων Λογιστών του Ηνωμένου Βασιλείου είναι διαθέσιμη μετά την απόκτηση πτυχίου στη χρηματοδότηση ή τη λογιστική.

Απαιτείται διδακτορικό δίπλωμα για να μπορέσει να ασκήσει σταδιοδρομία στον ακαδημαϊκό χώρο της λογιστικής, για παράδειγμα να εργαστεί ως καθηγητής πανεπιστημίου στη λογιστική. Ο γιατρός της Φιλοσοφίας (PhD) και ο γιατρός της διοίκησης επιχειρήσεων (DBA) είναι οι πιο δημοφιλείς βαθμοί.

Το διδακτορικό είναι το πιο κοινό πτυχίο για όσους επιθυμούν να ακολουθήσουν μια σταδιοδρομία στον ακαδημαϊκό χώρο, ενώ τα προγράμματα DBA γενικά επικεντρώνονται στον εξοπλισμό στελεχών επιχειρήσεων για επαγγελματικές ή δημόσιες σταδιοδρομίες που απαιτούν ερευνητικές δεξιότητες και προσόντα (Chua & Taylor, 2008).

Επαγγελματικά Προσόντα

Τα επαγγελματικά λογιστικά προσόντα περιλαμβάνουν τους ορισμούς του Ορκωτού Ελεγκτή Λογιστή και άλλα προσόντα, συμπεριλαμβανομένων των πιστοποιητικών

και των διπλωμάτων. Στο Ηνωμένο Βασίλειο, οι ορκωτοί λογιστές του ICAEW υποβάλλονται σε ετήσια εκπαίδευση και δεσμεύονται από τον κώδικα δεοντολογίας του ICAEW και υπόκεινται στις πειθαρχικές διαδικασίες του.

Στις Ηνωμένες Πολιτείες, οι απαιτήσεις για συμμετοχή στο AICPA ως Ορκωτός Ελεγκτής Λογιστής καθορίζονται από το Λογιστικό Συμβούλιο κάθε κράτους και τα μέλη συμφωνούν να συμμορφώνονται με τον Κώδικα Επαγγελματικής Δεοντολογίας και Κανονισμών του AICPA (Chua, et al. 2012).

Στην Ινδία το όργανο λογιστικής Arrex που συγκροτείται από το κοινοβούλιο της Ινδίας είναι το "Ινστιτούτο Ορκωτών Λογιστών της Ινδίας" (ICAI) όπου είναι γνωστό για την αυστηρή εκπαίδευση και τη μεθοδολογία μελέτης για τη χορήγηση του προσόντος.

Η ACCA είναι ο μεγαλύτερος παγκόσμιος φορέας λογιστικής με περισσότερα από 320.000 μέλη και ο οργανισμός παρέχει μια «ροή ΔΠΧΠ» και μια «ροή του Ηνωμένου Βασιλείου». Οι φοιτητές πρέπει να περάσουν συνολικά 14 εξετάσεις, οι οποίες είναι διατεταγμένες σε τρία έγγραφα (Collins, et al. 2012).

1.6 Η λογιστική έρευνα

Η λογιστική έρευνα είναι η έρευνα σχετικά με τις επιπτώσεις των οικονομικών γεγονότων στη λογιστική διαδικασία και τις επιπτώσεις που έχουν οι αναφερόμενες πληροφορίες στα οικονομικά γεγονότα. Περιλαμβάνει ένα ευρύ φάσμα τομέων έρευνας, όπως η χρηματοοικονομική λογιστική, η λογιστική διαχείρισης, ο έλεγχος και η φορολόγηση.

Η λογιστική έρευνα διεξάγεται τόσο από ακαδημαϊκούς ερευνητές όσο και από εξειδικευμένους λογιστές. Οι μεθοδολογίες στην ακαδημαϊκή λογιστική έρευνα μπορούν να ταξινομηθούν σε (Daske, et al. 2008) :

- Αρχειακή έρευνα, η οποία εξετάζει "αντικειμενικά δεδομένα που συλλέγονται από αποθετήρια".
- Πειραματική έρευνα, η οποία εξετάζει τα δεδομένα "ο ερευνητής συγκεντρώθηκε με τη χορήγηση θεραπειών σε υποκείμενα".
- Αναλυτική έρευνα, η οποία "βασίζεται στην πράξη της τυπικής μοντελοποίησης των θεωριών ή της τεκμηρίωσης ιδεών με μαθηματικούς όρους".

Η ταξινόμηση αυτή δεν είναι εξαντλητική. Υπάρχουν και άλλες πιθανές μεθοδολογίες που περιλαμβάνουν τη χρήση μελετών περιπτώσεων, προσομοιώσεις ηλεκτρονικών υπολογιστών και έρευνα πεδίου (Daske, et al. 2013).

Οι εμπειρικές μελέτες γνωστοποιούν ότι τα κορυφαία λογιστικά περιοδικά δημοσιεύουν συνολικά λιγότερα ερευνητικά άρθρα από τα συγκρίσιμα περιοδικά στα οικονομικά και άλλους κλάδους και, κατά συνέπεια, οι μελετητές των λογιστικών βιβλίων είναι σχετικά λιγότερο επιτυχημένοι σε ακαδημαϊκές εκδόσεις από ότι λόγω των διαφορετικών ποσοστών δημοσίευσης μεταξύ λογιστικών και άλλων επιχειρήσεων.

Μια πρόσφατη μελέτη βασισμένη στην ακαδημαϊκή κατάταξη συγγραφέων καταλήγει στο συμπέρασμα ότι η ανταγωνιστική αξία μιας ενιαίας έκδοσης σε ένα κορυφαίο περιοδικό είναι υψηλότερη στη λογιστική και χαμηλότερη στο μάρκετινγκ (Dimitropoulos, et al. 2013).

1.7 Το λογιστικό σύστημα πληροφοριών

Πολλές λογιστικές πρακτικές έχουν γίνει πιο απλές με τη βοήθεια λογισμικού που βασίζεται στον υπολογιστή. Ένα σύστημα προγραμματισμού πόρων για επιχειρήσεις (Enterprise Resource Planning - ERP) χρησιμοποιείται συνήθως για έναν μεγάλο οργανισμό και παρέχει μια ολοκληρωμένη, συγκεντρωτική και ολοκληρωμένη πηγή πληροφόρησης που οι εταιρείες μπορούν να χρησιμοποιήσουν για να διαχειριστούν καλύτερα όλες τις σημαντικές επιχειρηματικές διαδικασίες , από την αγορά έως τον τομέα των ανθρώπινων πόρων της(Durocher & Gendron, 2011).

Τα λογιστικά πληροφοριακά συστήματα έχουν συμβάλει στην μείωση του κόστους συσσώρευσης, αποθήκευσης και αναφοράς διοικητικών λογιστικών πληροφοριών των οργανισμών και έχουν καταστήσει δυνατή την παραγωγή λεπτομερέστερου λογαριασμού όλων των δεδομένων που εισάγονται σε οποιοδήποτε δεδομένο σύστημα.

1.8 Τα λογιστικά σκάνδαλα

Το 2001 σημειώθηκε μια σειρά από απάτες στις οικονομικές πληροφορίες που εμπλέκονταν η Enron, η ελεγκτική εταιρεία Arthur Andersen, η εταιρεία τηλεπικοινωνιών WorldCom, η Qwest και η Sunbeam, μεταξύ άλλων γνωστών εταιρειών (El-Gazzar & Fornaro, 2012).

Τα προβλήματα αυτά ανέδειξαν την ανάγκη που υπήρχε για αναθεώρηση της αποτελεσματικότητας των λογιστικών προτύπων, των κανονισμών ελέγχου και των αρχών εταιρικής διακυβέρνησης.

Σε ορισμένες περιπτώσεις, η διοίκηση χειρίστηκε τα στοιχεία που εμφανίζονται στις οικονομικές εκθέσεις για να δείξει καλύτερη οικονομική απόδοση. Σε κάποιες περιπτώσεις, τα φορολογικά και κανονιστικά κίνητρα ενθάρρυναν την υπερβολική μόχλευση των εταιρειών και τις αποφάσεις να φέρουν έκτακτο και αδικαιολόγητο κίνδυνο (Eng, et al. 2014).

Το σκάνδαλο της Enron συνέβαλε στην ανάπτυξη νέων κανονισμών για τη βελτίωση της αξιοπιστίας της χρηματοοικονομικής πληροφόρησης και την ευαισθητοποίηση του κοινού σχετικά με τη σημασία της ύπαρξης λογιστικών προτύπων που να μπορεί να δείξει την οικονομική πραγματικότητα των επιχειρήσεων και την αντικειμενικότητα και την ανεξαρτησία των ελεγκτικών εταιρειών.

Εκτός από τη μεγαλύτερη αναδιοργάνωση πτώχευσης στην αμερικανική ιστορία, το σκάνδαλο της Enron είναι αναμφισβήτητα η μεγαλύτερη αποτυχία ελέγχου. Περιλάμβανε ένα οικονομικό σκάνδαλο της Enron Corporation και των ελεγκτών της εταιρείας Arthur Andersen, το οποίο αποκαλύφθηκε στα τέλη του 2001 (Fifield, et al. 2011).

Το σκάνδαλο έγινε η αιτία για την διάλυση της εταιρείας Arthur Andersen, η οποία ήταν τότε μία από τις πέντε μεγαλύτερες λογιστικές επιχειρήσεις στον κόσμο. Ύστερα από μια σειρά από αποκαλύψεις που αφορούσαν τις παράτυπες λογιστικές

διαδικασίες που έγιναν καθ 'όλη τη δεκαετία του '90, η Enron κατέθεσε την προστασία της πτώχευσης του κεφαλαίου 11 τον Δεκέμβριο του 2001.

Ένα αποτέλεσμα αυτών των γεγονότων ήταν το πέρασμα του νόμου Sarbanes-Oxley στις Ηνωμένες Πολιτείες το 2002, που είχε ως αποτέλεσμα την πρώτη αναγνώριση δόλιας συμπεριφοράς της Enron.

Η πράξη αυτή εγείρει σημαντικές ποινικές κυρώσεις για την απάτη των κινητών αξιών, την καταστροφή, τη μεταβολή ή την κατασκευή αρχείων σε ομοσπονδιακές έρευνες ή οποιοδήποτε σχέδιο ή απόπειρα εξαπάτησης των μετόχων (Floropoulos, 2006).

Κεφάλαιο 2ο : Τα χαρακτηριστικά της Ελεγκτικής

2.1 Εισαγωγή στον τομέα της ελεγκτικής

Στο παρόν κεφάλαιο θα αναλύσουμε: το εννοιολογικό πλαίσιο της ελεγκτικής ,θα προσδιορίσουμε τους στόχους που έχει η ελεγκτική, θα αναφερθούν τα θέματα που εξετάζει, θα παρουσιάζονται οι διακρίσεις και οι σημαντικότεροι κλάδοι που έχει η ελεγκτική

Ελεγκτική είναι ο κλάδος της οικονομικής των επιχειρήσεων που αναλύει τους κανόνες που πρέπει να υπάρχουν ώστε να γίνεται η συγκέντρωση των τεκμηρίων σχετικά με ορισμένες πληροφορίες ώστε να μπορεί να διαπιστωθεί κατά πόσο υπάρχει συμφωνία μεταξύ των πληροφοριών και κάποιων κριτηρίων ώστε να μπορέσει να διατυπωθεί η σχετική γνώμη από τον ελεγκτή μέσα από την έκθεση ελέγχου. Επομένως εύκολα μπορεί κάποιος να συμπεράνει ότι το βασικό αντικείμενο της ελεγκτικής είναι διαπίστωση της ορθής ή μη παρουσίασης των οικονομικών καταστάσεων, της οικονομικής θέσης και οικονομικών καταστάσεων μιας οικονομικής μονάδας σύμφωνα πάντα με τις γενικώς παραδεκτές λογιστικές αρχές.

Συγκεκριμένα η λογιστική αποτυπώνει τις πράξεις της οικονομικής μονάδας ενώ η ελεγκτική έρχεται να τις ελέγξει ώστε να διαπιστωθούν τυχόν λάθη ή απάτες που έχουν γίνει. Όπως καταλαβαίνουμε δηλαδή στην εποχή μας είναι απαραίτητη η ύπαρξη της ελεγκτικής και ο έλεγχος των ελεγκτών. Παρακάτω θα αναλύσουμε την ιστορική της αναδρομή όπου θα συμπεράνουμε ότι η ύπαρξή της ήταν απαραίτητη ακόμα και σε παλαιότερα έτη.

2.2 Η ιστορία και η ανάπτυξη της Ελεγκτικής

Η αρχική θεσμοθέτηση της ελεγκτικής φαίνεται ότι ήταν απόρροια της αναγκαιότητας για έλεγχο των καθημερινών συναλλαγών και της διαφάνειας στην διαχείριση των δημόσιων χρημάτων. Η ύπαρξη των πρώτων λογιστών-ελεγκτών χρονολογείται στα 3000 π.χ. όπου ο “γραμματέας” χρησιμοποιούσε μια πλάκα για τις μεγάλες συναλλαγές όπου τις σημείωνε και νομιμοποιούσε με την υπογραφή του ο οποίος αντιπροσωπεύει τον σημερινό ελεγκτή.

Η ανάπτυξη της Ελεγκτικής ως ξεχωριστό επαγγελματικό πεδίου στα πλαίσια της οικονομικής και εμπορικής δραστηριότητας των ανθρώπων έγινε μόνο όταν η Λογιστική ανέλαβε με την οργάνωση και ανάπτυξη μετά το Μεσαίωνα και ιδιαίτερα στις Αγγλοσαξονικές κοινωνίες (Dimitropoulos, et al. 2013).

Έτσι βλέπουμε ότι στην Μ. Βρετανία που θεωρείται και η πατρίδα της Ελεγκτικής, να αναγνωρίζεται σύμφωνα με τον εταιρικό νόμο του 1862 η ανάγκη για ακριβή και αξιόπιστη χρηματοοικονομική πληροφόρηση των οικονομικών συνόλων της κοινωνίας μέσα από ένα οργανωμένο λογιστικό σύστημα, καθώς την ανάγκη για ανεξάρτητο έλεγχο των λογαριασμών μικρών και μεγαλύτερων οικονομικών μονάδων. Έπειτα η Λογιστική και η Ελεγκτική ακολούθησαν το δρόμο μιας μεγάλης αναπτυξιακής πορείας του στην Αμερική και στην Δυτική Ευρώπη, όπου το 1912 διαβάζουμε στον R. Montgomery ότι οι κύριοι στόχοι του ελέγχου είναι (Ball & Craig, 2010) :

- Η αποκάλυψη εσκεμμένων ή μη εσκεμμένων λαθών, εξαπατήσεων (προληπτικός έλεγχος).
- Η αποκάλυψη, η καταστολή αβούλητων ή ιδιόβουλων λαθών και εξαπατήσεων (κατασταλτικός έλεγχος).

Η προσπάθεια της μετατόπισης της έμφασης του εξωτερικού ελέγχου από την ανακάλυψη της απάτης και των λαθών διαχείρισης στην εκτίμηση της πραγματικής χρηματοοικονομικής κατάστασης και των αποτελεσμάτων μιας οικονομικής μονάδας συνεχίστηκε, ακόμη και μετά το τέλος του Β' Παγκόσμιου Πολέμου στις

Αγγλοσαξονικές χώρες, οπότε η πλάστιγγα έγειρε τελικά κατά τη δεκαετία του 1960 υπέρ και διαμορφώθηκε ότι ο σκοπός της ελεγκτικής πρέπει να είναι η διαμόρφωση γνώμης περί της ορθής παρουσίας της χρηματοοικονομικής θέσεως και των αποτελεσμάτων της ελεγχόμενης οικονομικής μονάδος (Ball, et al. 2015).

Η πρώτη εμφάνιση οργανωμένου ελεγκτικού επαγγέλματος στην Ελλάδα έγινε το 1956 όπου με την σύσταση του Σώματος Ορκωτών Λογιστών (ΣΟΛ), με τη βοήθεια Άγγλων ορκωτών λογιστών και υπό τον έμμεσο έλεγχο του κράτους. Τούτο είχε και τις ευθύνες αλλά και το προνόμιο της εκπλήρωσης των επιβαλλόμενων διατάξεων περί ελέγχου στον ιδιωτικό τομέα μέχρι της αναδιοργάνωσης του το 1992 με το Π.Δ. 226/92, με το οποίο δόθηκε η ευκαιρία ίδρυσης διάφορων ελεγκτικών εταιρειών σύμφωνα με τα αμερικάνικα και αγγλικά πρότυπα και κατ' εντολή της Ευρωπαϊκής Κοινότητας μέσω της 8^{ης} Οδηγίας.

Μέχρι και το 1995, ο έλεγχος που γινόταν στον ιδιωτικό τομέα θεωρούνταν ελλιπής μέχρι και προκατασκευασμένη γνωμάτευση, που τελικά υπέγραφε ο ελεγκτής. Κύριος λόγος για αυτά τα συμπεράσματα ήταν ότι οι ελεγκτές αυτοί ορίζονταν αλλά και άλλαζαν κατά βούληση της διοίκησης της ελεγχόμενης οικονομικής μονάδας και μόνο, σε αντίθεση με τις Αγγλοσαξονικές χώρες, όπου η αντικατάσταση μιας ελεγκτικής εταιρείας αξιολογούνταν και θα έπρεπε να γίνει δεκτή τόσο από την εταιρεία που θα την αντικαθιστούσε όσο και από το σώμα. Ένας ακόμη σοβαρός λόγος ήταν η κακή διαχείριση των ελεγκτικών εταιρειών και η ελλιπής ή σχεδόν ανύπαρκτη εκπαίδευση που είχαν οι ελεγκτές (Ballas, et al. 2010).

Ακόμη και σήμερα είναι ευάλωτος σε παρόμοιες αδυναμίες ο έλεγχος, ο οποίος γίνεται όχι από πάντα από εκπαιδευμένους επαγγελματίες ορκωτούς ελεγκτές, αλλά και από απλούς απόφοιτους οικονομικών σχολών χωρίς τις απαραίτητες γνώσεις των ορκωτών λογιστών. Επιτρέπεται να γίνεται έλεγχος στις οικονομικές μονάδες εκείνες του άρθρου 3 του Π.Δ. 226/92, παρ. 1δ, οι οποίες μπορεί να μην πληρούν δύο εκ των τριών τουλάχιστον, προϋποθέσεων για έλεγχο από ορκωτούς ελεγκτές οι οποίες έχουν ένα μέσο όρο πενήντα υπαλλήλων, συγκεκριμένο ύψος πωλήσεων και συγκεκριμένο ύψος ενεργητικού. Το ύψος των μεγεθών αυτών των προϋποθέσεων μεταβάλλεται.

2.3 Το αντικείμενο του ελέγχου και οι στόχοι της Ελεγκτικής

Ο προσωπικός σκοπός του ελεγκτή είναι η διαμόρφωση της γνώμης αναφορικά με την σωστή ή μη παρουσίαση της χρηματοοικονομικής καταστάσεων και αποτελεσμάτων χρήσεως της οικονομικής μονάδες στις αντίστοιχες λογιστικές καταστάσεις, σύμφωνα με τις γενικώς παραδεκτές λογιστικές αρχές., παρουσίασης της χρηματοοικονομικής θέσης, δηλαδή την περιουσιακή και κεφαλαιακή διάρθρωση της μονάδας την συγκεκριμένη ημερομηνία καθώς και τις μεταβολές αυτής ως αποτέλεσμα της δραστηριοποίησης της, κατά τη διαρρεύσασα περίοδο, αλλά και άλλων ενεργειών (Barth & Landsman, 2010).

Η Ελεγκτική ως γνωστικό αντικείμενο ευθύνεται επιπλέον για την προώθηση και την ανάπτυξη των αρχών και μεθόδων λειτουργίας της επιπλέον των υπόλοιπων γνωστικών αντικειμένων και τις άμεσα ή έμμεσα συνδεόμενες συναφείς επιστήμες και τεχνολογίες (Beneish, et al. 2015).

Ο ενδιαφερόμενος και ο έχων έννομο ενδιαφέρον για τη διεξαγωγή ενός ελέγχου είναι ο πλέον άμεσα οικονομικά συνδεόμενος με την ελεγχόμενη οικονομική μονάδα και όλοι οι ελεγκτές ορίζονται κάθε χρόνο από τη γενική συνέλευση των μετόχων και με τη μεσολάβηση και ευθύνη του διοικητικό συμβούλιου, όμως, εμμέσως σαφώς, για τον έλεγχο ενδιαφέρονται και τόσοι άλλοι όπως είναι οι κοινωνικοί εταίροι της μονάδας (Bengtsson, 2011).

Βάση του νόμου, Π.Δ. 226/92, άρθρο 16, παράγραφος 1^α, εντολέας του ελέγχου και αποδέκτης της έκθεσης του ελεγκτή είναι για τις ιδιωτικές επιχειρήσεις(η γενική συνέλευση μετόχων ή εταίρων) και για τις μονάδες όπου η πλειοψηφία των κεφαλαίων ανήκει στο Δημόσιο ο αποδέκτης της έκθεσης είναι αναλόγως της μορφής της μονάδας, (το ανώτατο εποπτικό όργανο αυτής, η γενική συνέλευση, η διοικούσα επιτροπή, το εποπτικό συμβούλιο, το διοικητικό συμβούλιο και ο υπεύθυνος υπουργός) Όσο αναφορά τα νομικά πρόσωπα δημοσίου δικαίου, η έκθεση

υποβάλλεται στην αρμόδια εποπτεύουσα αρχή π.χ. για τους δήμους στον υπουργό Εσωτερικών, αν και βασικός αποδέκτης της έκθεσης θα πρέπει να είναι το δημοτικό συμβούλιο. (Botzem, 2014).

Μέχρι σήμερα η σημαντικότερη πηγή των χρηματοοικονομικών πληροφοριών παραμένει στο λογιστικό σύστημα της επιχείρησης, στηριζόμενο στις αρχές της λογιστικής ιστορικού κόστους.

Οι λογιστικές καταστάσεις είναι το προϊόν ολόκληρου του λογιστικού συστήματος κάθε οικονομικής μονάδας, βάση της λογιστικής διαδικασίας ιστορικού κόστους και αποτυπώνεται με την τεχνική της διπλογραφικής μεθόδου. Η Λογιστική κυρίως γνωστή για την τεχνική της πλευρά, παρά από την θεωρητική. Επομένως πολλοί αναφέρονται σ αυτήν ως τη διαδικασία της συλλογής, ομαδοποίησης ,καταχώρησης ,παρουσίασης των αποτελεσμάτων, των μετρήσεων και άλλων πληροφοριών για την οικονομική κατάσταση και δραστηριότητα της οικονομικής μονάδας, με κύριο στόχο να συμβάλλουν στην εκτίμηση των αποτελεσμάτων, της παρούσας οικονομικής θέσεως, καθώς και στην πρόβλεψη των προοπτικών που έχει (Botzem & Dobusch, 2012).

2.4 Οι διακρίσεις στην Ελεγκτική

Το αντικείμενο του ελέγχου στα πλαίσια της χρηματοοικονομικής ελεγκτικής είναι κατά κύριο λόγο ο τρόπος και η αξιοπιστία που γίνεται η μέτρηση και παρουσίαση των χρηματοοικονομικών μεγεθών που δίνουν την εικόνα και το αποτέλεσμα μιας οικονομικής μονάδας. (Collins, et al. 2012).

Έτσι, η Ελεγκτική ως λειτουργία σε ευρύτερη διοίκηση (ιδιωτική και κρατική) μιας κοινωνίας διακρίνεται βασικά σε (Bratton & Cunningham, 2009) :

- Εξωτερική, ανεξάρτητη Ελεγκτική.
- Εσωτερική Ελεγκτική.
- Κρατική Ελεγκτική.

Εξωτερική, ανεξάρτητη Ελεγκτική : Ο εξωτερικός έλεγχος γίνεται κατά βάση του λογιστικού συστήματα μιας οικονομικής μονάδας, αλλά και σε άλλα επιμέρους οργανωτικά συστημάτων, τα οποία μπορούν εμμέσως να επηρεάζουν την αποτελεσματικότητα και λειτουργία όλου του λογιστικού συστήματος, με τη χρήση κάποιων ελεγκτικών standards αλλά και άλλων ελεγκτικών μεθόδων τα οποία είναι απαραίτητα στον ελεγκτή. Ο εσωτερικός έλεγχος στοχεύει στην αξιολόγηση διαδικασιών λειτουργίας και παρέχει πληροφορίες στη διοίκηση. Ανεξάρτητος έλεγχος θεωρείται είναι αυτός ο οποίος διενεργείται από ειδικά εκπαιδευμένους γι' αυτό το σκοπό λογιστές/ελεγκτές, που αναφέρονται ως ορκωτοί λογιστές ή ορκωτοί ελεγκτές. (Brüggemann, et al. 2013).

Οι μέτοχοι, παρόλο που έχουν την δυνατότητα ως οι ιδιοκτήτες της ελεγχόμενης οικονομικής μονάδας, δεν μπορούν, στις σημερινές επιχειρήσεις-κολοσσούς να έχουν πλεονεκτικότερη θέση πλην όμως ορισμένων οικογενειακών επιχειρήσεων, ίσως όμως και εκεί, η (αναγκαστική για την διοίκηση μεγάλων τέτοιων μονάδων) μεσολάβηση των managers μπορεί να αποδυναμώσει το σχετικό πλεονέκτημα της «κλειστής» ιδιοκτησίας. Η χρήση εσωτερικών ελεγκτών στο τελευταίο παράδειγμα

δεν καθιστά, τελικά, πλεονασματική τη χρησιμοποίηση εξωτερικών ελεγκτών, για διάφορους λόγους (Byard, et al. 2011).

Εσωτερική Ελεγκτική : Ο εσωτερικός έλεγχος ορίζεται από το ευρωπαϊκό Ινστιτούτο Εσωτερικών Ελεγκτών ως «μία ανεξάρτητη διαδικασία που λειτουργεί μέσα στον οργανισμό για να εξετάσει και να εκτιμήσει τις διαδικασίες που παρέχουν σημαντική υπηρεσία σε αυτόν». Επίσης επιφορτίζεται με το συνεχή έλεγχο και αξιολόγηση όλων των συστημάτων οργάνωσης και διαδικασιών λειτουργίας της οικονομικής μονάδας (Chua & Taylor, 2008).

Οι βασικοί στόχοι της Εσωτερικής Ελεγκτικής είναι (Chalmers, et al. 2011) :

- Η καθορισμένη λειτουργία των επιμέρους τμημάτων της μονάδας (μαζί και του Λογιστηρίου) σύμφωνα με πολιτική που υπάρχει και ισχύει, όπως αυτή περιγράφεται στα συγκεκριμένα εγχειρίδια λειτουργίας και διαδικασιών.
- Η διαπίστωση της αποτελεσματικότητας της εσωτερικής οργάνωσης (Internal Control).
- Η διαπίστωση τυχόν απάτης , λαθών και αδυναμιών στην οργάνωση και λειτουργία (προσωπικού και συστημάτων), που έχουν ως αποτέλεσμα την κακή διαχείριση των περιουσιακών στοιχείων της οικονομικής μονάδας ή και τη φθορά και απώλεια αυτών.
- Η εκτίμηση και αναφορά της αποτελεσματικότητας και της επάρκειας των συστημάτων οργάνωσης και διοίκησης της οικονομικής μονάδας.
- Η αναφορά και εκτίμηση των διαπιστώσεων του ελέγχου στη διοίκηση της μονάδας και η εισήγηση μέτρων ώστε να διορθωθεί η να βελτιωθεί, όπου αυτό κρίθει απαραίτητο.

Απο τους στόχους του εσωτερικού ελέγχου διαφαίνεται ήδη ποια είναι η διαφορά μεταξύ αυτού και του εξωτερικού ελέγχου. Θα μπορούσαμε να προσθέσουμε και κάποιες ακόμα διαφορές, οι οποίες αποτελούν τα μειονεκτήματα του εσωτερικού

ελέγχου όχι με την έννοια ότι ο εξωτερικός έλεγχος θα πρέπει να προτιμάται αντί του εσωτερικού, διότι οι δύο είναι απαραίτητοι και χρήσιμοι για τ ξεχωριστό λόγο που παίζει ο καθένας μακροοικονομικά και μικροοικονομικά αντιστοίχως, αλλά με σκοπό να τονιστεί ότι ο εξωτερικός έλεγχος δεν μπορεί να αντικαταστήσει τον εσωτερικό, δεδομένου ότι ο εσωτερικός έλεγχος είναι πιο απλός(και όχι τόσο σύνθετος όσο ο εξωτερικός) και υπάρχει σε κάποια μορφή, σε όλες τις επιχειρήσεις (Chen & Sami, 2008).

Κρατική Ελεγκτική : Εδώ περιλαμβάνονται όλοι οι έλεγχοι που γίνονται από διάφορους κρατικούς φορείς. Οι σημαντικότεροι έλεγχοι που γίνονται είναι οι φορολογικοί έλεγχοι που διενεργούνται από την εφορία και συγκεντρώνεται στην αναζήτηση και επαλήθευση των φορολογητέων εισοδημάτων των υποκείμενων σε φορολογία νομικών και φυσικών προσώπων. Επίσης έλεγχος από την Τράπεζα της Ελλάδος για νομισματικά και πιστωτικά θέματα και ο έλεγχος από το Ελεγκτικό Συνέδριο του κράτους που αφορά τους λογαριασμούς στο δημόσιο τομέα,

Σε σχέση με την εξωτερική ελεγκτική η κρατική ελεγκτική είναι αρκετά πιο συνοπτική και η διαφορά τους είναι ότι δεν ακολουθεί γενικώς παραδεκτά λογιστικά και ελεγκτικά πρότυπα. Και οι δύο αυτές μορφές ελέγχου, όμως, ψάχνουν τα τεκμήρια και προβαίνουν στις ανάλογες επαληθεύσεις μεγεθών και διαδικασιών. Θεωρείτο βέβαιο πάντως ότι η κρατική ελεγκτική θα μπορούσε σαφώς να αντικατασταθεί, από την εξωτερική ελεγκτική και εσωτερικά από σύστημα που αναφέραμε των εσωτερικών ελεγκτών

Διαχειριστικός έλεγχος : Για κάποιο χρονικό διάστημα, στην βιβλιογραφία περί εξωτερικού ελέγχου, δηλαδή εκείνου που γίνεται από το ΣΟΛ, είχε παρεμφρήσει η ιδέα και του διαχειριστικού ελέγχου, κατά τον οποίο ερευνούν κατά πόσον η οικονομική μονάδα διαχειρίζεται την περιουσία της σύμφωνα με τις διατάξεις που ισχύουν από τους νόμους, το καταστατικό και τους κανόνες διαχείρισης της εταιρείας. (Chen, et al. 2013)

2.5 Κλάδοι της Ελεγκτικής

Σε συνάρτηση με τον διαχωρισμό της ελεγκτικής σε διακριτές κατηγορίες, κλάδοι της ελεγκτικής μπορεί να θεωρηθούν οι παρακάτω:

- Η ιδιωτική ελεγκτική, η οποία ασχολείται με τον έλεγχο των ιδιωτικών επιχειρήσεων.
- Η δημόσια ελεγκτική, η οποία ασχολείται καθαρά με τον έλεγχο των νομικών Προσώπων δημοσίου δικαίου και των οργανισμών του δημοσίου (Παπάς, 1999).

Κεφάλαιο 3ο : Η εξόρυξη δεδομένων (Data Mining)

3.1 Ο ορισμός της εξόρυξης δεδομένων

Η εξόρυξη δεδομένων είναι η διαδικασία εξεύρεσης μιας ή περισσοτέρων πληροφοριών που μας ενδιαφέρουν με την χρήση των αλγορίθμων από μεγάλες βάσεις δεδομένων που περιλαμβάνουν διάφορες μεθόδους ώστε να μπορέσουν να αναζητήσουν και να βρουν πολύ γρήγορα και λεπτομερειακά βάσεις δεδομένων για την εξεύρεση τυχόν κρυμμένων προτύπων(patterns). Ένα διεπιστημονικό υποπεδίο της επιστήμης των υπολογιστών είναι μια βασική διαδικασία, όπου εφαρμόζονται έξυπνες μέθοδοι για την εξαγωγή δεδομένων όπου ο γενικός στόχος είναι η εξαγωγή των πληροφοριών από δεδομένα από τα οποία όταν εξαχθούν να μπορούν να μετατραπούν σε κατανοητή δομή προς τον άνθρωπο για περαιτέρω χρήση και να τον βοηθήσουν να πάρει τις κατάλληλες αποφάσεις.(Alles, et al. 2006). Έτσι λοιπόν απλώς μπορούμε να πούμε ότι η εξόρυξη δεδομένων είναι μία διαδικασία κατά την οποία μπορούμε να εξάγουμε καλά κρυμμένες πληροφορίες από μεγάλες βάσεις δεδομένων.

Εκτός από το πρώτο στάδιο ανάλυσης, περιλαμβάνει επίσης διάφορες πτυχές διαχείρισης βάσεων δεδομένων και πληροφοριών, προεπεξεργασία των δεδομένων, εκτιμήσεις μοντέλων και συμπερασμάτων, μετρήσεις ενδιαφερόντων, εκτιμήσεις πολυπλοκότητας, οπτικοποίηση και ηλεκτρονική ενημέρωση. Η εξόρυξη δεδομένων είναι το βήμα της ανάλυσης της διαδικασίας «ανεύρεση γνώσεων σε βάσεις δεδομένων»(Knowledge discovery in databases, KDD).

Γενικότερα ο όρος εξόρυξη δεδομένων είναι μια λάθος ονομασία, γιατί ο στόχος είναι η εξαγωγή μοτίβων και γνώσεων από μεγάλα βάσεις δεδομένων και όχι η εξαγωγή (εξόρυξη) των ίδιων των δεδομένων. Είναι ωστόσο και μια λέξη-κλειδί η οποία εφαρμόζεται συχνά σε όλες τις μορφές επεξεργασίας δεδομένων ή πληροφοριών μεγάλης κλίμακας (συλλογή, εξαγωγή, αποθήκευση, ανάλυση και στατιστικά

στοιχεία) καθώς και οποιαδήποτε εφαρμογή συστήματος υποστήριξης αποφάσεων υπολογιστών, μηχανική μάθηση και επιχειρηματική ευφυΐα (Du & Roohani, 2007).

Βιβλίο Εξόρυξης των δεδομένων : Τα πρακτικά εργαλεία και οι τεχνικές εκμάθησης μηχανών με την Java (που καλύπτει κυρίως το υλικό εκμάθησης μηχανών) αρχικά ονομαζόταν πρακτική εκμάθηση μηχανών και μετέπειτα απλώς προστέθηκε ο όρος εξόρυξη δεδομένων για λόγους μάρκετινγκ. Συχνά οι πιο γενικές (μεγαλύτερης έκτασης) αναλύσεις και αναλύσεις δεδομένων ή όταν γίνεται αναφορά σε πραγματικές μεθόδους, η τεχνητή νοημοσύνη και η μηχανική μάθηση, είναι πιο κατάλληλες λόγω του όγκου των πληροφοριών.

Η πραγματική εργασία που χρησιμοποιείτε η εξόρυξη των δεδομένων είναι η ημιαυτόματη ή αυτόματη ανάλυση μεγάλων ποσοτήτων δεδομένων για την εξαγωγή προηγούμενων άγνωστων ενδιαφερόντων προτύπων, όπως ομάδες αρχείων δεδομένων (ανάλυση συμπλέγματος), ασυνήθιστων αρχείων (ανίχνευση ανωμαλιών) και εξαρτήσεων (εξόρυξη κανόνα σύνδεσης, διαδοχική εξόρυξη προτύπων) (Kirkos, et al. 2007).

Αυτό συνήθως έχει ως επακόλουθο την χρήση διαφόρων τεχνικών βάσης δεδομένων όπως των χωρικών δεικτών. Αυτά τα πρότυπα μπορούν να θεωρηθούν ως ένα είδος περίληψης των δεδομένων εισόδου και μπορούν να χρησιμοποιηθούν για κάποια περαιτέρω ανάλυση ή να χρησιμοποιηθούν στα υποπεδία της μηχανικής μάθησης και της προγνωστικής ανάλυσης.

Για παράδειγμα, το βήμα εξόρυξης δεδομένων μπορεί να εντοπίσει αρκετές ομάδες στα δεδομένα, τα οποία μπορούν αργότερα να χρησιμοποιηθούν ώστε να επιτευχθούν ακριβέστερα αποτελέσματα πρόβλεψης από ένα σύστημα υποστήριξης αποφάσεων. Ούτε η ερμηνεία, η συλλογή δεδομένων, η προετοιμασία δεδομένων και η αναφορά των αποτελεσμάτων μπορούν να αποτελέσουν μέρος του βήματος της εξόρυξης δεδομένων, αλλά ανήκουν στο σύνολο της διαδικασίας KDD ως πρόσθετα βήματα.

Οι σχετικοί όροι εξόρυξης δεδομένων, αλειίας δεδομένων και snooping δεδομένων αναφέρονται στην χρήση των διαφόρων μεθόδων εξόρυξης δεδομένων για δειγματοληπτικό έλεγχο του τμήματος με μεγάλη συλλογή δεδομένων που είναι (ή μπορεί να είναι) πολύ μικρό για αξιόπιστα στατιστικά συμπεράσματα σχετικά με την

εγκυρότητα οποιονδήποτε σχεδίων που ανακαλύφθηκαν. Αυτές οι μέθοδοι επίσης μπορούν να χρησιμοποιηθούν για τη δημιουργία νέων υποθέσεων για να δοκιμαστούν ενάντια στους μεγαλύτερους πληθυσμούς δεδομένων (Kotsiantis, et al. 2007).

3.2 Η ετυμολογία

Στη δεκαετία του 1960, οι στατιστικολόγοι και οι οικονομολόγοι χρησιμοποίησαν όρους όπως η αλιεία δεδομένων ή η βυθοκόρηση δεδομένων για να αναφερθούν σε αυτά που θεωρούσαν την κακή πρακτική της ανάλυσης δεδομένων χωρίς μια υπόθεση a priori. Ο όρος "εξόρυξη δεδομένων" χρησιμοποιήθηκε κατά παρόμοιο τρόπο από τον οικονομολόγο Michael Lovell σε άρθρο που δημοσιεύτηκε στην αναθεώρηση των Οικονομικών Μελετών του 1983 (Martens, et al. 2008).

Ο όρος εξόρυξη δεδομένων εμφανίστηκε γύρω στο 1990 στην κοινότητα βάσεων δεδομένων, γενικά με θετικούς συνειρμούς. Για ένα μικρό χρονικό διάστημα στη δεκαετία του '80 χρησιμοποιήθηκε μια φράση "εξόρυξη βάσεων δεδομένων", αλλά δεδομένου ότι η εμπορική επωνυμία της HNC, μιας εταιρείας με έδρα το Σαν Ντιέγκο, έφτασε στο σταθμό βάσης δεδομένων, οι ερευνητές στράφηκαν προς την εξόρυξη δεδομένων. Άλλοι όροι που χρησιμοποιήθηκαν περιλαμβάνουν την αρχαιολογία δεδομένων, τη συλλογή πληροφοριών, την ανακάλυψη πληροφοριών, την εξαγωγή γνώσης κλπ.

Ο Gregory Piatetsky-Shapiro εισήγαγε τον όρο "ανακάλυψη γνώσεων σε βάσεις δεδομένων" για το πρώτο εργαστήριο με το ίδιο θέμα (KDD-1989) στην κοινότητα AI και μηχανικής μάθησης. Ωστόσο, ο όρος "εξόρυξη δεδομένων" έγινε πιο δημοφιλής στις κοινότητες επιχειρήσεων και του Τύπου. Σήμερα, οι όροι "εξόρυξη δεδομένων" και "ανακάλυψη γνώσης" χρησιμοποιούνται εναλλακτικά (Vasarhelyi & Kuenkaikaew, 2010).

Στην ακαδημαϊκή κοινότητα, τα σημαντικότερα φόρουμ έρευνας άρχισαν το 1995, όταν ξεκίνησε το πρώτο διεθνές συνέδριο για την εξόρυξη δεδομένων και την ανακάλυψη γνώσης (KDD-95) στο Μόντρεαλ με χορηγία AAAI. Προήδρευαν από κοινού ο Usama Fayyad και ο Ramasamy Uthurusamy. Ένα χρόνο αργότερα, το 1996, ο Usama Fayyad ξεκίνησε το περιοδικό από τον Kluwer που ονομάζεται Data Mining and Knowledge Discovery ως ιδρυτικός διευθυντής. Αργότερα ξεκίνησε το SIGKDD Newsletter SIGKDD Explorations.

Το συνέδριο KDD International έγινε το πρωτεύον συνέδριο υψηλότερης ποιότητας στην εξόρυξη δεδομένων με ποσοστό αποδοχής των υποβαλλομένων ερευνητικών εγγράφων κάτω από 18%. Το περιοδικό Data Mining and Knowledge Discovery είναι το κύριο ερευνητικό περιοδικό του πεδίου (Vasarhelyi, et al. 2010).

3.3 Η έρευνα και τα πρότυπα

Ο κορυφαίος επαγγελματικός οργανισμός σε αυτόν τον τομέα είναι η Ομάδα Ειδικών Συμφερόντων (SIG) της Ένωσης Υπολογιστικών Μηχανημάτων (ACM) για την Ανακάλυψη της Γνώσης και την Εξόρυξη Δεδομένων (SIGKDD). Από το 1989, το ACM SIG φιλοξένησε το ετήσιο διεθνές συνέδριο και δημοσίευσε τις εργασίες του και από το 1999 έχει δημοσιεύσει ένα εξαμηνιαίο ακαδημαϊκό περιοδικό με τίτλο "SIGKDD Explorations" (Du & Roohani, 2007).

Οι διασκέψεις επιστήμης υπολογιστών σχετικά με την εξόρυξη δεδομένων περιλαμβάνουν (Alles, et al. 2006) :

- Συνδιάσκεψη CIKM - Διάσκεψη ACM για τη διαχείριση πληροφοριών και γνώσεων.
- Ευρωπαϊκή Διάσκεψη για την εκμάθηση μηχανών, αρχές και πρακτική της αποκάλυψης γνώσεων σε βάσεις δεδομένων.
- Συνδιάσκεψη KDD - Συνέδριο ACM SIGKDD για την ανακάλυψη γνώσης και την εξόρυξη δεδομένων.

Τα θέματα εξόρυξης δεδομένων είναι επίσης παρόντα σε πολλές διασκέψεις διαχείρισης δεδομένων / βάσεων δεδομένων, όπως η διάσκεψη ICDE, το συνέδριο SIGMOD και το διεθνές συνέδριο για τις πολύ μεγάλες βάσεις δεδομένων.

Έχουν καταβληθεί αρκετές προσπάθειες για τον ορισμό προτύπων για τη διαδικασία εξόρυξης δεδομένων, όπως για παράδειγμα η Πρότυπη Διαδικασία της Ευρωπαϊκής Στατιστικής για την Εξόρυξη Δεδομένων (CRISP-DM 1.0) και το Πρότυπο Εξόρυξης Δεδομένων Java (JDM 1.0) του 2004. Η ανάπτυξη στους διαδόχους αυτών των διαδικασιών (CRISP-DM 2.0 και JDM 2.0) ήταν ενεργή το 2006, αλλά έχει σταματήσει από τότε. Το JDM 2.0 αποσύρθηκε χωρίς να καταφέρει να φτάσει σε κάποιο τελικό σχέδιο (Kirkos, et al. 2007).

Για την ανταλλαγή των εξαγόμενων μοντέλων, ειδικά για χρήση σε αναλυτικές προβλέψεις, το βασικό πρότυπο είναι η γλώσσα πρόβλεψης προτύπων μοντέλων (PMML), η οποία είναι μια γλώσσα που βασίζεται σε XML που αναπτύχθηκε από την Ομάδα Εξόρυξης Δεδομένων (DMG) και υποστηρίζεται ως μορφή ανταλλαγής από πολλές εφαρμογές εξόρυξης δεδομένων.

Όπως υποδηλώνει και το όνομα, καλύπτει μόνο τα μοντέλα πρόβλεψης, ένα συγκεκριμένο έργο εξόρυξης δεδομένων που έχει μεγάλη σημασία για επιχειρηματικές εφαρμογές. Ωστόσο, έχουν προταθεί επεκτάσεις για την κάλυψη (για παράδειγμα) της ομαδοποίησης υποσυνόλων ανεξάρτητα από την DMG (Kirkos, et al. 2010).

3.4 Η προστασία των προσωπικών δεδομένων και η ηθική

Αν και η ορολογία "εξόρυξη δεδομένων" δεν μπορούμε να πούμε πως έχει ηθικές συνέπειες, συχνά συνδέεται με την εξόρυξη πληροφοριών σε σχέση με την συμπεριφορά των ανθρώπων (ηθική και άλλα).

Τους τρόπους με τους οποίους μπορεί να χρησιμοποιηθεί η εξόρυξη δεδομένων μπορεί σε ορισμένες περιπτώσεις και σε περιβάλλοντα να εγείρουν ερωτήματα σχετικά με την ιδιωτικότητα, τη νομιμότητα και την ηθική. Ειδικότερα οι κυβερνητικές ή εμπορικές βάσεις δεδομένων για να μπορέσουν να εξορύξουν δεδομένα για σημαντικούς σκοπούς όπως είναι η εθνική ασφάλεια, η επιβολή του νόμου, όπως στο Πρόγραμμα Συνολικής Πληροφόρησης, έχουν προκαλέσει ανησυχίες σχετικά με την προστασία της ιδιωτικής ζωής (Kotsiantis, et al. 2007).

Η εξόρυξη δεδομένων απαιτεί την προετοιμασία δεδομένων, η οποία μπορεί να αποκαλύψει πληροφορίες ή πρότυπα που ενδέχεται να θέσουν σε κίνδυνο τις υποχρεώσεις εμπιστευτικότητας και προστασίας της ιδιωτικής ζωής. Ένας συνηθισμένος τρόπος για να συμβεί αυτό είναι η συνάθροιση δεδομένων. Η συνάθροιση δεδομένων συνεπάγεται με τη συγκέντρωση δεδομένων (πιθανώς από διάφορες πηγές) με τρόπο που διευκολύνει την ανάλυση (αλλά μπορεί επίσης να καταστήσει δυνατό τον προσδιορισμό ιδιωτικών δεδομένων ατομικού επιπέδου ή άλλων αντιλήψεων) (Vasarhelyi & Kuenkaikaw, 2010).

Αυτό δεν είναι εξόρυξη δεδομένων καθ'εαυτό, αλλά αποτέλεσμα της προετοιμασίας δεδομένων πριν, και για τους σκοπούς, της ανάλυσης. Η απειλή για την ιδιωτική ζωή ενός ατόμου τίθεται υπό αμφισβήτηση όταν τα δεδομένα, μετά τη σύνταξη, προκαλούν στον άνθρωπο ή σε οποιονδήποτε έχει πρόσβαση στο νεοσυσταθέν σύνολο δεδομένων να είναι σε θέση να εντοπίσει συγκεκριμένα άτομα, ειδικά όταν τα δεδομένα ήταν αρχικά ανώνυμα (Martens, et al. 2008).

Συνιστάται να γνωρίζει ένα άτομο τα ακόλουθα πριν συλλέξει τα δεδομένα (Vasarhelyi, et al. 2010) :

- Ο σκοπός της συλλογής δεδομένων και οποιονδήποτε (γνωστών) σχεδίων εξόρυξης δεδομένων.
- Πώς θα χρησιμοποιηθούν τα δεδομένα.
- Τα άτομα που ασχολούνται θα μπορούν να εξορύσσουν τα δεδομένα και να χρησιμοποιούν και τα παράγωγά τους.
- Το καθεστώς ασφάλειας που περιβάλλει την πρόσβαση στα δεδομένα.
- Τα δεδομένα που συλλέγονται μπορούν να ενημερωθούν.

Τα δεδομένα μπορούν επίσης να τροποποιηθούν έτσι ώστε να καταστούν ανώνυμα, έτσι ώστε τα άτομα να μην μπορούν εύκολα να αναγνωριστούν. Ωστόσο, ακόμη και τα σύνολα δεδομένων "de-identified" / "anonymized" ενδέχεται να περιέχουν επαρκείς πληροφορίες για να επιτρέπουν την ταυτοποίηση ατόμων, όπως συνέβη όταν οι δημοσιογράφοι κατάφεραν να βρουν πολλά άτομα βασισμένα σε μια σειρά ιστοριών αναζήτησης που απελευθερώθηκαν άθελα από την AOL (Alles, et al. 2006).

Η αποκάλυψη προσωπικών αναγνωρίσιμων πληροφοριών που οδηγούν στον πάροχο παραβιάζει τις ορθές πρακτικές πληροφόρησης. Αυτή η αδιακρισία μπορεί να προκαλέσει οικονομική, συναισθηματική ή σωματική βλάβη στο υποδεικνυόμενο άτομο. Σε μια περίπτωση παραβίασης της ιδιωτικής ζωής, οι προστάτες της Walgreens υπέβαλαν αγωγή κατά της εταιρείας το 2011 για την πώληση πληροφοριών σε εταιρείες εξόρυξης δεδομένων, οι οποίες με τη σειρά τους παρείχαν τα δεδομένα σε φαρμακευτικές εταιρείες.

Η κατάσταση στην Ευρώπη

Η Ευρώπη έχει μάλλον από τους πιο ισχυρούς νόμους περί απορρήτου και καταβάλλονται προσπάθειες για την επιπλέον ενίσχυση των δικαιωμάτων των καταναλωτών. Οι αρχές ασφαλούς δεδομένων εκθέτουν επί του παρόντος αποτελεσματικά τους ευρωπαίους χρήστες στην εκμετάλλευση της ιδιωτικής ζωής από εταιρείες των ΗΠΑ. Ως επακόλουθο της δημοσιοποίησης της παγκόσμιας παρακολούθησης του Edward Snowden, υπήρξε η αυξημένη συζήτηση για την ανάκληση αυτής της συμφωνίας, καθώς τα δεδομένα θα είναι πλήρως εκτεθειμένα στον Οργανισμό Εθνικής Ασφάλειας και οι προσπάθειες επίτευξης συμφωνίας θα έχουν αποτύχει (Du & Roohani, 2007).

Η κατάσταση στις Ηνωμένες Πολιτείες

Στις Ηνωμένες Πολιτείες, οι ανησυχίες για την προστασία της ιδιωτικής ζωής αντιμετωπίστηκαν από το Κογκρέσο των ΗΠΑ μέσω της θέσπισης κανονιστικών ελέγχων όπως ο νόμος περί φορητότητας και λογοδοσίας για τις ασφάλειες υγείας (HIPAA). Η HIPAA απαιτεί από τα άτομα να δώσουν την «ενημερωμένη συγκατάθεσή» τους σχετικά με τις πληροφορίες που παρέχουν και τις προβλεπόμενες παρούσες και μελλοντικές χρήσεις.

Σύμφωνα με άρθρο στο Biotech Business Week, «στην πράξη, η HIPAA μπορεί να μην προσφέρει μεγαλύτερη προστασία απ ότι οι μακρόχρονοι κανονισμοί στην ερευνητική σκηνή», λέει ο AAHC. Το πιο σημαντικό είναι ότι ο στόχος του κανόνα για προστασία μέσω ενημερωμένης συγκατάθεσης που προσεγγίζουν ένα επίπεδο ακατανόητης γνησιότητας των μέσων ατόμων. Αυτό υπογραμμίζει την αναγκαιότητα ανωνυμίας των δεδομένων σε πρακτικές συνάθροισης δεδομένων και εξορυκτικές πρακτικές (Kirkos, et al. 2007).

Η νομοθεσία περί απορρήτου πληροφοριών των ΗΠΑ, όπως η HIPAA και το Family Educational Rights and Privacy Act (FERPA), ισχύουν μόνο για τους συγκεκριμένους τομείς που αντιμετωπίζει κάθε τέτοιος νόμος. Η χρήση της εξόρυξης δεδομένων από την πλειονότητα των επιχειρήσεων στις ΗΠΑ δεν ελέγχεται από καμία νομοθεσία.

3.4.1 Αλγόριθμοι

Αν και στην εξόρυξη δεδομένων δεν χρειάζεται να εισάγουμε μεγάλες ποσότητες δεδομένων, είναι καλύτερο και πιο αποτελεσματικό να χρησιμοποιούμε αυτές τις τεχνικές σε μεγαλύτερους όγκους δεδομένων. Στην εξόρυξη δεδομένων είναι καλύτερα να χρησιμοποιείται για δομές δεδομένων και συσχετίσεις μεταξύ τους, που θα ήταν δύσκολο να γίνουν αντιληπτές από τον άνθρωπο. Άρα, είναι προτιμότερο να χρησιμοποιήσουμε τέτοιους αλγόριθμων στα στάδια της κατανόησης και προετοιμασίας των δεδομένων μας. Τέτοια παραδείγματα αλγόριθμων είναι οι εξής: οι αλγόριθμοι γενίκευσης, σύνοψης, εντοπισμού και διόρθωσης διαφόρων ελλειπών δεδομένων που μπορεί να υπάρχουν. Η διάκριση μεταξύ τεχνικών εξόρυξης δεδομένων δεν είναι ευκρινής διότι αρκετές από τις τεχνικές ΕΔ έχουν αντληθεί από τον τομέα που αφορά την στατιστική. Τυχόν εργαλεία και μέθοδοι στατιστικής μπορούν να χρησιμοποιηθούν ως τεχνικές ΕΔ. Οι μέθοδοι αυτοί αποτελούν την επίλυση προβλημάτων, τον τρόπο επιλογής των δεδομένων, την παρουσίαση του αποτελέσματος, καθώς και την αξιολόγηση του. Στην στατιστική υπήρχε μία σταθερή θεωρητική βάση όπου τα αποτελέσματα από τη χρήση των μεθόδων ερμηνεύονται αρκετά δύσκολα καθώς είναι απαραίτητες οι κινήσεις που πρέπει να γίνουν και προς ποια κατεύθυνση θα αναλυθούν τα όποια δεδομένα εξαχθούν. Στην εξόρυξη δεδομένων από την άλλη μπορούν να συνυπάρξουν τα δεδομένα που θα αναλύσει ο Η/Υ και η γνώση του εμπειρογνώμονα.

Κάποιες άλλες τεχνικές και δείκτες που υπάρχουν στην στατιστική και μπορούν να χρησιμοποιηθούν και στην ανάλυση ΕΔ είναι η παλινδρόμηση, ο διάμεσος, ο μέσος, η συσχέτιση, η διακύμανση, η τυπική απόκλιση, κ.α.

3.5 Ο Νόμος περί πνευματικών δικαιωμάτων

Η κατάσταση που επικρατεί στην Ευρώπη

Λόγω έλλειψης ευελιξίας στον ευρωπαϊκό νόμο περί πνευματικών δικαιωμάτων και βάσεων δεδομένων, η εξόρυξη δεδομένων που προστατεύονται από τα πνευματικά δικαιώματα, όπως είναι η εξόρυξη ιστού χωρίς την άδεια του κατόχου των πνευματικών δικαιωμάτων, είναι παράνομη. Όταν τα δεδομένα μιας βάσης είναι καθαρά δεδομένα στην Ευρώπη, είναι πιθανόν να μην υπάρχουν δικαιώματα πνευματικής ιδιοκτησίας, αλλά ενδέχεται να υπάρχουν δικαιώματα βάσης δεδομένων, έτσι ώστε η εξόρυξη δεδομένων να υπόκειται σε κανονισμούς βάσει των σχετικών οδηγιών για τις βάσεις δεδομένων (Kirkos, et al. 2010).

Με τη σύσταση της αναθεώρησης του Hargreaves αυτό οδήγησε την κυβέρνηση του Ηνωμένου Βασιλείου να τροποποιήσει το νόμο περί πνευματικών δικαιωμάτων το 2014 ώστε να επιτρέψει την εξόρυξη περιεχομένου ως περιορισμό και εξαίρεση. Μόνο η δεύτερη χώρα στον κόσμο μπόρεσε να το πράξει μετά την Ιαπωνία, η οποία εισήγαγε μια εξαίρεση το 2009 για την εξόρυξη δεδομένων. Ωστόσο, λόγω του περιορισμού της Οδηγίας περί Πνευματικής Ιδιοκτησίας, η εξαίρεση του Ηνωμένου Βασιλείου επιτρέπει μόνο την εξόρυξη περιεχομένου για τους μη εμπορικούς σκοπούς (Vasarhelyi & Kuenkaikaew, 2010).

Η νομοθεσία που έχει το Ηνωμένο Βασίλειο περί πνευματικής ιδιοκτησίας δεν επιτρέπει την παραβίαση αυτής της διάταξης από συμβατικούς όρους και προϋποθέσεις. Η Ευρωπαϊκή Επιτροπή διευκόλυνε τη συζήτηση των ενδιαφερόμενων φορέων σχετικά με την εξόρυξη κειμένου και δεδομένων το 2013, με τίτλο "Άδειες για την Ευρώπη".

Η επικέντρωση στην επίλυση αυτού του νομικού ζητήματος, όπως οι άδειες και όχι οι περιορισμοί και οι εξαιρέσεις, οδήγησαν τους εκπροσώπους των βιβλιοθηκών, πανεπιστημίων, ερευνητών, ομάδων της κοινωνίας των πολιτών να εγκαταλείψουν το διάλογο με τους ενδιαφερόμενους τον Μάιο του 2013.

Η κατάσταση που επικρατεί στις Ηνωμένες Πολιτείες

Αντίθετα με την Ευρώπη, ο ελαστικός χαρακτήρας που έχει ο αμερικανικός νόμος περί πνευματικής ιδιοκτησίας, και ειδικά η δίκαιη χρήση, σημαίνει ότι η εξόρυξη περιεχομένου στην Αμερική, καθώς και άλλες χώρες δίκαιης χρήσης όπως είναι το Ισραήλ, η Νότια Κορέα η Ταϊβάν είναι νόμιμες. Καθώς η εξόρυξη περιεχομένου είναι μετασχηματιστική, δηλαδή δεν αντικαθιστά το πρωτότυπο έργο, θεωρείται νόμιμη υπό δίκαιη χρήση (Vasarhelyi, et al. 2010).

Δίνοντας ένα παράδειγμα, στο πλαίσιο του διακανονισμού του Βιβλίου της Google, ο προεδρεύων δικαστής έκρινε ότι το πρόγραμμα ψηφιοποίησης της Google για βιβλία που προστατεύονται από πνευματικά δικαιώματα ήταν νόμιμο εν μέρει λόγω των μετασχηματιστικών χρήσεων που παρουσίασε το έργο ψηφιοποίησης, όπου είναι η εξόρυξη κειμένου και δεδομένων (Kirkos, et al. 2010).

3.6 Τα λογισμικά προγράμματα για την εξόρυξη δεδομένων

Οι παρακάτω εφαρμογές είναι διαθέσιμες υπό άδειες ελεύθερου / ανοιχτού κώδικα. Η δημόσια πρόσβαση στον πηγαίο κώδικα της εφαρμογής είναι επίσης διαθέσιμη (Alles, et al. 2006).

- Carrot2
- Chemicalize.org
- ELKI
- GATE
- KNIME
- Massive Online Analysis (MOA)
- MEPX
- ML-Flex
- MLPACK library
- NLTK (Natural Language Toolkit)
- OpenNN
- Orange
- R
- Torch
- UIMA
- Weka

Το ιδιόκτητο λογισμικό και οι εφαρμογές εξόρυξης δεδομένων

Οι παρακάτω εφαρμογές είναι διαθέσιμες υπό άδειες ιδιοκτησίας (Du & Roohani, 2007).

- Angoss
- Clarabridge:
- KXEN Modeler

- LIONsolver
- Megaputer Intelligence
- Microsoft Analysis Services
- NetOwl
- OpenText Big Data Analytics
- Oracle Data Mining
- PSeven
- Qlucore Omics Explorer
- RapidMiner
- SAS Enterprise Miner
- SPSS Modeler
- STATISTICA Data Miner
- Tanagra
- Vertica

Έρευνες αγοράς

Αρκετοί οργανισμοί και ερευνητές έχουν πραγματοποιήσει ανασκοπήσεις των εργαλείων εξόρυξης δεδομένων και των ερευνών για τα δεδομένα. Αυτά προσδιορίζουν μερικά από τα πλεονεκτήματα και τις αδυναμίες που υπάρχουν στα πακέτα λογισμικών. Επίσης μπορούν να παρέχουν μια γενική εικόνα των συμπεριφορών, των προτιμήσεων και των απόψεων της εξόρυξης δεδομένων (Kirkos, et al. 2010).

Ορισμένες από αυτές τις αναφορές περιλαμβάνουν :

- Δείκτης Hurwitz Victory : Είναι μια έκθεση για προχωρημένους αναλυτές, ως εργαλείο αξιολόγησης της αγοράς, υπογραμμίζει τόσο τις ποικίλες χρήσεις για την προηγμένη τεχνολογία ανάλυσης όσο και τους προμηθευτές που κάνουν τις εφαρμογές αυτές δυνατές.
- Έρευνες Data Miner Rexer Analytics (2007-2015).

- 2011 Wiley Διεπιστημονικές ανασκοπήσεις : Εξόρυξη δεδομένων και ανακάλυψη γνώσης.
- Σύμφωνα με την έκθεση Forrester Research 2010, η προβλεπτική ανάλυση και οι λύσεις εξόρυξης δεδομένων.
- Έκθεση Gartner 2008 "Magic Quadrant".
- Τρίτη σειρά άρθρων του Robert A. Nisbet για το 2006 "Εργαλεία Εξόρυξης Δεδομένων : Ποιο είναι το καλύτερο για το CRM".
- Η αναθεώρηση των Πακέτων Λογισμικού Εξόρυξης Δεδομένων για την Αμερικάνικη Στατιστική Υπηρεσία.
- Goebel & Gruenwald 1999 "Έρευνα της εξόρυξης δεδομένων ενός εργαλείου λογισμικού ανεύρεσης γνώσης" στις Εξερευνήσεις SIGKDD.

Κεφάλαιο 4ο : Το λογισμικό για την εξόρυξη δεδομένων Weka

4.1 Εισαγωγή

Το Weka είναι ένα πρόγραμμα για υπολογιστές γραμμένο σε γλώσσα Java που σχεδιάστηκε και αναπτύχθηκε στο Πανεπιστήμιο Waikato της Νέας Ζηλανδίας. Πρόκειται για ένα δωρεάν λογισμικό που μπορεί να έχει πρόσβαση σε αυτό ο καθένας.

Το Weka περιλαμβάνει αρκετές εργασίες της εξόρυξης δεδομένων όπως είναι η συσταδοποίηση, επεξεργασία δεδομένων, παλινδρόμηση, διάφορα γραφήματα και οπτική παρουσίαση. Η χρήση του βασίζεται αρκετά στην βέλτιστη χρήση του υπολογιστική ισχύ και έχει ως σκοπό την εκπαίδευση της μηχανικής εκμάθησης και να μας παρέχει συσχετίσεις και πρότυπα.

Η αρχική μη έκδοση Java του Weka ήταν Tcl / Tk front-end σε αλγόριθμους μοντελοποίησης που εφαρμόστηκαν σε κάποιες άλλες γλώσσες προγραμματισμού, καθώς και σε βοηθητικά προγράμματα προεπεξεργασίας δεδομένων στο C και σύστημα βασισμένο σε Makefile για πειράματα μηχανικής μάθησης (Hall, 1999).

Αρχικά το WEKA σχεδιάστηκε αρχικά ως το εργαλείο για την ανάλυση των δεδομένων στους αγροτικούς τομείς, αλλά η πρόσφατη έκδοσή του (Weka 3), βασισμένη σε Java της οποίας η ανάπτυξη είχε ξεκινήσει το 1997, πλέον χρησιμοποιείται σε πολλές διαφορετικές εφαρμογές, κυρίως για έρευνα και εκπαιδευτικούς σκοπούς.

Τα πλεονεκτήματα που έχει το Weka είναι τα εξής:

- Δωρεάν διαθεσιμότητα βάσει της Γενικής Δημόσιας Άδειας GNU.

- Φορητότητα, δεδομένου ότι υλοποιείται πλήρως στη γλώσσα προγραμματισμού Java άρα είναι συμβατό σχεδόν σε όλες τις σύγχρονες πλατφόρμες υπολογιστών.
- Μια ολοκληρωμένη συλλογή τεχνικών προεπεξεργασίας και μοντελοποίησης των δεδομένων.
- Ευκολία στη χρήση λόγω των γραφικών διεπαφών χρήστη.

Το Weka υποστηρίζει πολλές τυπικές εργασίες εξόρυξης δεδομένων, ειδικότερα, προεπεξεργασίας δεδομένων, ταξινόμησης, ομαδοποίησης, παλινδρόμησης, οπτικοποίησης και επιλογής χαρακτηριστικών (Larose, 2006).

Οι τεχνικές που χρησιμοποιεί Weka βασίζονται κυρίως στην υπόθεση ότι τα δεδομένα είναι διαθέσιμα ως ένα κοινό αρχείο ή σχέση, όπου κάθε σημείο δεδομένων περιγράφεται από έναν σταθερό αριθμό χαρακτηριστικών (κανονικά, ονομαστικά ή αριθμητικά χαρακτηριστικά, αλλά υποστηρίζονται επίσης μερικοί άλλοι τύποι χαρακτηριστικών).

Η πρόσβαση που παρέχει το WEKA είναι σε βάσεις δεδομένων SQL χρησιμοποιώντας Java Database Connectivity όπου μπορεί να επιστρέψει το αποτέλεσμα από ένα ερώτημα βάσης δεδομένων αφού το επεξεργαστεί. Το Weka έχει την δυνατότητα να παρέχει πρόσβαση σε βαθιά μάθηση με το DeepLearning4j.

Το WEKA δεν έχει την δυνατότητα να κάνει εξόρυξη πολλαπλών σχέσεων αλλά σε αυτήν την περίπτωση μπορεί ο χρήστης να χρησιμοποιήσει ξεχωριστό λογισμικό για να μετατρέψει μια συλλογή συνδεδεμένων πινάκων βάσης δεδομένων σε έναν πίνακα που θα είναι ενιαίος και θα μπορεί να τον διαχειριστεί. Επίσης οι αλγόριθμοι του WEKA δεν καλύπτουν έναν σημαντικό τομέα που είναι η μοντελοποίηση των ακολουθιών.

Ιστορία

- Το 1993, το Πανεπιστήμιο Waikato στη Νέα Ζηλανδία ξεκίνησε την ανάπτυξη της αρχικής έκδοσης του Weka, το οποίο έγινε μίγμα των Tcl / Tk, C και Makefiles.
- Το 1997, αποφάσισαν να γίνει αναδιαμόρφωση από την αρχή του Weka σε γλώσσα Java, που συμπεριλάμβανε και τις εφαρμογές των αλγορίθμων μοντελοποίησης.
- Το 2005, το Weka πήρε το βραβείο SIGKDD Data Mining and Discovery Service Discovery Service.
- Το 2006, η εταιρεία Pentaho Corporation πήρε την αποκλειστική άδεια χρήσης του Weka για επιχειρηματική ευφυΐα. Αποτελεί την συνιστώσα εξόρυξης δεδομένων και προγνωστικής ανάλυσης της πλατφόρμας επιχειρησιακών πληροφοριών Pentaho.

4.2 Οι διεπαφές του χρήστη

Το WEKA έχει ως κύριο περιβάλλον εργασίας το Explorer, αλλά ουσιαστικά η ίδια λειτουργικότητα είναι προσβάσιμη μέσω της διασύνδεσης ροής γνώσης που βασίζεται σε στοιχεία από την γραμμή των εντολών. Επίσης είναι και ο πειραματιστής, ο οποίος επιτρέπει τη σύγκριση της πρόβλεψης των αλγορίθμων μηχανικής μάθησης του Weka σε μια συλλογή συνόλου δεδομένων (Puthran & Shah, 2012).

Η διεπαφή του Explorer διαθέτει διάφορα πάνελ που παρέχουν πρόσβαση στα κυριότερα στοιχεία του πάγκου εργασίας τα οποία είναι τα εξής (Rochester, 2013) :

- Ο πίνακας προεργασίας διαθέτει δυνατότητες για την εισαγωγή δεδομένων από μια βάση δεδομένων , ένα αρχείο διαχωρισμένων τιμών με κομμάτια (CSV)κ.λπ. και για την προεπεξεργασία αυτών των δεδομένων χρησιμοποιώντας έναν αλγόριθμο φιλτραρίσματος. Αυτά τα φίλτρα μπορούν να χρησιμοποιηθούν για να μετασχηματίζουν τα δεδομένα (π.χ. μετατρέποντας τα αριθμητικά χαρακτηριστικά σε διακεκριμένα) και να επιτρέπουν τη διαγραφή στιγμιότυπων και χαρακτηριστικών σύμφωνα με συγκεκριμένα κριτήρια.
- Ο πίνακας ταξινόμησης επιτρέπει την εφαρμογή αλγορίθμων ταξινόμησης και παλινδρόμησης (αδιακρίτως ονομαζόμενων ταξινομητών του Weka) στο σύνολο δεδομένων που προκύπτει, για την εκτίμηση της ακρίβειας του προγνωστικού μοντέλου που προκύπτει και για την απεικόνιση λανθασμένων προβλέψεων, καμπυλών χαρακτηριστικών κ.λπ.
- Ο πίνακας "Συνεργάτες" προσφέρει πρόσβαση σε άτομα με κανόνες σύνδεσης που θέλουν να εντοπίσουν όλες τις σημαντικές αλληλεπιδράσεις μεταξύ χαρακτηριστικών στα δεδομένα.

- Ο πίνακας συμπλέγματος προσφέρει πρόσβαση στις τεχνικές συσσωμάτωσης στο Weka, π.χ., ο απλός αλγόριθμος k-mean. Υπάρχει επίσης μια χρήση του αλγόριθμου μεγιστοποίησης προσδοκίας για την εκμάθηση ενός συνόλου κανονικών κατανομών.
- Ο πίνακας "Επιλογή χαρακτηριστικών" παρέχει αλγόριθμους για τον εντοπισμό των πιο προγνωστικών χαρακτηριστικών σε ένα σύνολο δεδομένων.
- Ο πίνακας "Οπτικοποίηση" μας παρουσιάζει μια γραφική παράσταση ενός πίνακα, όπου μπορούν να επιλεγούν και να διευρυνθούν μεμονωμένα διαγράμματα σκέδασης και να αναλυθούν περαιτέρω χρησιμοποιώντας διάφορους χειριστές επιλογής.

4.3 Τα εργαλεία της παλινδρόμησης και τα πακέτα επέκτασης του λογισμικού

Το Weka διαθέτει έναν μεγάλο αριθμό εργαλείων παλινδρόμησης και ταξινόμησης. Τα εγγενή πακέτα είναι αυτά που περιλαμβάνονται στο εκτελέσιμο λογισμικό Weka, ενώ άλλες μη μητρικές μπορούν να μεταφορτωθούν και να χρησιμοποιηθούν στο περιβάλλον R.Weka. Μεταξύ των εγγενών πακέτων, το διασημότερο εργαλείο είναι το πακέτο δέντρων μοντέλου M5p (Witten, et al. 2011).

Είναι διαθέσιμος ο πλήρης κατάλογος των. Μερικά από τα εργαλεία παλινδρόμησης είναι τα εξής(Bouman & Van Dongen, 2009) :

- M5Rules
- DecisionStump
- M5P
- RandomForest
- RepTree
- ZeroR
- DecisionRules
- Γραμμική παλινδρόμηση
- SMOreg
- SimpleLinearRegression
- MultiLayerPerceptron
- GaussianProcesses

Πακέτα επέκτασης : Στην έκδοση 3.7.2, προστέθηκε ένας διαχειριστής πακέτων για να διευκολυνθεί η εγκατάσταση των πακέτων επέκτασης. Κάποιες από τις λειτουργίες που χρησιμοποιήθηκαν για να συμπεριληφθούν στο Weka πριν από αυτή την έκδοση έχουν μεταφερθεί από τότε σε αυτά τα πακέτα επέκτασης, αλλά αυτή η αλλαγή διευκολύνει επίσης άλλους να συνεισφέρουν επεκτάσεις στο Weka και να διατηρούν το λογισμικό, καθώς αυτή η αρθρωτή αρχιτεκτονική επιτρέπει να γίνονται ανεξάρτητες ενημερώσεις τον πυρήνα του Weka και τις επιμέρους επεκτάσεις (Maimon&Rokach,2010).

4.4 Τα σχετικά εργαλεία του λογισμικού

- Το αυτοματοποιημένο σύστημα εκμάθησης μηχανών για το Weka είναι το Auto-WEKA
- Το περιβάλλον για DeveLoping : Είναι οι εφαρμογές KDD που υποστηρίζονται από τις δομές ευρετηρίου (ELKI) είναι ένα παρόμοιο έργο με το Weka με επίκεντρο την ανάλυση συμπλέγματος, δηλαδή τις μεθόδους χωρίς επίβλεψη.
- Το KNIME είναι ένα λογισμικό εκμάθησης μηχανών και εξόρυξης δεδομένων το οποίο υλοποιείται στην Java.
- Το Massive Online Analysis (MOA) είναι ένα έργο ανοιχτού κώδικα για την εξόρυξη ροών δεδομένων μεγάλης κλίμακας, που αναπτύχθηκε επίσης στο Πανεπιστήμιο του Waikato της Νέας Ζηλανδίας.
- Το Neural Designer είναι ένα λογισμικό εξόρυξης δεδομένων που βασίζεται σε τεχνικές βαθιάς εκμάθησης γραμμένες σε C ++.
- Το Orange είναι ένα παρόμοιο έργο ανοιχτού κώδικα για την εξόρυξη δεδομένων, την εκμάθηση μηχανών και την οπτικοποίηση που γράφεται σε Python και C ++.
- Το RapidMiner είναι ένα εμπορικό μηχάνημα μάθησης που υλοποιείται στην Java που ενσωματώνει το Weka (Puthran & Shah, 2012).

Επίλογος

Αν και η ενσωμάτωση μεταξύ τεχνικών εξόρυξης δεδομένων και διαδικασιών ελέγχου είναι ένας σχετικά νέος τομέας, η εξόρυξη δεδομένων θεωρείται χρήσιμη και συμβάλλει στη μείωση της πίεσης κόστους σε πολλές επιχειρηματικές εφαρμογές που σχετίζονται με τον έλεγχο. Οι τεχνικές εξόρυξης δεδομένων ενδέχεται να είναι σε θέση να συμβάλουν σε κάτι τέτοιο, ακόμη και όταν τα δεδομένα δεν είναι τόσο ολοκληρωμένα και όλες οι αναλύσεις βασίστηκαν σε υποθέσεις. Ωστόσο, δεν αποδεικνύει ότι η εξόρυξη δεδομένων είναι κατάλληλη για κάθε εργασία ελέγχου. Με αποτέλεσμα, αυτό να απαιτεί σημαντική προσπάθεια για να μάθουν πώς να χρησιμοποιούν τις τεχνικές εξόρυξης δεδομένων και να κατανοούν τις συνέπειες των αποτελεσμάτων. Ωστόσο, εάν οι ελεγκτικές εταιρείες διαθέτουν τεράστιες ποσότητες δεδομένων για έλεγχο και οι ελεγκτές είναι εξοικειωμένοι με τη φύση των συναλλαγών και τα αναμενόμενα πρότυπα σφαλμάτων, τότε η εξόρυξη δεδομένων παρέχει ένα αποτελεσματικό μέσο για επιφανειακά ενδιαφέροντα θέματα. Οι τεχνικές εξόρυξης δεδομένων διαθέτουν δυνατότητες ταξινόμησης και πρόβλεψης που μπορούν να διευκολύνουν τη λήψη αποφάσεων σε οικονομικά προβλήματα. Οι μέθοδοι εξόρυξης δεδομένων που χρησιμοποιούνται στη συλλεγμένη βιβλιογραφία περιλαμβάνουν τα Νευρωνικά Δίκτυα, τους Γενετικούς Αλγορίθμους, τα Δέντρα Αποφάσεων, τη Θεωρία Rough Set, τη Θεμελιώδη Λήψη Υποθέσεων και τον Μαθηματικό Προγραμματισμό. Οι περισσότερες από τις έρευνες φαίνεται να προτιμούν το μοντέλο Νευρωνικού Δικτύου. Παρόλο που ένα σημαντικό ποσό ερευνητικής προσπάθειας έχει αντιμετωπίσει την εφαρμογή των τεχνικών εξόρυξης δεδομένων στη χρηματοδότηση, υπάρχουν πολλές εύφορες περιοχές για περαιτέρω έρευνα.

Βιβλιογραφία

Agoglia, T. Doupkin, S. and Tsakumis, T. (2011). Principles-based versus rules-based accounting standards: The influence of standard precision and audit committee strength on financial reporting decisions. *The Accounting Review*. Vol. 86 No. 3, pp. 747-767.

Alles, M. Brennan, G. Kogan, A. & Vasarhelyi, A. (2006). Continuous monitoring of business process controls: A pilot implementation of a continuous auditing system at Siemens. *International Journal of Accounting Information Systems*. 7(2): 137-161.

Alon, A. & Dwyer, D. (2014). Early adoption of IFRS as a strategic response to transnational and local influences. *The International Journal of Accounting*. 49(3), 348-370.

Armstrong, S. Barth, E. Jagolinzer, D. and Riedl, J. (2010). Market reaction to the adoption of IFRS in Europe. *The Accounting Review*. Vol. 85 No. 1, pp. 31-61.

Atwood, J. Drake, S. Myers, N. and Myers, A. (2011). Do earnings reported under IFRS tell us more about future earnings and cash flows?. *Journal of Accounting and Public Policy*. Vol.30 No. 2, pp. 103-121.

Ball, A. & Craig, R. (2010). Using neo-institutionalism to advance social and environmental accounting. *Critical Perspectives on Accounting*. 21(4), 283-293.

Ball, R. Li, X. & Shivakumar, L. (2015). Contractibility and transparency of financial statement information prepared under IFRS: Evidence from debt contracts around IFRS adoption. *Journal of Accounting Research*. 53(5), 915-963.

Ballas, A. Skoutela, D. & Tzovas, C. (2010). The relevance of IFRS to an emerging market: evidence from Greece. *Managerial Finance*. 36(11), 931-948.

Balsmeier, B. & Vanhaverbeke, S. (2016). International financial reporting standards and private firms' access to bank loans. *European Accounting Review*.

Barth, E. & Landsman, R. (2010). How did financial reporting contribute to the Financial Crisis? *European Accounting Review*. 19(3): 399-423.

Beneish, M. Miller, B. and Lombardi-Yohn, T. (2015). Macroeconomic evidence on the impact of mandatory IFRS adoption on equity and debt markets. *Journal of Accounting and Public Policy*. Vol 34 No. 1, pp. 1-27.

Bengtsson, E. (2011). Repoliticalization of accounting standard setting- The IASB, the EU and the global financial crisis. *Critical Perspectives on Accounting*. 22(6), 567-580.

Botzem, S. (2014). Transnational standard setting in accounting: Organizing expertise-based self-regulation in times of crises. *Accounting, Auditing & Accountability Journal*. 27(6), 933-955.

Botzem, S., & Dobusch, L. (2012). Standardization cycles: A process perspective on the formation and diffusion of transnational standards. *Organization Studies*. 33(5-6), 737-762.

Bouman, R. & Van Dongen, J. (2009). *Pentaho Solutions: Business Intelligence and Data Warehousing with Pentaho and MySQL*. Indianapolis, IN: Wiley Publishing Inc.

Bratton, W. & Cunningham, A. (2009). Treatment differences and political realities in the GAAP-IFRS debate. *Virginia Law Review*. 95(4), 989-1023.

Brüggemann, U. Hitz, J. & Sellhorn, T. (2013). Intended and unintended consequences of mandatory IFRS adoption: A review of extant evidence and suggestions for future research. *European Accounting Review*. 22(1), 1-37.

Byard, D. Li, Y. and Yu, Y. (2011). The effect of mandatory IFRS adoption on financial analysts' information environment. *Journal of Accounting Research*. (49). 69-96.

Chalmers, K. Clinch, G. and Godfrey, J. (2011). Changes in value relevance of accounting information from IFRS adoption: Evidence from Australia. *Australian Journal of Management*. Vol 36 No. 2, pp. 151-173.

Chen, H. and Sami, H. (2008). Trading volume reaction to the earnings reconciliation from IAS to U.S. GAAP. *Contemporary Accounting Research*. Vol. 25 No. 1. 15-53.

Chen, F. Hope, K. Li, Q. & Wang, X. (2011). Financial reporting quality and investment efficiency of private firms in emerging markets. *The Accounting Review*. 86(4), 1255-1288.

Chen, X. Cheng, Q. and Lo K. (2013) Accounting restatements and external financing choices. *Contemporary Accounting Research*. Vol 30(2), pp. 750-779.

Chen, H. Deng, S. Gupta, P. and Sami, H. (2015). The SEC's elimination of 20-F reconciliation and information asymmetry. *Journal of Contemporary Accounting & Economics*. 11, 75-87.

Chen, H. and Khurana. K. (2016). The impact of eliminating the Form 20-F reconciliation on shareholder wealth: Evidence from U.S. cross-listed firms. *The Accounting Review*. Vol. 90, No.199-228.

Chua, F. & Taylor, L. (2008). The rise and rise of IFRS: An examination of IFRS diffusion. *Journal of Accounting and Public Policy*. 27(6), 462-473.

Chua, Y. Cheong, C. and Gould, G. (2012). The impact of mandatory IFRS adoption on accounting quality: Evidence from Australia. *Journal of International Accounting Research*. Vol 11 No. 1, pp. 119-146.

Collins, L. Pasewark, R. and Riley, E. (2012). Rules-based and principles-based accounting standards. *Accounting Horizons*. Vol 26 No. 4, pp. 681-705.

Daske, H. Hail, L. Leuz, C. and Verdi, R. (2008). Mandatory IFRS reporting around the world: Early evidence on the economic consequences. *Journal of Accounting Research*. Vol 46 No. 5, pp. 1085-1142.

Daske, H. Hail, L. Leuz, C. & Verdi, R. (2013). Adopting a label: Heterogeneity in the economic consequences around IAS/IFRS adoptions. *Journal of Accounting Research*. 51(3), 495-547.

Dimitropoulos, E. Asteriou, D. Kousenidis, D. & Leventis, S. (2013). The impact of IFRS on accounting quality: Evidence from Greece. *Advances in Accounting*. 29 (1), 108-123.

Du, H. & Roohani, S. (2007). Meeting Challenges and Expectations of Continuous Auditing in the Context of Independent Audits of Financial Statements. *International Journal of Auditing*. Vol. 11, No. 2, pp. 133-146, July 2007.

Durocher, S. & Gendron, Y. (2011). IFRS: on the docility of sophisticated users in preserving the ideal of comparability. *European Accounting Review*. 20(2), 233-262.

El-Gazzar, M. and Fornaro, M. (2012). Voluntary versus mandatory corporate disclosures on management responsibilities for financial reporting: An empirical investigation. *Global Review of Accounting and Finance*. Vol 3 No. 2, pp. 74-93.

Eng, L. Sun, L. and Vichitsarawong, T. (2014). Are international financial reporting standards-based and U. S. GAAP-based accounting amounts comparable? Evidence from U.S. ARDs, *Journal of Accounting Auditing and Finance*. Vol 29 (2), pp 1630187.

Fifield, S. Finningham, G. Fox, A. Power, D. & Veneziani, M. (2011). A cross-country analysis of IFRS reconciliation statements. *Journal of Applied Accounting Research*. 12(1), 26-42.

Floropoulos, I. (2006). IFRS. First time users: Some empirical evidence from Greek companies. *Spoudai: Journal of Economics and Business*. 56(3), 39-70.

Hall, A. (1999). *Correlation-Based Feature Selection for Machine Learning* (Ph.D.). University of Waikato.

Kaluza, B. (2013). *Instant WEKA How-to*. Birmingham. UK: Packt Publishing Ltd.

Kirkos, E. Spathis, C. & Manolopoulos, Y. (2010). Audit-firm group appointment : an artificial intelligence approach. *Intelligent Systems in Accounting, Finance & Management*. 17(1): 1-17.

Kotsiantis, S. Koumanakos, E. Tzelepis, D. & Tampakas, V. (2007). Forecasting Fraudulent Financial Statements using Data Mining. *International Journal of Computational Intelligence*. 3(2).

Larose, T. (2006). *Data Mining Methods and Models*. Hoboken. NJ: John Wiley & Sons Inc.

Maimon, O. & Rokach, L. (2010). *Data Mining and Knowledge Discovery Handbook*. New York, NY: Springer + Business Media.

Martens, D. Bruynseels, L. Baesens, B. Willekens, M. & Vanthienen, J. (2008). Predicting going concern opinion with data mining. *Decision Support Systems*. 45(4): 765-777.

Puthran, S. & Shah, K. (2012). *Intrusion Detection System using Datamining Techniques*. Saarbrücken: LAP LAMBERT Academic Publishing.

Rochester, E. (2013). *Clojure Data Analysis Cookbook*. Birmingham, UK: Packt Publishing Ltd.

Vasarhelyi, A. & Kuenkaikaew, S. (2010). Continuous auditing and continuous control monitoring: case studies from leading organizations: Rutgers Business School. *Rutgers Accounting Research Center*.

Vasarhelyi, A. Teeter, A. & Krahel, J. (2010). Audit Education and the Real-Time Economy. *Issues in Accounting Education*. 25(3).

Witten, I. H. Frank, E. & Hall, A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques*. Burlington, MA: Morgan Kaufmann Publishers.