



# **ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ**

## **ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ**

### **ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ**

**Πρόγραμμα Μεταπτυχιακών Σπουδών**

**Επιστήμη και Τεχνολογία της Πληροφορικής και των Υπολογιστών**

**Ειδίκευση Λογισμικού και Πληροφοριακών Συστημάτων**

#### **ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ**

**Εφαρμογή τεχνικών εξόρυξης διεργασιών στον τομέα της Εκπαίδευσης  
εξάγοντας αρχεία καταγραφής δεδομένων από βάση δεδομένων**

**Ελένη Ανδριάνη**

**A.M. MCSE19061**

**Επιβλέποντες:**

**Γεώργιος Μιαούλης, Ομότιμος Καθηγητής**

**Γεωργία Θεοδωροπούλου, Υποψήφια Διδάκτωρ**

Αθήνα, Ιανουάριος 2023



## **ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ**

**Εφαρμογή τεχνικών εξόρυξης διεργασιών στον τομέα της Εκπαίδευσης  
εξάγοντας αρχεία καταγραφής δεδομένων από βάση δεδομένων**

**Ελένη Ανδριάνη  
Α.Μ. MCSE19061**

**Εισηγητές: Μιαούλης Γεώργιος, Ομότιμος Καθηγητής  
Θεοδωροπούλου Γεωργία, Υποψήφια Διδάκτωρ**

**Εξεταστική Επιτροπή:**

**Μιαούλης Γεώργιος, Ομότιμος Καθηγητής  
Μπαρδής Γεώργιος, Επίκουρος Καθηγητής  
Μπουσδέκης Αλέξανδρος, Πανεπιστημιακός Υπότροφος ΠΑ.Δ.Α.**

**Ημερομηνία εξέτασης      27 / 01 / 2023**



## ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΜΕΤΑΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ

Η κάτωθι υπογεγραμμένη Ελένη Ανδριάνη του Νεοκλή, με αριθμό μητρώου MCSE19061 φοιτήτρια του Προγράμματος Μεταπτυχιακών Σπουδών Επιστήμη και Τεχνολογία της Πληροφορικής και των Υπολογιστών, ειδίκευση Λογισμικού και Πληροφοριακών Συστημάτων του Τμήματος Μηχανικών Πληροφορικής και Υπολογιστών της Σχολής Μηχανικών του Πανεπιστημίου Δυτικής Αττικής, δηλώνω ότι:

«Είμαι συγγραφέας αυτής της μεταπτυχιακής εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της, είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, οι όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε ακριβώς είτε παραφρασμένες, αναφέρονται στο σύνολό τους, με πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο. Επίσης, βεβαιώνω ότι αυτή η εργασία έχει συγγραφεί από μένα αποκλειστικά και αποτελεί προϊόν πνευματικής ιδιοκτησίας τόσο δικής μου, όσο και του Ιδρύματος.

Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του πτυχίου μου».

Επιθυμώ την απαγόρευση πρόσβασης στο πλήρες κείμενο της εργασίας μου μέχρι 30/06/2023 και έπειτα από αίτηση μου στη Βιβλιοθήκη και έγκριση του επιβλέποντα καθηγητή.

Η Δηλούσα



Ελένη Ανδριάνη



## ΕΥΧΑΡΙΣΤΙΕΣ

Η παρούσα διπλωματική εργασία ολοκληρώθηκε μετά από επίμονες προσπάθειες, σε ένα ενδιαφέρον γνωστικό αντικείμενο, όπως αυτό της εξόρυξης διαδικασιών και σε μια δύσκολη συγκυρία για όλους. Την προσπάθειά μου αυτή υποστήριξαν οι επιβλέποντες κ.κ. Γεώργιος Μιαούλης και Γεωργία Θεοδωροπούλου, τους οποίους θα ήθελα να ευχαριστήσω για την εμπιστοσύνη και τη στήριξη.

Επιπλέον θα ήθελα να ευχαριστήσω τους αγαπημένους μου φίλους και συγγενείς οι οποίοι συνέβαλαν στην ολοκλήρωση της παρούσας εργασίας με τη συνεχή στήριξή τους όλο αυτό το διάστημα, καθώς και τους συναδέλφους των Διευθύνσεων Εκπαιδευτικής Τεχνολογίας & Καινοτομίας και Ανάπτυξης Πληροφοριακών Συστημάτων του ΥΠΑΙΘ που ανταποκρίθηκαν προκειμένου να αποκτηθούν τα απαραίτητα δεδομένα από τα πληροφοριακά συστήματα.





## ΠΕΡΙΛΗΨΗ

Στην σύγχρονη εποχή πολύ μεγάλες ποσότητες δεδομένων και διαδικασιών καταγράφονται και αποθηκεύονται και συνεπώς προκύπτει η ανάγκη ανάλυσής τους προκειμένου να απαντηθούν ερωτήματα όπως τι έχει συμβεί, γιατί έχει συμβεί και πώς αυτό θα μπορούσε να βελτιωθεί μελλοντικά.

Η ανάγκη αυτή έχει οδηγήσει στην ανάπτυξη του επιστημονικού τομέα της εξόρυξης διαδικασιών (process mining) που μας παρέχει τη δυνατότητα να απαντήσουμε σε ερωτήματα που σχετίζονται με την απόδοση των διαδικασιών αναλύοντας τα δεδομένα και τα συμβάντα που καταγράφονται κατά την εκτέλεσή τους, δημιουργώντας μοντέλα που αποτυπώνουν τις διαδικασίες αυτές και πραγματοποιώντας έλεγχο συμμόρφωσης της καταγεγραμμένης με την αναμενόμενη διαδικασία.

Ένας επιμέρους τομέας της εξόρυξης διαδικασιών είναι και η εξόρυξη εκπαιδευτικών διαδικασιών (educational process mining), όπου χρησιμοποιούνται αρχεία καταγραφής συμβάντων από εκπαιδευτικά πληροφοριακά συστήματα προκειμένου να διευκολυνθεί η καλύτερη κατανόηση της εκάστοτε διαδικασίας από τους ειδικούς που ασχολούνται με τη βελτίωση των διαδικασιών στον τομέα της εκπαίδευσης.

Στην παρούσα εργασία παρουσιάζονται αρχικά οι τεχνικές εξόρυξης διαδικασιών καθώς και οι εφαρμογές αυτών σε εκπαιδευτικές διαδικασίες. Στη συνέχεια πραγματοποιείται η εφαρμογή της εξόρυξης διαδικασιών στη διαδικασία των Πανελλαδικών Εξετάσεων όπως αυτή οργανώνεται και υλοποιείται από το Υπουργείο Παιδείας και Θρησκευμάτων, η οποία πραγματοποιείται με τα εργαλεία ProM και Disco (Academic Licence). Τέλος, καταγράφονται οι παρατηρήσεις και τα συμπεράσματα που προέκυψαν κατά την αξιολόγηση των αποτελεσμάτων της εξόρυξης, καθώς και μελλοντικές προοπτικές που αφορούν στη συγκεκριμένη διαδικασία.

**Λέξεις Κλειδιά:** Εξόρυξη διαδικασιών, Εξόρυξη εκπαιδευτικών διαδικασιών, Πανελλαδικές εξετάσεις, ProM, Ανακάλυψη διαδικασιών, Έλεγχος συμμόρφωσης

## **ABSTRACT**

In modern era, large amounts of data and processes are recorded and saved. As a result of this, the need to analyze them is arisen in order for questions such as what has happened, why it has happened and how this process can be improved in the future, to be answered.

The aforementioned need had led to the development of the scientific field of process mining which enables us to respond to questions about the effectiveness of the processes by analyzing the data and events recorded during their execution, creating models that reflect these processes and obtaining conformance checking between the recorded and expected process.

A subfield of process mining is the educational process mining in which event logs from educational information systems are used so as to boost a deeper understanding of each process by experts who aim at the improvement of processes in education.

In this dissertation, the techniques of process mining are presented as well as their application in several educational processes. Moreover, with the help of the tools ProM and Disco (Academic License), process mining is applied to the National Exams exactly as they are organized in compliance with Greek Ministry of Education and Religion. Finally, all the observations and arisen conclusions from the evaluation of the results of mining are included in the hope that potential uses of this specific process are revealed in the future.

**Keywords:** Process mining, Educational process mining, National exams, ProM, Process discovery, Conformance checking

## ΠΕΡΙΕΧΟΜΕΝΑ

1.	ΕΙΣΑΓΩΓΗ .....	17
2.	ΔΙΑΧΕΙΡΙΣΗ ΚΑΙ ΕΞΟΡΥΞΗ ΔΙΑΔΙΚΑΣΙΩΝ (PROCESS MINING) .....	19
2.1	Εισαγωγή .....	19
2.2	Διαδικασίες.....	21
2.3	Αρχεία Καταγραφής Συμβάντων (Event Logs) .....	22
2.4	Εξόρυξη Διαδικασιών (Process Mining) .....	23
2.5	Τεχνικές εξόρυξης διαδικασιών .....	24
2.6	Αξιολόγηση (Evaluation): Διάγνωση διαφορών μεταξύ παρατηρούμενης συμπεριφοράς και μοντελοποιημένης συμπεριφοράς.....	25
2.7	Ερωτήματα και προκλήσεις που προκύπτουν κατά την εξόρυξη διαδικασιών .....	27
2.8	Εργαλεία για την Εξόρυξη Διαδικασιών .....	30
3.	ΕΞΟΡΥΞΗ ΕΚΠΑΙΔΕΥΤΙΚΩΝ ΔΙΑΔΙΚΑΣΙΩΝ (EDUCATIONAL PROCESS MINING).....	33
3.1	Η εξόρυξη διαδικασιών στον τομέα της Εκπαίδευσης .....	33
3.2	Εννοιολογικό πλαίσιο της εξόρυξης εκπαιδευτικών διαδικασιών .....	35
3.3	Τομείς εφαρμογών της εξόρυξης εκπαιδευτικών διαδικασιών .....	37
3.4	Μελλοντικές προκλήσεις στην εξόρυξη εκπαιδευτικών διαδικασιών – Εκπαιδευτική Διοίκηση.....	40
3.5	Εξόρυξη Εκπαιδευτικών Δεδομένων για την Ανάλυση της Συμπεριφοράς των Εκπαιδευόμενων.....	43
4.	ΕΞΟΡΥΞΗ ΔΙΑΔΙΚΑΣΙΩΝ ΠΑΝΕΛΛΑΔΙΚΩΝ ΕΞΕΤΑΣΕΩΝ.....	45
4.1	Διαδικασία διεξαγωγής Πανελλαδικών Εξετάσεων για την πρόσβαση στην Τριτοβάθμια Εκπαίδευση .....	45
4.2	Καταγραφή διαδικασίας των Πανελλαδικών εξετάσεων .....	45
4.3	Βασικοί στόχοι της ανάλυσης και εξόρυξης της διαδικασίας Πανελλαδικών Εξετάσεων.....	49
4.4	Γενική μεθοδολογία εξόρυξης διαδικασιών .....	49
4.5	Λήψη δεδομένων των Πανελλαδικών Εξετάσεων από το Υπουργείο Παιδείας και προετοιμασία τους για αξιοποίηση με τεχνικές εξόρυξης.....	50
4.5.1	Προγραμματισμός και εξαγωγή δεδομένων.....	52
4.5.2	Συλλογή δεδομένων .....	53
4.6	Προετοιμασία δεδομένων για εισαγωγή στα εργαλεία εξόρυξης .....	55

5.	ΥΛΟΠΟΙΗΣΗ ΚΑΙ ΕΦΑΡΜΟΓΗ .....	57
5.1	Εισαγωγή των αρχείων στο ProM και επιλογές οπτικοποίησής τους .....	57
5.1.1	Οπτικοποίηση (Visualization) πρώτου αρχείου καταγραφής συμβάντων .....	58
5.1.2	Οπτικοποίηση (Visualization) δεύτερου αρχείου καταγραφής συμβάντων ...	64
5.2	Εξερεύνηση του πρώτου αρχείου καταγραφής συμβάντων με το εργαλείο γραφήματος Dotted Chart .....	66
5.3	Ανακάλυψη Διαδικασιών (Process Discovery) μέσω της εκτέλεσης αλγορίθμων εξόρυξης διαδικασιών .....	70
5.3.1	Alpha Miner .....	71
5.3.2	Heuristics Miner.....	74
5.3.3	Inductive Miner .....	75
5.3.4	Inductive Visual Miner .....	77
5.3.5	Fuzzy Miner.....	80
5.3.6	Disco .....	81
5.3.7	Παρατηρήσεις από τη διαδικασία του Process Discovery.....	88
5.4	Έλεγχος Συμμόρφωσης (Conformance Checking).....	89
5.4.1	Έλεγχος συμμόρφωσης του μοντέλου που προέκυψε από τον Inductive Miner με φίλτρο θορύβου 20% .....	90
5.4.2	Έλεγχος συμμόρφωσης του μοντέλου που προέκυψε από τον Inductive Miner με φίλτρο θορύβου 10% .....	95
5.4.3	Έλεγχος συμμόρφωσης του μοντέλου που προέκυψε από τον Inductive Miner με φίλτρο θορύβου 0% .....	96
5.4.4	Έλεγχος συμμόρφωσης του μοντέλου που προέκυψε από τον Heuristics Miner.....	97
5.4.5	Παρατηρήσεις από τη διαδικασία του Conformance Checking .....	99
5.5	Ανάλυση Απόδοσης (Performance Analysis).....	100
5.6	Σύνθεση συμπερασμάτων .....	101
6.	ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΜΕΛΛΟΝΤΙΚΗ ΕΡΓΑΣΙΑ .....	105
7.	ΒΙΒΛΙΟΓΡΑΦΙΑ - ΑΝΑΦΟΡΕΣ .....	107

## ΠΙΝΑΚΑΣ ΕΙΚΟΝΩΝ

Εικόνα 1: Internet of events (van der Aalst W. (2016) Data Science in Action. In: ProcessMining. Springer, Berlin, Heidelberg, σ. 5) .....	19
Εικόνα 2: Η εξόρυξη διαδικασιών ως γέφυρα (van der Aalst W. (2016) Data Science in Action. In: Process Mining. Springer, Berlin, Heidelberg, σ. 18).....	20
Εικόνα 3: Το πλαίσιο αναφοράς για την εξόρυξη διαδικασιών (van der Aalst W. (2016) Data Science in Action. In: Process Mining. Springer, Berlin, Heidelberg, σ. 32) .....	24
Εικόνα 4: Εξισορρόπηση των 4 παραμέτρων (van der Aalst W. (2016) Data Science in Action. In: Process Mining. Springer, Berlin, Heidelberg, σ. 189).....	25
Εικόνα 5: Το "flower model" το οποίο μπορεί να αναπαράγει κάθε ίχνος που περιέχει τις δραστηριότητες [a,b,...,h] ( <a href="http://mlwiki.org/index.php/Process_Mining">http://mlwiki.org/index.php/Process_Mining</a> ) .....	26
Εικόνα 6: Το πλαίσιο αναφοράς για την εξόρυξη εκπαιδευτικών διαδικασιών (Bogarín Alejandro & Cerezo Rebeca & Romero Cristóbal, A survey on educational process mining, 2017.) .....	35
Εικόνα 7: Η αρχιτεκτονική της πλατφόρμας SoftLearn (Barreiros BV, Lama M, Mucientes M, Vidal JC. Softlearn: a process mining platform for the discovery of learning paths, 2014) .....	41
Εικόνα 8: Διάγραμμα διαδικασιών Πανελλαδικών Εξετάσεων στο λογισμικό ADONIS.....	48
Εικόνα 9: Επισκόπηση της μεθοδολογίας για την εξόρυξη διαδικασιών (van Eck M.L., 2015, σ. 3) 50	
Εικόνα 10: Διαδικασία αίτησης-δήλωσης Πανελλαδικών Εξετάσεων και καταχώρησης βαθμολογίας στο ΠΣ-MySchool .....	55
Εικόνα 11: Κεφαλίδες αρχείου καταγραφής συμβάντων.....	57
Εικόνα 12: Επιλογή βασικών παραμέτρων .....	58
Εικόνα 13: Οπτικοποίηση πρώτου αρχείου καταγραφής συμβάντων .....	59
Εικόνα 14: Σύνοψη των γεγονότων του πρώτου αρχείου καταγραφής συμβάντων .....	59
Εικόνα 15: Σύνοψη των γεγονότων έναρξης-λήξης στο πρώτο αρχείο καταγραφής συμβάντων .....	60
Εικόνα 16: Στατιστικά στοιχεία απεικόνισης Explore Event Log.....	61
Εικόνα 17: Διαδρομές δραστηριοτήτων με φθίνουσα συχνότητα (1 <sup>ο</sup> event log) .....	61
Εικόνα 18: Διαδρομές δραστηριοτήτων με αύξον μήκος μονοπατιού (1 <sup>ο</sup> event log) .....	63
Εικόνα 19: Διαδρομές δραστηριοτήτων με φθίνον μήκος μονοπατιού (1 <sup>ο</sup> event log) .....	63
Εικόνα 20: Σύνοψη των γεγονότων του δεύτερου αρχείου καταγραφής συμβάντων .....	64
Εικόνα 21: Σύνοψη των γεγονότων έναρξης-λήξης στο δεύτερο αρχείο καταγραφής συμβάντων ..	65
Εικόνα 22: Διαδρομές δραστηριοτήτων με φθίνουσα συχνότητα (2 <sup>ο</sup> event log).....	66
Εικόνα 23: Αρχική απεικόνιση του αρχείου καταγραφής συμβάντων με τη χρήση του Dotted Chart.....	67
Εικόνα 24: Απεικόνιση αρχείου καταγραφής συμβάντων με Dotted Chart, ανά CaseID (ApplicationID) με χρήση διαφορετικού χρώματος ανά συμβάν .....	68
Εικόνα 25: Μελέτη προβληματικής αίτησης-δήλωσης I.....	69
Εικόνα 26: : Μελέτη προβληματικής αίτησης-δήλωσης II.....	69
Εικόνα 27: Κλασική έκδοση Alpha Miner.....	71
Εικόνα 28: Έκδοση Alpha+ του Alpha Miner.....	72
Εικόνα 29: Έκδοση Alpha++ του Alpha Miner .....	72
Εικόνα 30: Κλασική έκδοση Alpha Miner στο event log με στοιχεία από τα BK .....	73
Εικόνα 31: Έκδοση Alpha+ του Alpha Miner στο event log με στοιχεία από τα BK .....	73
Εικόνα 32: Αρχικό μοντέλο Heuristics Miner .....	74
Εικόνα 33: Heuristics Miner (petri-net).....	74

Εικόνα 34: Heuristics Miner στο event log με στοιχεία από τα ΒΚ.....	75
Εικόνα 35: Κατώφλι θορύβου στον Inductive Miner .....	76
Εικόνα 36: Inductive Miner με φίλτρο θορύβου 20%.....	76
Εικόνα 37: Inductive Miner με φίλτρο θορύβου 20% στο event log με στοιχεία από τα ΒΚ .....	76
Εικόνα 38: Inductive Miner με φίλτρο θορύβου 0%.....	77
Εικόνα 39: Inductive Miner με φίλτρο θορύβου 0% στο event log με στοιχεία από τα ΒΚ .....	77
Εικόνα 40: Inductive Visual Miner - Μονοπάτια και Αποκλίσεις.....	78
Εικόνα 41: Inductive Visual Miner - Μονοπάτια και Αποκλίσεις (α' μέρος).....	78
Εικόνα 42: Inductive Visual Miner - Μονοπάτια και Αποκλίσεις (β' μέρος).....	78
Εικόνα 43: Inductive Visual Miner - Μονοπάτια και Χρόνοι Αναμονής.....	79
Εικόνα 44: Inductive Visual Miner - Μονοπάτια και Χρόνοι Αναμονής (α' μέρος) .....	79
Εικόνα 45: Inductive Visual Miner - Μονοπάτια και Χρόνοι Αναμονής (β' μέρος) .....	79
Εικόνα 46: Inductive Visual Miner - Ροή Συμβάντων.....	80
Εικόνα 47: Fuzzy Model του πρώτου event log .....	80
Εικόνα 48: Fuzzy Model του event log με τα δεδομένα των ΒΚ (Significance cutoff 0%).....	80
Εικόνα 49: Επιλογή παραμέτρων στο Disco.....	81
Εικόνα 50: Γενική επισκόπηση στο Disco.....	82
Εικόνα 51: Κατανομή των συμβάντων ανά έτος γέννησης του υποψηφίου.....	83
Εικόνα 52: Κατανομή των συμβάντων ανά είδος Λυκείου.....	83
Εικόνα 53: Χάρτης διαδικασίας (100% Activities - 100% Paths).....	84
Εικόνα 54: Χάρτης διαδικασίας (100% Activities - 20% Paths).....	84
Εικόνα 55: Χάρτης μέσης διάρκειας δραστηριοτήτων .....	85
Εικόνα 56: Χάρτης διαδικασίας για τα Επαγγελματικά Λύκεια (100% Activities - 20% Paths).....	86
Εικόνα 57: Χάρτης διαδικασίας για το αρχείο με τα δεδομένα των ΒΚ (100% Activities - 20% Paths).....	87
Εικόνα 58: Χάρτης μέσης διάρκειας δραστηριοτήτων για το αρχείο με τα δεδομένα των ΒΚ.....	88
Εικόνα 59: Επιλογή αλγορίθμου ελέγχου συμμόρφωσης .....	90
Εικόνα 60: Παραμετροποίηση κοστών ελέγχου συμμόρφωσης .....	91
Εικόνα 61: Έλεγχος συμμόρφωσης πρώτου αρχείου καταγραφής συμβάντων με μοντέλο Inductive Miner και φίλτρο θορύβου 20% .....	91
Εικόνα 62: Έλεγχος συμμόρφωσης πρώτου αρχείου καταγραφής συμβάντων με μοντέλο Inductive Miner και φίλτρο θορύβου 20% (α' μέρος) .....	91
Εικόνα 63: Έλεγχος συμμόρφωσης πρώτου αρχείου καταγραφής συμβάντων με μοντέλο Inductive Miner και φίλτρο θορύβου 20% (β' μέρος) .....	92
Εικόνα 64: Εμφάνιση αποκλίσεων στην ΚΑΤΑΧΩΡΗΣΗ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ.....	93
Εικόνα 65: Εμφάνιση αποκλίσεων στην ΑΚΥΡΩΣΗ ΑΙΤΗΣΗΣ.....	93
Εικόνα 66: Συνολικά στατιστικά ελέγχου συμμόρφωσης (Inductive Miner – 20%).....	95
Εικόνα 67: Έλεγχος συμμόρφωσης πρώτου αρχείου καταγραφής συμβάντων με μοντέλο Inductive Miner και φίλτρο θορύβου 10%.....	95
Εικόνα 68: Συνολικά στατιστικά ελέγχου συμμόρφωσης (Inductive Miner – 10%).....	96
Εικόνα 69: Έλεγχος συμμόρφωσης πρώτου αρχείου καταγραφής συμβάντων με μοντέλο Inductive Miner .....	97
Εικόνα 70: Συνολικά στατιστικά ελέγχου συμμόρφωσης (Inductive Miner – 0%).....	97
Εικόνα 71: Έλεγχος συμμόρφωσης πρώτου αρχείου καταγραφής συμβάντων με μοντέλο Heuristics Miner (α').....	98

Εικόνα 72: Έλεγχος συμμόρφωσης πρώτου αρχείου καταγραφής συμβάντων με μοντέλο Heuristics Miner (β')	98
Εικόνα 73: Συνολικά στατιστικά ελέγχου συμμόρφωσης (Heuristics Miner)	99
Εικόνα 74: Ανάλυση απόδοσης δεύτερου αρχείου καταγραφής συμβάντων με την πληροφορία των ΒΚ, με μοντέλο Inductive Miner και φίλτρο θορύβου 20% (α')	100
Εικόνα 75: Ανάλυση απόδοσης δεύτερου αρχείου καταγραφής συμβάντων με την πληροφορία των ΒΚ, με μοντέλο Inductive Miner και φίλτρο θορύβου 20% (β')	101
Εικόνα 76: Διαγραμματική αναπαράσταση ενιαίας πλατφόρμας διαχείρισης διαδικασιών Πανελλαδικών Εξετάσεων	103





## 1. ΕΙΣΑΓΩΓΗ

Στη σύγχρονη εκπαίδευση, χρησιμοποιούνται διάφορα συστήματα πληροφοριών για την υποστήριξη εκπαιδευτικών διαδικασιών. Στις περισσότερες περιπτώσεις, αυτά τα συστήματα έχουν δυνατότητα καταγραφής για έλεγχο και παρακολούθηση των διαδικασιών που υποστηρίζουν. Τα πληροφοριακά συστήματα που χρησιμοποιούνται συλλέγουν και αποθηκεύουν δεδομένα σχετικά με τους μαθητές, την εγγραφή τους σε συγκεκριμένα προγράμματα και μαθήματα, τις προτιμήσεις τους καθώς και τις επιδόσεις τους. Αυτά τα δεδομένα μπορούν να αναλυθούν σε διάφορα επίπεδα και από διάφορες προοπτικές, δείχνοντας έτσι ποικίλες πτυχές της οργάνωσης του εκπαιδευτικού συστήματος και δίνοντάς μας περισσότερες πληροφορίες για τη συνολική εκπαιδευτική διαδικασία.

Πρόσφατα, οι παραδοσιακές τεχνικές εξόρυξης δεδομένων έχουν εφαρμοστεί εκτενώς για εύρεση μοτίβων που παρουσιάζουν ενδιαφέρον καθώς και δημιουργία περιγραφικών και προγνωστικών μοντέλων από μεγάλους όγκους δεδομένων που συσσωρεύονται μέσω της χρήσης διαφόρων πληροφοριακών συστημάτων. Τα αποτελέσματα της εξόρυξης δεδομένων μπορούν να χρησιμοποιηθούν για την καλύτερη κατανόηση της εκπαιδευτικής διαδικασίας, για τη δημιουργία συστάσεων και συμβουλών προς τους μαθητές, για βελτίωση της διαχείρισης πόρων κ.τ.λ. (Trcka N., 2011).

Ωστόσο, οι περισσότερες από τις παραδοσιακές τεχνικές εξόρυξης δεδομένων δεν επικεντρώνονται στη διαδικασία στο σύνολό της. Δεν στοχεύουν στην ανακάλυψη ή την ανάλυση της πλήρους εκπαιδευτικής διαδικασίας και έτσι δεν είναι σαφές για παράδειγμα πώς, δεδομένης μίας συγκεκριμένης διεργασίας, θα μπορούσε κανείς να ελέγξει κατά πόσο οι μαθητές την ακολουθούν πάντα. Δεν είναι επίσης δυνατό να έχουμε μια σαφή οπτική αναπαράσταση ολόκληρης της διαδικασίας. Για να επιτραπεί αυτού του είδους η ανάλυση (στην οποία η διαδικασία παίζει τον κεντρικό ρόλο), δημιουργήθηκε μία νέα κατηγορία εξόρυξης δεδομένων, η οποία ονομάζεται εξόρυξη διαδικασιών (Process Mining).

Η εξόρυξη διαδικασιών προέκυψε από την επιχειρηματική κοινότητα. Επικεντρώνεται στην ανάπτυξη ενός συνόλου έξυπνων εργαλείων και τεχνικών που αποσκοπούν στην εξαγωγή γνώσης σχετικής με τις διαδικασίες από αρχεία καταγραφής συμβάντων (event logs).

Στην παρούσα εργασία εξετάζεται η δυνατότητα εφαρμογής της εξόρυξης διαδικασιών, με τη βοήθεια δύο εργαλείων εξόρυξης (ProM και Disco–Academic License), σε εκπαιδευτικές διαδικασίες σχετικές με τις πανελλαδικές εξετάσεις. Η συγκεκριμένη μελέτη περίπτωσης εστιάζει στην ανακάλυψη ενός μοντέλου που αφορά στη συνολική εκπαιδευτική διαδικασία οργάνωσης των πανελλαδικών εξετάσεων και στην εξέταση συμμόρφωσης με τη διαδικασία όπως αυτή προκύπτει από τα αρχεία καταγραφής συμβάντων. Στόχος είναι να παρακολουθήσουμε τη διαδικασία των πανελλαδικών εξετάσεων μέσω των δεδομένων που έχουν καταγραφεί ηλεκτρονικά στα

πληροφοριακά συστήματα και να ανακαλύψουμε εάν αυτό που πραγματικά συμβαίνει συνάδει με αυτό που θα έπρεπε να συμβαίνει ή εάν υπάρχουν αποκλίσεις από την καταγεγραμμένη διαδικασία καθώς και σημαντικές καθυστερήσεις σε ορισμένα σημεία. Λαμβάνοντας υπόψη τις ανωτέρω παρατηρήσεις θα διατυπωθούν προτάσεις βελτίωσης του συνόλου της διαδικασίας ή/και των επιμέρους διεργασιών της. Είναι αξιοσημείωτο πως η εξόρυξη μιας τέτοιας διαδικασίας που αφορά μεν στην εκπαίδευση αλλά στον διοικητικό τομέα αυτής κι όχι σε αυτόν της μάθησης, είναι κάτι που δεν είναι σύνηθες καθώς ο κύριος όγκος της διαθέσιμης βιβλιογραφίας και αρθρογραφίας αφορά την εξόρυξη διαδικασιών στον τομέα της μάθησης (learning).

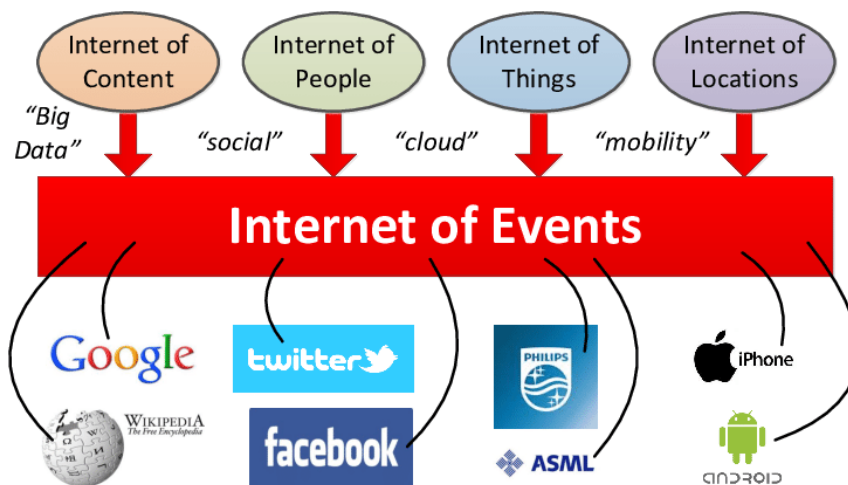
Η δομή της εργασίας έχει ως εξής:

Αρχικά, στο Κεφάλαιο 2, γίνεται μία εισαγωγή στις έννοιες της εξόρυξης διαδικασιών, μία αξιολόγησή της και τέλος μια παρουσίαση των διαθέσιμων εργαλείων εξόρυξης. Στη συνέχεια, στο Κεφάλαιο 3, πραγματοποιείται μια επικέντρωση στην έννοια της εξόρυξης εκπαιδευτικών διαδικασιών και στους τομείς εφαρμογών της. Οδεύοντας προς το σκοπό της παρούσας εργασίας, στο Κεφάλαιο 4 αρχικά περιγράφονται οι εκπαιδευτικές διαδικασίες οργάνωσης των Πανελλαδικών Εξετάσεων, όπως αυτές τηρούνται και εφαρμόζονται από το Υπουργείο Παιδείας και Θρησκευμάτων, και στη συνέχεια τα βήματα για τη συλλογή και την προετοιμασία των διαθέσιμων δεδομένων των διαδικασιών αυτών προκειμένου να γίνουν αντικείμενο επεξεργασίας από τα εργαλεία εξόρυξης. Ακολουθεί στο Κεφάλαιο 5 η παρουσίαση των βημάτων για την εξόρυξη της διαδικασίας των Πανελλαδικών Εξετάσεων, η ανακάλυψη, ο έλεγχος συμμόρφωσης και η ανάλυση των αποτελεσμάτων που ακολουθείται από προτάσεις βελτίωσης. Τέλος, στο Κεφάλαιο 6 καταγράφονται οι παρατηρήσεις και τα συμπεράσματα που εξήχθησαν καθώς και οι κατευθύνσεις για πιθανή μελλοντική εργασία.

## 2. ΔΙΑΧΕΙΡΙΣΗ ΚΑΙ ΕΞΟΡΥΞΗ ΔΙΑΔΙΚΑΣΙΩΝ (PROCESS MINING)

### 2.1 Εισαγωγή

Σήμερα πολλοί άνθρωποι λένε ότι «τα δεδομένα είναι το νέο πετρέλαιο», και αυτό δείχνει τις απίστευτες ποσότητες δεδομένων που συλλέγονται και την τεράστια αξία τους. Στην καθημερινή ζωή καταγράφονται πολλά δεδομένα και γεγονότα, π.χ. η παραγγελία ενός βιβλίου, η πραγματοποίηση μιας τηλεφωνικής κλήσης, η πραγματοποίηση μιας ιατρικής εξέτασης, μία αγορά με την πιστωτική μας κάρτα κ.ο.κ. Αυτά και πολλά άλλα παραδείγματα αποδεικνύουν ότι τα δεδομένα αποθηκεύονται συνεχώς και με κάθε είδους τρόπο (van der Aalst W., 2016).



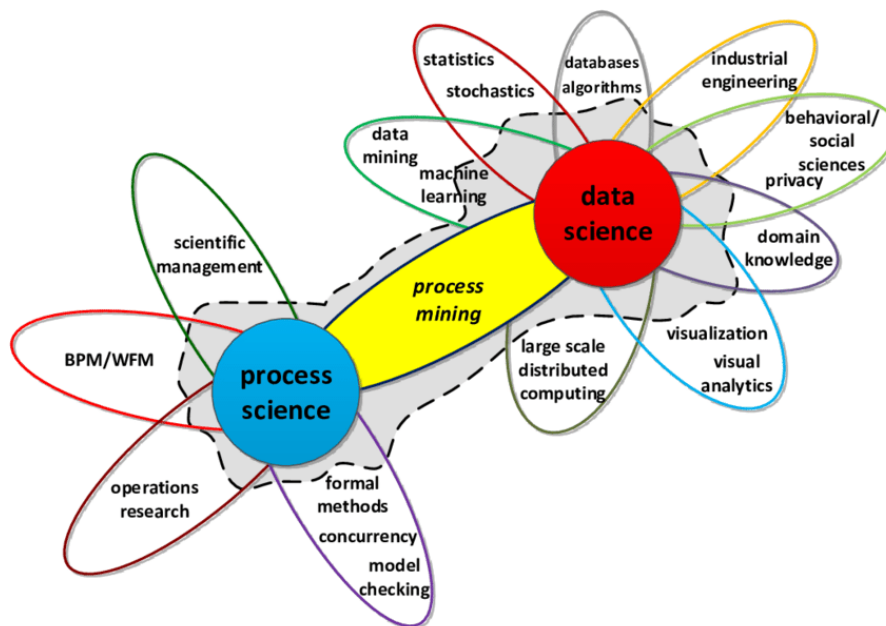
Εικόνα 1: Internet of events (van der Aalst W. (2016) Data Science in Action. In: ProcessMining. Springer, Berlin, Heidelberg, σ. 5)

Για το λόγο αυτό πολλοί μιλούν σήμερα για «μεγάλα δεδομένα» (Big Data), τα οποία είναι στην ουσία οι τεράστιες ποσότητες δεδομένων-συμβάντων που καταγράφονται και αποθηκεύονται. Υπάρχουν πολλές πηγές που είχαν προβλέψει την εκθετική αύξηση των δεδομένων έως το 2020 και στη συνέχεια, ωστόσο η πρόκληση σήμερα δεν είναι να δημιουργήσουμε περισσότερα δεδομένα αλλά να μετατρέψουμε αυτά τα δεδομένα σε πραγματική αξία.

Όσον αφορά στην επιστήμη των δεδομένων υπάρχουν ερωτήματα όπως τι συνέβη, γιατί συνέβη, τι θα συμβεί στο μέλλον και τέλος ποιο είναι το καλύτερο που θα μπορούσε να συμβεί. Υπάρχει επομένως η ανάγκη για ανάλυση των δεδομένων ώστε να μπορεί κανείς να προτείνει πράγματα που θα βελτιώσουν μια κατάσταση ή μια διαδικασία.

Η ανάγκη αυτή οδήγησε στην ανάπτυξη της έννοιας της εξόρυξης διαδικασιών (process mining), η οποία γεφυρώνει το χάσμα μεταξύ της κλασικής ανάλυσης μοντέλου διεργασιών (process model analysis) και της ανάλυσης προσανατολισμένης σε δεδομένα όπως η εξόρυξη δεδομένων (data mining). Η εξόρυξη διαδικασιών γεφυρώνει

αυτό το κενό, επειδή εστιάζει μεν στις διαδικασίες, αλλά ταυτόχρονα χρησιμοποιώντας τα πραγματικά δεδομένα (van der Aalst W., 2016).



Εικόνα 2: Η εξόρυξη διαδικασιών ως γέφυρα (van der Aalst W. (2016) Data Science in Action. In: Process Mining. Springer, Berlin, Heidelberg, σ. 18)

### Εφαρμογές της Εξόρυξης Διαδικασιών

Στη σύγχρονη ψηφιακή οικονομία μας, οι εταιρείες και οι οργανισμοί απαιτούν ευέλικτες διαδικασίες προκειμένου να είναι ανταγωνιστικές. Αυτή η ευελιξία μπορεί να έρθει μόνο μέσω μιας βαθιάς κατανόησης του τρόπου λειτουργίας και των δυνατοτήτων που αυτός παρέχει. Πραγματοποιώντας εξόρυξη διαδικασιών έχουμε τη δυνατότητα να απαντήσουμε σε ερωτήσεις που σχετίζονται με την απόδοση των διαδικασιών καθώς και ερωτήσεις σχετικά με τη συμμόρφωση του μοντέλου διεργασιών με τα πραγματικά συμβάντα. Βγάζουμε πολύτιμα συμπεράσματα για τους λόγους για τους οποίους υπάρχουν ορισμένα σημεία συμμόρφωσης στις διαδικασίες, για το πώς μπορούν να αφαιρεθούν ή, τέλος, γιατί ορισμένες φορές υπάρχουν παρεκκλίσεις από το μοντέλο. Επιπλέον, μέσω της εξόρυξης διαδικασιών μπορούμε να αξιολογούμε γρήγορα τα δεδομένα μας και να ελέγχουμε τις διαδικασίες, με στόχο την βελτιστοποίησή τους.

Γενικότερα, όλοι οι οργανισμοί χρησιμοποιούν διαδικασίες για τον χειρισμό των διεργασιών και των περιπτώσεών τους (cases). Μερικές φορές αυτές οι διαδικασίες επιβάλλονται από το σύστημα πληροφοριών. Ωστόσο, στις περισσότερες περιπτώσεις, δεν υπάρχει τεκμηρίωση αυτών των διαδικασιών και ακόμη κι όταν υπάρχει, η πραγματικότητα μπορεί να είναι πολύ διαφορετική. Αυτό καθιστά σημαντική την ανακάλυψη των πραγματικών διαδικασιών χρησιμοποιώντας δεδομένα καταγραφής συμβάντων. Τα ανακαλυφθέντα μοντέλα διεργασιών θα μπορούσαν να

χρησιμοποιηθούν για τη συζήτηση προβλημάτων μεταξύ των ενδιαφερομένων μερών (για την επίτευξη συμφωνίας και προόδου είναι σημαντικό να υπάρχει άποψη για τις πραγματικές διαδικασίες), για την ανάπτυξη ιδεών βελτίωσης της διαδικασίας (βλέποντας την πραγματική διαδικασία και τα προβλήματά της διεγείρονται προσπάθειες επανασχεδιασμού), για βελτίωση του υπάρχοντος μοντέλου διεργασιών (π.χ. ανάλυση των σημείων συμφόρησης) και τέλος για τη διαμόρφωση ενός μοντέλου διεργασιών το οποίο μπορεί να χρησιμεύσει ως πρότυπο (van der Aalst W., 2012).

Οι τεχνικές ελέγχου συμμόρφωσης συσχετίζουν συμβάντα στο αρχείο καταγραφής με δραστηριότητες στο μοντέλο διεργασιών. Με αυτόν τον τρόπο είναι δυνατόν να συγκριθεί η συμπεριφορά που παρατηρήθηκε στο αρχείο καταγραφής συμβάντων με τη συμπεριφορά όπως αυτή αναπαριστάται βάσει των μοντέλων. Μπορεί κανείς, για παράδειγμα, να ποσοτικοποιήσει τις διαφορές (π.χ., το 80% των παρατηρούμενων περιπτώσεων είναι δυνατό σύμφωνα με το μοντέλο) και να διαγνώσει αποκλίσεις (π.χ., στην πραγματικότητα η δραστηριότητα  $x$  παραλείπεται αν και το μοντέλο δεν το επιτρέπει). Ο έλεγχος συμμόρφωσης μπορεί να χρησιμοποιηθεί για να ελεγχθεί η ποιότητα των τεκμηριωμένων διαδικασιών (κατά πόσο δηλαδή περιγράφουν την πραγματικότητα με ακρίβεια), για τον εντοπισμό περιπτώσεων που παρεκκλίνουν και την εξέταση και κατανόηση κοινών τους χαρακτηριστικών, για τον προσδιορισμό τμημάτων της διαδικασίας στα οποία συμβαίνουν οι περισσότερες αποκλίσεις, για να κριθεί η ποιότητα ενός ανακαλυφθέντος μοντέλου διεργασίας, για την καθοδήγηση εξελικτικών αλγορίθμων ανακάλυψης διαδικασίας (υπάρχουν π.χ. γενετικοί αλγόριθμοι οι οποίοι πρέπει να αξιολογούν συνεχώς την ποιότητα των νέων μοντέλων χρησιμοποιώντας έλεγχο συμμόρφωσης), και τέλος ως σημείο εκκίνησης για την μελλοντική βελτίωση του μοντέλου (van der Aalst W., 2012).

## 2.2 Διαδικασίες

Μια διαδικασία είναι μια σειρά ενεργειών ή βημάτων που επαναλαμβάνονται σε εξέλιξη από μία καθορισμένη «αρχή» σε ένα καθορισμένο «τέλος». Ο σκοπός μιας διαδικασίας είναι να καθιερώσει και να διατηρήσει μια κοινώς κατανοητή ροή ενεργειών για να επιτρέψει την ολοκλήρωση μιας εργασίας όσο το δυνατόν πιο αποτελεσματικά και με συνέπεια.

Συνήθως, σε πρώτο επίπεδο οι διαδικασίες φαίνονται πολύ δομημένες και σαφείς αλλά στην πραγματικότητα υπάρχουν πολλές αποκλίσεις και διαφοροποιήσεις μεταξύ της επίσημα καταγεγραμμένης διαδικασίας και αυτής που εντέλει καταγράφεται στα υπάρχοντα συστήματα.

Η εξόρυξη διαδικασιών έρχεται ως λύση για τη λήψη πληροφοριών από την κατάλληλη επεξεργασία των καταγεγραμμένων δεδομένων, την απάντηση ερωτήσεων και την επίλυση προβλημάτων. Συνήθως κατά τη διαδικασία της εξόρυξης η αναζήτηση είναι προσανατολισμένη προς συγκεκριμένες προκλήσεις ή εμπόδια.

Οι τεχνικές εξόρυξης διαδικασιών λειτουργούν καλά σε δομημένες διαδικασίες με μικρή αποκλίνουσα συμπεριφορά και ισχυρές εξαρτήσεις μεταξύ των βημάτων της διαδικασίας. Ωστόσο, οι περισσότερες και πιο ενδιαφέρουσες διαδικασίες είναι συνήθως πολύ πιο δύσκολο να εξορυχτούν. Οι μη δομημένες διαδικασίες είναι συνήθως πολύπλοκες στη διαχείριση αλλά έχουν μεγάλες δυνατότητες βελτίωσης κι επομένως ο στόχος της εξόρυξης διαδικασιών είναι να ανακαλύψει και να περιγράψει τη δομή σε μη δομημένες διαδικασίες (van der Aalst W., 2007).

Στον τομέα της εκπαίδευσης για παράδειγμα, οι περισσότερες διαδικασίες είναι μη δομημένες, καθώς υπεισέρχεται σε αυτές σε μεγάλο βαθμό και ο ανθρώπινος παράγοντας ο οποίος δημιουργεί περισσότερες και ορισμένες φορές απρόβλεπτες εναλλακτικές. Ωστόσο η διοικητική υποστήριξη της διαδικασίας οργάνωσης των Πανελλαδικών Εξετάσεων είναι μια δομημένη διαδικασία, η οποία υπάγεται σε συγκεκριμένο πλαίσιο.

### **2.3 Αρχεία Καταγραφής Συμβάντων (Event Logs)**

Τα ψηφιακά δεδομένα είναι παντού - σε κάθε τομέα, σε κάθε οικονομία, σε κάθε οργανισμό και σε κάθε σπίτι - και θα συνεχίσουν να αυξάνονται εκθετικά (Manyika J., 2011). Η πανταχού παρουσία τέτοιων δεδομένων είναι αυτή που επιτρέπει νέες μορφές ανάλυσης διεργασιών, όπως η εξόρυξη διαδικασιών, οι οποίες βασίζονται πλέον σε παρατηρούμενα γεγονότα και όχι σε κατασκευασμένα στο χέρι μοντέλα (van der Aalst W., 2012).

Για να είναι δυνατή η εξόρυξη διαδικασιών, οι απαιτούμενες πληροφορίες και τα δεδομένα πρέπει να είναι δομημένα σε μορφή ρητών αρχείων καταγραφής συμβάντων (event logs), δηλαδή σε αρχεία στα οποία καταγράφονται και αποθηκεύονται, συνήθως από τα πληροφοριακά συστήματα, τα διάφορα ψηφιακά δεδομένα και γεγονότα που πραγματοποιούνται σε συγκεκριμένες διαδικασίες.

Τα αρχεία καταγραφής συμβάντων συχνά αναφέρονται και με τον όρο «ιστορικό» και φανερώνουν τη σειρά των δραστηριοτήτων για την ολοκλήρωση μιας διαδικασίας και τη χρονική περίοδο που χρειάζεται για τη λήξη της. Κάθε διαδικασία αποτελείται από περιπτώσεις (cases) που αποτελούν στιγμιότυπα της διαδικασίας αυτής. Κάθε περίπτωση αποτελείται από γεγονότα τα οποία καταγράφονται με χρονική σειρά με το κάθε γεγονός να σχετίζεται με μια συγκεκριμένη περίπτωση. Τα γεγονότα μέσα σε μια περίπτωση μπορούν να έχουν και γνωρίσματα όπως η δραστηριότητα, ο χρόνος, το κόστος και η πηγή, καθώς και επιπλέον χαρακτηριστικά ανάλογα με το τι πληροφορία αποθηκεύεται στα αρχεία καταγραφής συμβάντων.

Σήμερα, όλες οι τεχνικές εξόρυξης διαδικασιών προϋποθέτουν ότι είναι δυνατή μια τέτοια καταγραφή της αλληλουχίας των σχετικών συμβάντων που σημειώθηκαν σε μια επιχείρηση ή έναν οργανισμό, έτσι ώστε κάθε συμβάν να αναφέρεται σε μια δραστηριότητα (δηλαδή, ένα καλά καθορισμένο βήμα κάποιας διαδικασίας) και να

σχετίζεται με μια συγκεκριμένη περίπτωση και την χρονική στιγμή που πραγματοποιείται. Τα συμβάντα ενδέχεται να έχουν επιπλέον πληροφορίες που αποθηκεύονται στα αρχεία καταγραφής συμβάντων και, όποτε είναι δυνατόν, οι τεχνικές εξόρυξης διαδικασιών χρησιμοποιούν αυτές τις επιπλέον πληροφορίες, όπως π.χ. την ακριβή χρονική σήμανση στην οποία το συμβάν έχει καταγραφεί (timestamp), τον πόρο (δηλαδή, άτομο ή συσκευή) που δημιούργησε το συμβάν, τον τύπο συμβάντος στο πλαίσιο του κύκλου ζωής συναλλαγής δραστηριότητας (π.χ. εάν η δραστηριότητα έχει ξεκινήσει, ακυρωθεί ή ολοκληρωθεί) ή διάφορα άλλα στοιχεία δεδομένων που καταγράφηκαν μαζί με το συμβάν (π.χ. το μέγεθος μιας παραγγελίας) (Janes A., 2017).

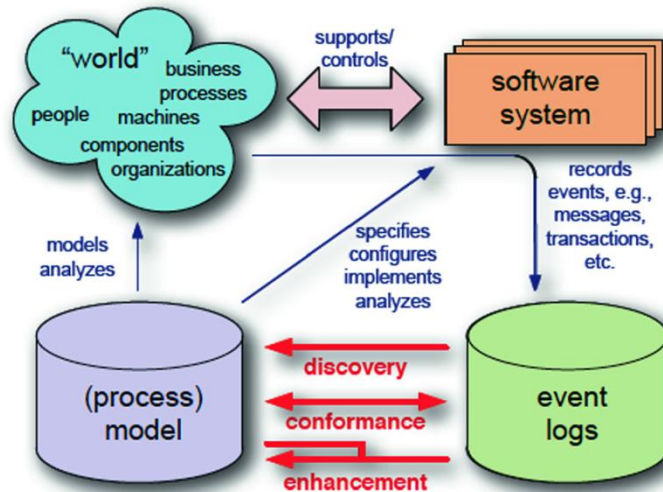
Υπάρχει επομένως ένας πραγματικός κόσμος όπου συμβαίνουν τα γεγονότα και συνήθως υπάρχουν αντίστοιχα πληροφοριακά συστήματα που, με κάποιον τρόπο, καταγράφουν τα γεγονότα αυτά. Στην ιδανική περίπτωση το πληροφοριακό σύστημα περιέχει πίνακες με μορφή συμβατή με τους αλγόριθμους εξόρυξης διαδικασιών, δηλαδή το case id, το event και το timestamp, οπότε στην ουσία υπάρχει έτοιμο και το αρχείο καταγραφής συμβάντων. Σε άλλες περιπτώσεις, όπου τα αρχεία του πληροφοριακού συστήματος έχουν διαφορετική δομή, είναι δυνατό να δημιουργηθούν ερωτήματα προς τη βάση δεδομένων και με τον τρόπο αυτό να εξαχθεί η πληροφορία σε μορφή συμβατή με την εξόρυξη διαδικασιών. Έτσι προκύπτουν τελικά τα αρχεία καταγραφής συμβάντων και στη συνέχεια μπορούμε να κάνουμε εξόρυξη διαδικασιών, μπορούμε να ανακαλύψουμε ένα μοντέλο διεργασιών, μπορούμε να κάνουμε έλεγχο συμμόρφωσης συγκρίνοντας το αρχείο καταγραφής συμβάντων με το μοντέλο διεργασιών που προέκυψε και τέλος μπορούμε να εμπλουτίσουμε το μοντέλο αυτό με πληροφορίες σχετικά με αποκλίσεις και επιδόσεις, όπως για παράδειγμα τα σημεία συμμόρφωσης σε μια διαδικασία.

## **2.4 Εξόρυξη Διαδικασιών (Process Mining)**

Η εξόρυξη διαδικασιών είναι ένας σχετικά νέος ερευνητικός τομέας που στοχεύει στην ανακάλυψη, παρακολούθηση και βελτίωση πραγματικών διαδικασιών με εξαγωγή της γνώσης από αρχεία καταγραφής συμβάντων που είναι άμεσα διαθέσιμα στα σύγχρονα πληροφοριακά συστήματα.

Ο όρος εξόρυξη διαδικασιών χρησιμοποιείται σ' ένα ευρύτερο πλαίσιο για να αναφέρεται όχι μόνο σε τεχνικές για την ανακάλυψη και προσομοίωση μοντέλων διεργασιών αλλά και σε τεχνικές για την ανάλυση απόδοσης με βάση τα αρχεία καταγραφής συμβάντων σε συνδυασμό με τα δεδομένα εξόρυξης εργασιών, τον έλεγχο συμμόρφωσης που αναφέρεται στην παρακολούθηση των αποκλίσεων μεταξύ πρότυπου μοντέλου διεργασίας και ιστορικού γεγονότων, την εξόρυξη κοινωνικών δικτύων/δικτύων οργανισμού, την επέκταση ή επιδιόρθωση των μοντέλων, την πρόγνωση περιστατικών και τις συστάσεις βάσει του ιστορικού (Aalst W. Van Der et al.,

2012). Κατά τη διάρκεια της εξόρυξης διαδικασιών εξειδικευμένοι αλγόριθμοι εξόρυξης δεδομένων εφαρμόζονται σε δεδομένα καταγραφής συμβάντων προκειμένου να εντοπιστούν τάσεις και μοτίβα που αφορούν τα συμβάντα. Έτσι, παρέχεται αντικειμενική και βασισμένη στα πραγματικά γεγονότα γνώση η οποία βοηθά στον έλεγχο, την ανάλυση και τη βελτίωση των υπάρχουσών διαδικασιών, απαντώντας σε ερωτήσεις που σχετίζονται με τη συμμόρφωση και την απόδοση (Aalst W. Van Der, 2011).



Εικόνα 3: Το πλαίσιο αναφοράς για την εξόρυξη διαδικασιών (van der Aalst W. (2016) Data Science in Action. In: Process Mining. Springer, Berlin, Heidelberg, σ. 32)

Από τη μία πλευρά λοιπόν η εξόρυξη διαδικασιών εξετάζει εννοιολογικά μοντέλα που περιγράφουν τις διαδικασίες και τα αντίστοιχα σχετικά δεδομένα και από την άλλη πλευρά επικεντρώνεται στην πραγματική εκτέλεση των διαδικασιών, όπως αυτή αντικατοπτρίζεται στο αποτύπωμα της πραγματικότητας που καταγράφεται και αποθηκεύεται από τα εκάστοτε πληροφοριακά συστήματα (Janes A., 2017).

## 2.5 Τεχνικές εξόρυξης διαδικασιών

Οι τρεις βασικοί τύποι τεχνικών εξόρυξης διαδικασιών επισημαίνονται στο κάτω μέρος της Εικόνας 3.

Η ανακάλυψη (discovery) ξεκινά από ένα αρχείο καταγραφής συμβάντων και παράγει αυτόματα ένα μοντέλο διεργασιών που αναπαριστά και εξηγεί τις διαφορετικές συμπεριφορές που παρατηρούνται στο αρχείο καταγραφής συμβάντων και χωρίς την προϋπόθεση ύπαρξης προηγούμενης γνώσης σχετικά με τη διαδικασία. Η συντριπτική πλειονότητα των αλγορίθμων ανακάλυψης διαδικασιών επικεντρώνεται στην ανακάλυψη της διαδικασίας ελέγχου-ροής (control-flow), δημιουργώντας ένα μοντέλο που υποδεικνύει τις επιτρεπόμενες ακολουθίες δραστηριοτήτων σύμφωνα με το αρχείο καταγραφής συμβάντων.

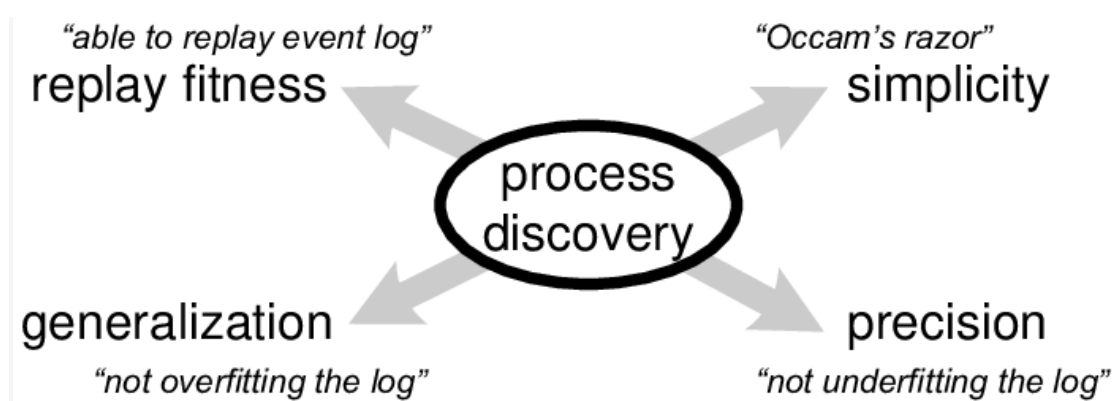


Ο έλεγχος συμμόρφωσης (conformance checking) συγκρίνει ένα υπάρχον μοντέλο διεργασιών και ένα αρχείο καταγραφής συμβάντων για τη διαδικασία αυτή, με κύριο στόχο την κατανόηση της παρουσίας πιθανών αποκλίσεων καθώς και της φύσης αυτών των αποκλίσεων. Οι τεχνικές ελέγχου συμμόρφωσης λαμβάνουν ως είσοδο ένα αρχείο καταγραφής συμβάντων και ένα μοντέλο διεργασιών και επιστρέφουν ενδείξεις σχετικά με την αντιστοιχία και την τήρηση των συμπεριφορών που περιέχονται στο αρχείο καταγραφής σε σχέση με αυτές τις οποίες αναπαριστά το μοντέλο.

Τέλος, η βελτίωση (enhancement) στοχεύει στον εμπλουτισμό ενός υπάρχοντος μοντέλου διεργασιών χρησιμοποιώντας πληροφορίες που έχουν εγγραφεί σε ένα αρχείο καταγραφής συμβάντων για αυτήν τη διαδικασία, έτσι ώστε το τελικό μοντέλο διεργασιών να αναπαριστά με πιο πιστό τρόπο την πραγματικότητα. Οι τεχνικές βελτίωσης λαμβάνουν ως είσοδο ένα μοντέλο διεργασιών και ένα αρχείο καταγραφής συμβάντων παράγουν ως έξοδο είναι ένα νέο μοντέλο διεργασιών που ενσωματώνει και αντανακλά νέες πληροφορίες εξάγονται από τα δεδομένα αυτά (Janes A., 2017).

## 2.6 Αξιολόγηση (Evaluation): Διάγνωση διαφορών μεταξύ παρατηρούμενης συμπεριφοράς και μοντελοποιημένης συμπεριφοράς

Προκειμένου να γίνει η εκτίμηση κατά πόσο το μοντέλο διεργασιών που έχουμε ανακαλύψει ανταποκρίνεται στην πραγματικότητα, συνήθως εξετάζονται τέσσερις ποιοτικές παράμετροι για τη σύγκριση του μοντέλου και του αρχείου καταγραφής συμβάντων: (α) καταλληλότητα (fitness), (β) απλότητα (simplicity), (γ) ακρίβεια (precision) και (δ) γενίκευση (generalization) (κεφ. 7, Aalst W. Van Der, 2011).

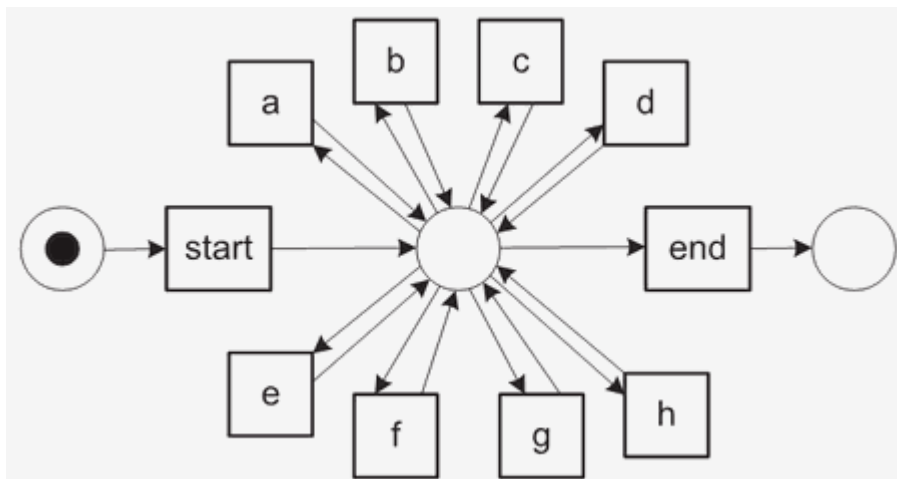


Εικόνα 4: Εξισορρόπηση των 4 παραμέτρων (van der Aalst W. (2016) Data Science in Action. In: Process Mining. Springer, Berlin, Heidelberg, σ. 189)

Σύμφωνα με το βιβλίο "Process Mining: Discovery, Conformance and Enhancement of Business Processes" και το άρθρο "Process Mining: Overview and Opportunities" του W. Van Der Aalst, οι τέσσερις αυτές παράμετροι αναλύονται ως εξής:

Ένα μοντέλο με μεγάλο βαθμό καταλληλότητας σημαίνει ότι το μοντέλο αυτό συνάδει σε μεγάλο βαθμό με το event log, αναπαράγει δηλαδή το μεγαλύτερο μέρος της συμπεριφοράς που εμφανίζεται στο αρχείο καταγραφής συμβάντων χωρίς να υπάρχουν αποκλίσεις. Ένα μοντέλο έχει τέλειο fitness εάν όλα τα ίχνη στο αρχείο καταγραφής συμβάντων μπορούν να αναπαραχθούν από το μοντέλο από την αρχή έως το τέλος. Συχνά το fitness περιγράφεται από έναν αριθμό μεταξύ 0 (πολύ κακό fitness) και 1 (τέλειο fitness).

Η παράμετρος της απλότητας υποδεικνύει το βαθμό πολυπλοκότητας του παραγόμενου μοντέλου κι ο στόχος είναι να είναι όσο το δυνατό πιο απλό το μοντέλο διεργασιών το οποίο έχει ανακαλυφθεί. Προφανώς, το απλούστερο μοντέλο που μπορεί να εξηγήσει τη συμπεριφορά που απεικονίζεται στο αρχείο καταγραφής συμβάντων είναι και το καλύτερο μοντέλο. Όσο πιο πολύπλοκη είναι η διαδικασία και όσο πιο πολλά δεδομένα περιέχει το αρχείο καταγραφής συμβάντων, τόσο πιο πιθανό είναι το μοντέλο που θα προκύψει να μην είναι εύκολα αναγνώσιμο ώστε να μπορεί κανείς να κατανοήσει και να εξηγήσει εύκολα τη βασική δομή των δραστηριοτήτων που συμβαίνουν, κάτι που είναι γνωστό ως φαινόμενο spaghetti στην εξόρυξη διαδικασιών.



Εικόνα 5: Το "flower model" το οποίο μπορεί να αναπαράγει κάθε ίχνη που περιέχει τις δραστηριότητες [a,b,...,h] ([http://mlwiki.org/index.php/Process\\_Mining](http://mlwiki.org/index.php/Process_Mining))

Οι παράμετροι fitness και simplicity από μόνες τους δεν αρκούν για να κριθεί η ποιότητα ενός ανακαλυφθέντος μοντέλου διεργασιών. Για παράδειγμα, είναι πολύ εύκολο να κατασκευάσετε ένα εξαιρετικά απλό μοντέλο ("flower model", Εικόνα 5) που μπορεί να αναπαράγει όλα τα ίχνη σε ένα αρχείο καταγραφής συμβάντων, αλλά και σε οποιοδήποτε άλλο αρχείο καταγραφής συμβάντων το οποίο αναφέρεται στο ίδιο σύνολο δραστηριοτήτων. Επίσης, δεν είναι επιθυμητό να υπάρχει ένα μοντέλο που επιτρέπει μόνο την ακριβή συμπεριφορά που εμφανίζεται στο αρχείο καταγραφής συμβάντων, διότι το συγκεκριμένο αρχείο καταγραφής περιέχει μόνο παραδείγματα συμπεριφορών και πολλά άλλα πιθανά ίχνη ενδέχεται να μην έχουν εμφανιστεί ακόμη,

παρόλα αυτά θα θέλαμε να υπάρχει η δυνατότητα αναπαραγωγής τους από το ανακαλυφθέν μοντέλο διεργασιών.

Ένα μοντέλο είναι ακριβές εάν δεν επιτρέπει να αναπαράγεται σ' αυτό «υπερβολική» συμπεριφορά, δηλαδή πάρα πολλά άλλα ίχνη εκτός από αυτά που περιέχονται στο αρχείο καταγραφής συμβάντων. Είναι σαφές ότι το «flower model» που απεικονίζεται παραπάνω στερείται ακρίβειας. Ένα μοντέλο που δεν είναι ακριβές λέγεται ότι είναι *underfitting*, πράγμα που σημαίνει ότι το μοντέλο αυτό υπεργενικεύει τη συμπεριφορά των παραδειγμάτων του αρχείου καταγραφής και επιτρέπει αρκετές συμπεριφορές πολύ διαφορετικές από αυτές που φαίνονται στο αρχείο αυτό.

Η παράμετρος της γενίκευσης υποδεικνύει το κατά πόσο το μοντέλο θα έχει τη δυνατότητα να «γενικεύει», δηλαδή να μην περιορίζει τη συμπεριφορά του μόνο στα παραδείγματα που εμφανίζονται στο αρχείο καταγραφής αλλά να μπορεί να αναπαράγει και συμπεριφορές που θα περιέχονται σε μελλοντικά αρχεία καταγραφής συμβάντων της ίδιας διαδικασίας. Ένα μοντέλο που δεν ικανοποιεί επαρκώς την παράμετρο της γενίκευσης λέγεται ότι είναι *overfitting*, δηλαδή πρόκειται για τη δημιουργία ενός πολύ συγκεκριμένου μοντέλου όπου είναι προφανές ότι μόνο το συγκεκριμένο αρχείο καταγραφής συμβάντων μπορεί να αναπαραχθεί αλλά κάποιο άλλο αρχείο που αφορά την ίδια διαδικασία δεν θα ικανοποιείται και θα παράγει ένα εντελώς διαφορετικό μοντέλο διεργασιών.

Είναι εμφανές ότι η παράμετρος της γενίκευσης είναι αντικρουόμενη μ' αυτή της ακρίβειας, ωστόσο ο στόχος είναι στο τέλος να υπάρχει μια ισορροπία μεταξύ αυτών και να έχουμε ένα μοντέλο ακριβές αλλά όχι τόσο ώστε να μην μπορεί να απεικονιστεί σ' αυτό και μια καινούρια συμπεριφορά η οποία ανήκει στην ίδια διαδικασία.

## **2.7 Ερωτήματα και προκλήσεις που προκύπτουν κατά την εξόρυξη διαδικασιών**

Η εξόρυξη διαδικασιών δημιουργεί μια σειρά από ενδιαφέροντα επιστημονικά ερωτήματα. Μερικά από αυτά έχουν απαντηθεί ενώ άλλα απαιτούν περαιτέρω έρευνα. Παρακάτω παρουσιάζονται συνοπτικά ορισμένα από αυτά (W.M.P. van der Aalst, 2004):

### **- Εξόρυξη κρυφών διεργασιών**

Μία από τις βασικές παραδοχές της εξόρυξης διαδικασιών είναι ότι κάθε συμβάν καταχωρείται στο αρχείο καταγραφής. Προφανώς, δεν είναι δυνατόν να εξαχθούν πληροφορίες σχετικά με εργασίες που δεν έχουν καταγραφεί. Ωστόσο, ορισμένες μορφές συμβολισμών είναι δυνατόν να περιέχουν «κρυφές διεργασίες», πράγμα που κάνει πιο δύσκολη αλλά όχι αδύνατη την αναπαράστασή τους στο μοντέλο διεργασιών.

- Εξόρυξη διπλών διεργασιών

Το πρόβλημα των διπλών διεργασιών αναφέρεται στην κατάσταση που μπορεί κανείς να έχει ένα μοντέλο διεργασιών (π.χ. ένα δίκτυο Petri) με δύο κόμβους που αναφέρονται στην ίδια διεργασία.

- Βρόχοι εξόρυξης (loops)

Σε μια διαδικασία μπορεί να είναι δυνατή η εκτέλεση της ίδιας διεργασίας πολλές φορές. Αυτό συνήθως αναπαρίσταται ως βρόχος στο αντίστοιχο μοντέλο. Πολλές τεχνικές εξόρυξης κάνουν κάποιες υποθέσεις σχετικά με τους βρόχους, οι οποίες περιορίζουν την κατηγορία των διαδικασιών που μπορούν να εξορυχτούν σωστά.

- Χρήση πληροφορίας χρόνου

Σε ένα αρχείο καταγραφής συμβάντων (event log) κάθε γραμμή αντιστοιχεί σε ένα συμβάν (δηλαδή, την εκτέλεση μιας εργασίας για μια συγκεκριμένη περίπτωση). Το αρχείο καταγραφής περιέχει επίσης πληροφορίες χρόνου, δηλαδή κάθε συμβάν έχει χρονική σήμανση. Οι πληροφορίες χρόνου μπορούν να χρησιμοποιηθούν για δύο σκοπούς: (1) προσθήκη της πληροφορίας χρονισμών στο μοντέλο της διαδικασίας και (2) βελτίωση της ποιότητας του ανακαλυφθέντος μοντέλου διεργασιών.

- Εξόρυξη με διαφορετικές προοπτικές

Η κυρίαρχη προοπτική της εξόρυξης διαδικασιών είναι η λεγόμενη προοπτική ελέγχου-ροής (control flow), όπου γίνεται διάταξη των διεργασιών και η οποία είναι δυνατό να επεκταθεί ώστε να συμπεριλάβει πληροφορίες χρονισμού (δηλαδή, τα συμβάντα να έχουν χρονικές σημάνσεις). Ωστόσο, εκτός από την προοπτική ελέγχου-ροής χρησιμοποιούνται κι άλλες προοπτικές, όπως πχ η προοπτική της οργάνωσης (organization perspective), στην οποία καθορίζεται η οργανωτική δομή και ο πληθυσμός και περιγράφονται οι σχέσεις μεταξύ των πόρων και άλλων αντικειμένων που αφορούν οργανωτικά ζητήματα (π.χ. ευθύνη και διαθεσιμότητα). Οι πόροι, δηλαδή οι άνθρωποι ή οι συσκευές, σχηματίζουν τον οργανωτικό πληθυσμό και κατανέμονται σε ρόλους και ομάδες. Επίσης χρησιμοποιείται η προοπτική πληροφοριών η οποία αφορά σε δεδομένα ελέγχου και παραγωγής. Τα δεδομένα ελέγχου είναι δεδομένα που εισάγονται αποκλειστικά για σκοπούς διαχείρισης διεργασιών, π.χ. μεταβλητές που εισάγονται για σκοπούς δρομολόγησης, ενώ τα δεδομένα παραγωγής είναι αντικείμενα πληροφοριών (π.χ. έγγραφα, φόρμες και πίνακες) των οποίων η ύπαρξη δεν εξαρτάται από τη διαχείριση διεργασιών. Τέλος, μπορεί κανείς να εξετάσει και την προοπτική εφαρμογής, η οποία ασχολείται με τις εφαρμογές που χρησιμοποιούνται για την εκτέλεση των διεργασιών (π.χ. τη χρήση ενός προγράμματος επεξεργασίας κειμένου). Σε ένα αρχείο καταγραφής συμβάντων μπορούμε να εντοπίσουμε ίχνη από όλες αυτές τις διαφορετικές προοπτικές.

- Αντιμετώπιση θορύβου

Οι περισσότεροι αλγόριθμοι εξόρυξης θεωρούν ότι οι πληροφορίες που εισάγονται είναι σωστές. Στις περισσότερες περιπτώσεις ωστόσο, το αρχείο καταγραφής ενδέχεται να περιέχει «θόρυβο», δηλαδή πληροφορίες που έχουν καταγραφεί εσφαλμένα. Για παράδειγμα, ορισμένες φορές ένα συμβάν δεν καταγράφεται ή καταγράφεται κάποια στιγμή μετά από την πραγματοποίησή του. Ο αλγόριθμος εξόρυξης πρέπει να είναι τέτοιος ώστε οι συσχετισμοί που εξάγει και αναπαριστά να μην βασίζονται σε μία μόνο παρατήρηση κι επομένως να μπορεί να διακρίνει τις εξαιρέσεις από την «κανονική ροή».

- Αντιμετώπιση της έλλειψης πληρότητας

Ένα αρχείο καταγραφής συμβάντων είναι ελλιπές εάν δεν περιέχει επαρκείς πληροφορίες οι οποίες να προκύπτουν από τη διαδικασία. Σε αυτήν την περίπτωση απαιτούνται ευρετικές μέθοδοι (heuristics) προκειμένου να αντιμετωπιστεί αυτό το πρόβλημα.

- Συγκέντρωση δεδομένων από ετερογενείς πηγές

Τα σύγχρονα πληροφοριακά συστήματα είναι ιδιαίτερα περίπλοκα και συνήθως αποτελούνται από μεγάλο αριθμό εφαρμογών / συστατικών στοιχείων. Σαν αποτέλεσμα οι πληροφορίες που απαιτούνται για την εξόρυξη διαδικασιών βρίσκονται διάσπαρτες στα πληροφοριακά συστήματα και τα συμβάντα ενδέχεται να καταγράφονται σε διάφορα επίπεδα ενός πληροφοριακού συστήματος. Ως εκ τούτου, η διαδικασία για τη συλλογή των αρχείων καταγραφής συμβάντων που χρησιμοποιούνται ως είσοδος για την εξόρυξη διαδικασιών, ορισμένες φορές καθίσταται κάθε άλλο παρά απλή.

- Οπτικοποίηση των αποτελεσμάτων

Μια άλλη πρόκληση είναι να παρουσιαστούν τα αποτελέσματα της εξόρυξης διαδικασιών με τέτοιο τρόπο ώστε οι άνθρωποι που εμπλέκονται να αποκτήσουν πραγματική εικόνα για τη διαδικασία και να έχουν έτσι τη δυνατότητα λήψης σωστών αποφάσεων για αλλαγές και βελτιώσεις. Οι σημαντικές πληροφορίες πρέπει να απεικονίζονται με τέτοιο τρόπο ώστε να είναι κατανοητές από όλους και η μη-σημαντικές ενδεχομένως να πρέπει να υποβιβαστούν προκειμένου η απεικόνιση που θα προκύψει να είναι όσο το δυνατόν απλούστερη.

- Ανάλυση Delta

Δεδομένου του γεγονότος ότι για μια διαδικασία μπορεί να προϋπάρχουν περιγραφικά ή κανονιστικά μοντέλα που έχουν δημιουργηθεί από ανθρώπους, είναι ενδιαφέρον να συγκρίνουμε αυτά τα μοντέλα με τα μοντέλα που προκύπτουν από τους αλγορίθμους και τις τεχνικές της εξόρυξης διαδικασιών. Η ανάλυση Delta χρησιμοποιείται για τη σύγκριση των δύο μοντέλων και την επεξήγηση των διαφορών.

Από τα παραπάνω ερωτήματα/προκλήσεις, η παρούσα διπλωματική αφορά σε μεγάλο βαθμό τη συγκέντρωση των δεδομένων από ετερογενείς πηγές, καθώς οι αιτήσεις-δηλώσεις για τις πανελλαδικές εξετάσεις καταχωρούνται σε όλες τις σχολικές μονάδες της χώρας. Αυτό σημαίνει πιθανή ανομοιογένεια στον τρόπο που αντιμετωπίζονται και καταγράφονται οι αιτήσεις στο πληροφοριακό σύστημα. Για το λόγο αυτό, η εξόρυξη της διαδικασίας των πανελλαδικών έρχεται αντιμέτωπη και με την αντιμετώπιση θορύβου, καθώς υπάρχουν πληροφορίες που έχουν καταγραφεί εσφαλμένα και οι αλγόριθμοι εξόρυξης πρέπει να είναι σε θέση να τις διακρίνουν από την κανονική ροή. Τέλος, στην παρούσα διαδικασία εξόρυξης πραγματοποιείται και εξόρυξη κρυφών διεργασιών, καθώς στο πληροφοριακό σύστημα καταγράφονται και διαδικασίες που κανονικά δεν θα έπρεπε να υπάρχουν, όπως π.χ. ημιτελείς αιτήσεις-δηλώσεις, διαφορετική σειρά αλληλουχίας συμβάντων, και οι οποίες εμφανίζονται μεν προκειμένου να εντοπιστούν οι λάθος χειρισμοί αλλά δεν είναι τέτοιες που να επηρεάζουν την ακριβή αναπαράσταση των μοντέλων που προκύπτουν.

## **2.8 Εργαλεία για την Εξόρυξη Διαδικασιών**

Μια πληθώρα εργαλείων, τεχνικών και τεχνολογιών εξόρυξης διεργασιών έχει αναπτυχθεί και χρησιμοποιηθεί επιτυχώς σε διάφορους τομείς εφαρμογών. Τα δύο πιο συνηθισμένα εργαλεία που χρησιμοποιούνται ευρέως είναι το ProM (Process Mining framework) και το Disco (Janes A., 2017).

- Το ProM (Process Mining framework) είναι ένα εργαλείο ανοιχτού κώδικα που περιλαμβάνει αλγορίθμους εξόρυξης διαδικασιών και βασίζεται σε JAVA. Παρέχει μια ολοκληρωμένη πλατφόρμα η οποία ενσωματώνει διάφορες προσθήκες (plug-ins) και την οποία οι χρήστες και οι προγραμματιστές μπορούν να εκμεταλλευτούν για να εκτελέσουν και να αναπτύξουν τεχνικές που αφορούν την εξόρυξη διαδικασιών. Η αρχιτεκτονική αυτή φιλοξενεί επί του παρόντος έναν πολύ μεγάλο αριθμό προσθηκών που καλύπτουν όλες τις διαφορετικές πτυχές της εξόρυξης διεργασιών, από την εισαγωγή δεδομένων έως την ανακάλυψη, τον έλεγχο συμμόρφωσης και τη βελτίωση κάνοντας χρήση διαφορετικών προοπτικών (van der Aalst W., 2016).

- Το Disco είναι ένα εμπορικό, αυτόνομο και ελαφρύ εργαλείο εξόρυξης διαδικασιών. Υποστηρίζει διάφορες μορφές αρχείων ως εισαγωγή, και ιδιαίτερα παρέχει εγγενή υποστήριξη για την εισαγωγή αρχείων CSV, τα οποία μπορούν να επισημαίνονται με πληροφορίες που αφορούν το συμβάν πριν από την εισαγωγή. Το Disco έχει μεγάλη χρηστικότητα, πιστότητα και αποδοτικότητα και καθιστά την εξόρυξη διεργασιών εύκολη και γρήγορη (Günther C. W., 2012).

Ωστόσο, το εργαλείο ανοιχτού κώδικα ProM υπήρξε το de-facto πρότυπο για την εξόρυξη διαδικασιών κατά την τελευταία δεκαετία. Η ανακάλυψη διεργασιών, ο έλεγχος συμμόρφωσης, η ανάλυση κοινωνικών δικτύων, η οργανωτική εξόρυξη, η

εξόρυξη αποφάσεων, η πρόβλεψη και η σύσταση βάσει ιστορικού κ.λπ. υποστηρίζονται από το ProM. Για παράδειγμα, υπάρχουν δεκάδες διαφορετικοί αλγόριθμοι ανακάλυψης διαδικασιών που υποστηρίζονται από το ProM, σε αντίθεση με άλλα εμπορικά εργαλεία που έχουν πιο περιορισμένη εργαλειοθήκη. Αυτό έχει ως αποτέλεσμα να καθιστά το ProM ένα εργαλείο με τεράστια λειτουργικότητα και να μην υπάρχει άλλο προϊόν που να προσφέρει ένα συγκρίσιμο σύνολο αλγορίθμων εξόρυξης διεργασιών (van der Aalst W., 2012).

Το ProM, για την εισαγωγή των αρχείων καταγραφής συμβάντων, χρησιμοποιεί το πρότυπο XES. Τα τελευταία χρόνια το πρότυπο XES (eXtensible Event Stream) εμφανίστηκε ως η κύρια μορφή αναφοράς για την αποθήκευση, την ανταλλαγή και την ανάλυση των αρχείων καταγραφής συμβάντων. Το XES, το οποίο βασίζεται σε XML, εμφανίστηκε για πρώτη φορά το 2009 ως διάδοχος του προτύπου MXML (Janes A., 2017). Τα αρχεία καταγραφής συμβάντων, όπως έχει ήδη αναφερθεί, μπορούν να πάρουν πληθώρα διαφορετικών μορφών και παραστάσεων. Κάθε αρχιτεκτονική συστήματος που περιλαμβάνει κάποιο είδος μηχανισμού καταγραφής έχει αναπτύξει τη δική της λύση για αυτό το σκοπό. Το XES είναι ένα πρότυπο για αρχεία καταγραφής συμβάντων σκοπός του οποίου είναι να παρέχει μια γενικά αναγνωρισμένη μορφή για την ανταλλαγή δεδομένων καταγραφής συμβάντων μεταξύ των εργαλείων και των διάφορων τομέων εφαρμογών και χρησιμοποιείται από πολλά λογισμικά. Ο πρωταρχικός του σκοπός ήταν η εξόρυξη διαδικασιών, δηλαδή η ανάλυση επιχειρησιακών διαδικασιών με βάση τα αρχεία καταγραφής συμβάντων. Ωστόσο, το πρότυπο XES έχει σχεδιαστεί ώστε να είναι επίσης κατάλληλο γενικότερα για εξόρυξη δεδομένων, για εξόρυξη κειμένου και για στατιστική ανάλυση (Günther C.W., 2014). Το πρότυπο XES έγινε το επίσημο πρότυπο της IEEE για την εξόρυξη διαδικασιών το 2016.

Στην παρούσα εργασία θα χρησιμοποιηθούν και τα δύο εργαλεία, κυρίως το εργαλείο ProM αλλά και το εργαλείο Disco – Academic License, προκειμένου να πραγματοποιηθεί εξόρυξη εκπαιδευτικών διαδικασιών που σχετίζονται με το θεσμό και την εφαρμογή στην πράξη της οργάνωσης των Πανελλαδικών εξετάσεων.





### **3. ΕΞΟΡΥΞΗ ΕΚΠΑΙΔΕΥΤΙΚΩΝ ΔΙΑΔΙΚΑΣΙΩΝ (EDUCATIONAL PROCESS MINING)**

Η εξόρυξη εκπαιδευτικών διαδικασιών (Educational Process Mining, EPM) είναι ένας αναδυόμενος επιμέρους τομέας της εξόρυξης εκπαιδευτικών δεδομένων (Educational Data Mining, EDM) που στοχεύει να κάνει σαφή γνώση την μη εκφρασμένη γνώση και να διευκολύνει την καλύτερη κατανόηση της εκπαιδευτικής διαδικασίας. Η EPM χρησιμοποιεί αρχεία καταγραφής συμβάντων που συλλέγονται ειδικά από εκπαιδευτικά περιβάλλοντα προκειμένου να ανακαλύψει, να αναλύσει και να παρέχει μια οπτική αναπαράσταση της πλήρους εκπαιδευτικής διαδικασίας (Bogarín A., 2017).

Η εξόρυξη εκπαιδευτικών διαδικασιών, σε αντιστοιχία με τις αρχές του process mining, στοχεύει (i) στην κατασκευή ολοκληρωμένων και συμπαγών μοντέλων εκπαιδευτικών διεργασιών που είναι σε θέση να αναπαράγουν όλες τις παρατηρούμενες συμπεριφορές εκπαιδευτών και κυρίων εκπαιδευόμενων (ανακάλυψη μοντέλου διεργασίας), (ii) στον έλεγχο κατά πόσο η μοντελοποιημένη συμπεριφορά (είτε είναι προκαθορισμένη είτε έχει ανακαλυφθεί από εκπαιδευτικά δεδομένα) ταιριάζει με την παρατηρούμενη συμπεριφορά (έλεγχος συμμόρφωσης) και (iii) στην προβολή πληροφοριών που εξάγονται από τα αρχεία καταγραφής στο μοντέλο, προκειμένου να αποσαφηνίσει τη γνώση και να διευκολύνει την καλύτερη κατανόηση της διαδικασίας από τους ειδικούς που ασχολούνται με τη βελτίωση των διαδικασιών στον τομέα της εκπαίδευσης (επέκταση μοντέλου διεργασίας) (Trcka N., 2009).

#### **3.1 Η εξόρυξη διαδικασιών στον τομέα της Εκπαίδευσης**

Σήμερα, με την ανάπτυξη και την αυξανόμενη δημοτικότητα των υποστηριζόμενων από την τεχνολογία περιβαλλόντων μάθησης, τα πληροφοριακά συστήματα μας επιτρέπουν να καταγράψουμε όλες τις εκδηλώσεις, τις δράσεις και τις δραστηριότητες των εκπαιδευόμενων. Η εξόρυξη διαδικασιών μπορεί να χρησιμοποιήσει αυτά τα αρχεία καταγραφής και τα ίχνη προκειμένου να ανακαλύψει, να παρακολουθήσει και να βελτιώσει τις εκπαιδευτικές διαδικασίες.

Όπως έχει ήδη αναφερθεί, η εξόρυξη διαδικασιών αποτελεί μια γέφυρα μεταξύ της εξόρυξης δεδομένων (Data Mining) και της μοντελοποίησης και ανάλυσης των διαδικασιών. Η εξόρυξη δεδομένων εφαρμόστηκε ευρέως με επιτυχία για να βρει ενδιαφέροντα μοτίβα από δεδομένα που συγκεντρώθηκαν σε εκπαιδευτικά περιβάλλοντα. Ωστόσο, η εξόρυξη εκπαιδευτικών δεδομένων (EDM) επικεντρώνεται σε εξαρτήσεις δεδομένων ή απλά πρότυπα και δεν παρέχει οπτική αναπαράσταση της συνολικής μαθησιακής διαδικασίας, δηλαδή δεν επικεντρώνεται στη διαδικασία στο σύνολό της. Προκειμένου λοιπόν να πραγματοποιηθεί μια ανάλυση όπου η διαδικασία παίζει τον κεντρικό ρόλο, εφαρμόστηκε η εξόρυξη εκπαιδευτικών διαδικασιών (EPM), η

οποία είναι η εφαρμογή τεχνικών εξόρυξης διαδικασιών σε ακατέργαστα εκπαιδευτικά δεδομένα.

Η EPM ενδιαφέρεται για τις διεργασίες και τη διαδικασία από την αρχή ως το τέλος (end-to-end process) κι όχι για τοπικά μοτίβα. Μπορεί να δημιουργήσει ολοκληρωμένα, συμπαγή μοντέλα εκπαιδευτικών διαδικασιών που είναι σε θέση να αναπαραγάγουν όλες τις παρατηρούμενες εκπαιδευτικές συμπεριφορές, να ελέγξουν αν η συμπεριφορά μοντελοποίησης ταιριάζει με τη συμπεριφορά που παρατηρείται και να εξαγάγει πληροφορίες από τις εγγραφές στο μοτίβο για να δημιουργήσει σαφή γνώση και για να διευκολύνει μια καλύτερη κατανόηση της όλης εκπαιδευτικής διαδικασίας (Bogarín A., 2017).

Όπως έχει αναφερθεί, τα εκπαιδευτικά συστήματα υποστηρίζουν μεγάλο όγκο δεδομένων, τα οποία προέρχονται από πολλαπλές πηγές και τα οποία αποθηκεύονται σε διάφορες μορφές. Τα δεδομένα προέρχονται από «πρόσωπο με πρόσωπο» εκπαιδευτικά συστήματα, όπως π.χ. οι παραδοσιακές αίθουσες διδασκαλίας, από εξ αποστάσεως εκπαίδευση που λαμβάνει χώρα σε διαδραστικά μαθησιακά περιβάλλοντα, από συνεργατική μάθηση με υπολογιστή κ.ο.κ. Για παράδειγμα, τα πρόσωπο με πρόσωπο εκπαιδευτικά συστήματα αποθηκεύουν μόνο διοικητικές και δημογραφικές πληροφορίες, δηλαδή τα προφίλ των εκπαιδευόμενων (π.χ. βαθμοί και στόχοι προγράμματος σπουδών), ποιος ακολουθεί ποιο πρόγραμμα, ποια μαθήματα διδάσκεται και με τι βαθμό αξιολογείται στις εξετάσεις. Τα ηλεκτρονικά εκπαιδευτικά συστήματα αποθηκεύουν πιο ακριβή δεδομένα, επειδή μπορούν να καταγράψουν όλες τις πληροφορίες σχετικά με τις ενέργειες και τις αλληλεπιδράσεις των μαθητών σε αρχεία καταγραφής συμβάντων και βάσεις δεδομένων. Αυτά τα δεδομένα περιλαμβάνουν αρχεία καταγραφής χρήσης πόρων (π.χ. φυλλάδια, εγγραφές βίντεο), δεδομένα αξιολόγησης και πολλές πληροφορίες για τις χρονικές στιγμές στις οποίες λαμβάνουν χώρα συγκεκριμένα γεγονότα. Είναι γεγονός πως η πρόκληση για επίτευξη της μέγιστης αποτελεσματικότητας με το μικρότερο δυνατό κόστος στα σύγχρονα εκπαιδευτικά συστήματα οδηγεί στην καταγραφή ολόενα και περισσότερων πληροφοριών σχετικά με α) τη βραχυπρόθεσμη ικανοποίηση των εκπαιδευόμενων από τα προγράμματα, τα μαθήματα και τους πόρους και β) τη μακροπρόθεσμη αίσθηση χρησιμότητας που λαμβάνουν από τα μαθήματα που έχουν παρακολουθήσει. Οι καταγεγραμμένες πληροφορίες στα εκπαιδευτικά συστήματα είναι δομημένες (αρχεία καταγραφής, πληροφορίες εγγραφής μαθητών, προφίλ χρήσης μαθητών, διοικητικές πληροφορίες κ.λπ.) ή μη δομημένες (αλληλεπίδραση με καθηγητές μέσω συνομιλίας, συνεργασία με άλλους μαθητές μέσω συνομιλίας κλπ.). Για να ανακαλυφθεί ένα σωστό μοντέλο εκπαιδευτικής διαδικασίας θεωρείται ότι τα αρχεία καταγραφής συμβάντων περιέχουν αντιπροσωπευτικό δείγμα συμπεριφοράς των εκπαιδευόμενων. Ωστόσο, η εφαρμογή τεχνικών ανακάλυψης διαδικασιών στην εκπαίδευση παρουσιάζει ορισμένες προκλήσεις, δεδομένου του τεράστιου όγκου και της ετερογένειας των ιχνών που συναντώνται συχνά σε εκπαιδευτικά σύνολα δεδομένων (Cairns A.H., 2015):

### 1) Μεγάλος όγκος δεδομένων - Μεγάλος αριθμός περιπτώσεων ή γεγονότων σε αρχεία καταγραφής συμβάντων

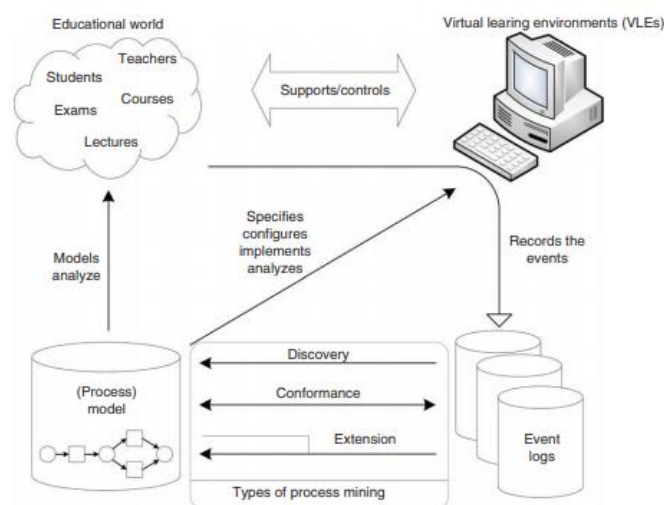
Τα αρχεία καταγραφής συμβάντων στον τομέα της εκπαίδευσης, ιδιαίτερα εκείνα που προέρχονται από περιβάλλοντα ηλεκτρονικής μάθησης, ενδέχεται να περιέχουν μεγάλο όγκο καταγραφών λεπτομερών γεγονότων και δεδομένων που σχετίζονται με την εκάστοτε εκπαιδευτική διαδικασία και οι περισσότερες από τις σύγχρονες τεχνικές / εργαλεία εξόρυξης διαδικασιών δεν είναι σε θέση να χειριστούν τεράστια αρχεία καταγραφής συμβάντων.

### 2) Ετερογένεια και πολυπλοκότητα: Μεγάλος αριθμός διακριτών ίχνων και δραστηριοτήτων σε αρχεία καταγραφής συμβάντων

Στην πράξη, οι εκπαιδευτικές διαδικασίες δεν είναι δομημένες και είναι ιδιαίτερα ευέλικτες, με αποτέλεσμα να υπάρχουν πολλά ετερογενή και ξεχωριστά ίχνη στα αρχεία καταγραφής, που αντικατοπτρίζουν την μεγάλη ποικιλία συμπεριφορών στις μαθησιακές διαδρομές των εκπαιδευόμενων. Κατά συνέπεια, οι υπάρχουσες τεχνικές εξόρυξης διαδικασιών δημιουργούν πολύπλοκα μοντέλα (spaghetti models) που συχνά είναι πολύ συγκεχυμένα και δυσνόητα. Επομένως πρέπει να υιοθετείται ένας συνδυασμός τεχνικών απλούστευσης και φιλτραρίσματος που μειώνουν την πολυπλοκότητα των αρχείων καταγραφής συμβάντων χωρίς όμως να χάνονται πληροφορίες που μας επιτρέπουν να ανακαλύπτουμε βασικές έννοιες και μοτίβα από αυτά τα αρχεία καταγραφής.

## 3.2 Εννοιολογικό πλαίσιο της εξόρυξης εκπαιδευτικών διαδικασιών

Μια επισκόπηση της εφαρμογής της εξόρυξης διαδικασιών στον εκπαιδευτικό τομέα φαίνεται στην παρακάτω εικόνα. Στην ουσία, αυτό το πλαίσιο αναφοράς της εκπαιδευτικής εξόρυξης διαδικασιών είναι μια προσαρμογή του γενικού πλαισίου της εξόρυξης διαδικασιών που απεικονίζεται στην Εικόνα 3.



Εικόνα 6: Το πλαίσιο αναφοράς για την εξόρυξη εκπαιδευτικών διαδικασιών (Bogarín Alejandro & Cerezo Rebeca & Romero Cristóbal, A survey on educational process mining, 2017.)

Τα κύρια στοιχεία του παραπάνω πλαισίου που το διαφοροποιούν από το γενικό πλαίσιο της εξόρυξης διαδικασιών είναι τα εξής:

- Εκπαιδευτικός κόσμος (Educational World): Βασικά δύο είδη συμμετεχόντων διαδραματίζουν σημαντικό ρόλο σε οποιαδήποτε εκπαιδευτική δραστηριότητα, μαθητές και εκπαιδευτικοί, όπου οι εκπαιδευτικοί παρέχουν κατάλληλους πόρους για να διασφαλίσουν την επιτυχία των μαθητών και οι μαθητές είναι το ουσιαστικό μέρος κάθε δραστηριότητας μάθησης, αλληλεπιδρώντας με άλλους συμμετέχοντες (μαθητές ή εκπαιδευτικούς) και με το ίδιο το σύστημα. Τέλος, υπάρχουν τα μαθήματα, οι διαλέξεις, οι εξετάσεις κ.λπ. που χρησιμοποιούνται ως πόροι για τους συμμετέχοντες.
- Εικονικό περιβάλλον μάθησης (Virtual Learning Environment): Αφορά κυρίως στις διαδικασίες ηλεκτρονικής μάθησης (e-learning) και παρέχει τις βασικές δομές και τους πόρους όπου πραγματοποιούνται οι μαθησιακές δράσεις και οι αλληλεπιδράσεις των συμμετεχόντων. Καταγράφει επίσης τα γεγονότα που συμβαίνουν κατά τη διάρκεια της διαδικασίας μάθησης. Τα περισσότερα παρέχουν στους καθηγητές ή τους ερευνητές βασικά εργαλεία για την ανάλυση της μάθησης των εκπαιδευόμενων (εξέλιξη βαθμολογιών, αριθμός δραστηριοτήτων που έχουν γίνει, συμμετοχή στο φόρουμ, τελευταία σύνδεση κ.λπ.), αλλά δεν παρέχουν συγκεκριμένα εργαλεία που θα επέτρεπαν στους εκπαιδευτικούς να αξιολογήσουν διεξοδικά τη συνολική μαθησιακή διαδικασία.
- Αρχεία καταγραφής συμβάντων (Event Logs): Αυτά είναι αρχεία που καταγράφουν τα γεγονότα που συμβαίνουν στα εικονικά περιβάλλοντα μάθησης και τα οποία συνήθως αποθηκεύονται σε βάσεις δεδομένων. Περιέχουν μεγάλο αριθμό ανεπεξέργαστων δεδομένων και πρέπει συνήθως να μετατραπούν σε μια συγκεκριμένη μορφή αρχείου προκειμένου να χρησιμοποιηθούν από τα εργαλεία εξόρυξης διαδικασιών.
- Μοντέλα διεργασιών (Process Models): Αποκαλύπτουν πολύτιμες πληροφορίες σχετικά με το πώς οι συμμετέχοντες του εκπαιδευτικού κόσμου αλληλεπιδρούν με το σύστημα, ξεκινώντας από τα αρχεία καταγραφής συμβάντων. Λαμβάνονται χρησιμοποιώντας διαφορετικές τεχνικές εξόρυξης εκπαιδευτικών διαδικασιών, όπως ανακάλυψη (discovery), όπου ο εκπαιδευτής μπορεί να απεικονίσει το μοντέλο συμπεριφοράς των μαθησιακών διαδρομών των εκπαιδευόμενων παρέχοντας γνώση της διαδικασίας αντί μόνο του μαθησιακού αποτελέσματος, έλεγχο συμμόρφωσης (conformance checking), όπου ο εκπαιδευτής μπορεί να αναλύσει εάν το παραγόμενο μοντέλο αντιστοιχεί στο μοντέλο συμπεριφοράς που βρίσκεται στα αρχεία καταγραφής συμβάντων και επέκταση (enhancement ή extension), όπου π.χ. ο εκπαιδευτής μπορεί να εξάγει γνώση για σημεία συμφόρησης ή για σχέσεις μεταξύ των εκπαιδευόμενων σε ένα μάθημα (Bogarín A., 2017).

### 3.3 Τομείς εφαρμογών της εξόρυξης εκπαιδευτικών διαδικασιών

Η εξόρυξη στις εκπαιδευτικές διαδικασίες έχει χρησιμοποιηθεί σε ένα ευρύ φάσμα εφαρμογών στον τομέα της εκπαίδευσης προκειμένου να αντιμετωπίσει ποικίλα εκπαιδευτικά προβλήματα (Bogarín A., 2017):

#### Περιβάλλοντα MOOC και LMS

Τα μαζικά ανοιχτά διαδικτυακά μαθήματα (Massive Open Online Courses - MOOCs), τα συστήματα διαχείρισης εκμάθησης (Learning Management Systems - LMS) και άλλα παρόμοια διαδικτυακά μαθησιακά περιβάλλοντα παρέχουν δωρεάν ευκαιρίες μάθησης στην διαδικτυακή κοινότητα. Τα αρχεία καταγραφής που δημιουργούνται από αυτά τα συστήματα παρέχουν μια εικόνα για το πώς οι άνθρωποι παρακολουθούν το μάθημα όταν παρακολουθούν βίντεο ή διαλέξεις και όταν παραδίδουν εργασίες, μεταξύ άλλων. Η περισσότερη έρευνα σχετικά με την εφαρμογή του Process Mining έχει γίνει σε αυτούς τους τύπους μαθησιακού περιβάλλοντος. Οι Trcka et al. (Trcka N., 2011) επεσήμαναν την εφαρμογή του Process Mining σε εκπαιδευτικά δεδομένα και συζήτησαν ορισμένες από τις δυνατότητές του για την εξαγωγή γνώσης από περιβάλλοντα LMS, λαμβάνοντας υπόψη μόνο τα ίχνη εξέτασης των μαθητών. Επίσης, οι Bogarín et al. (Bogarín A., 2014) χρησιμοποίησαν δεδομένα από αρχεία καταγραφής σε περιβάλλον Moodle και πρότειναν τη χρήση ομαδοποίησης (clustering) προκειμένου να αποκτήσουν πιο συγκεκριμένα και ακριβή μοντέλα διεργασιών για τη συμπεριφορά των μαθητών. Σε άλλες έρευνες, οι Mukala et al. (Mukala P., 2015) χρησιμοποίησαν τεχνικές εξόρυξης διαδικασιών προκειμένου να εντοπίσουν και να αναλύσουν τις μαθησιακές συνήθειες των μαθητών με βάση δεδομένα σε περιβάλλον MOOC. Τα αποτελέσματα έδειξαν ότι οι επιτυχημένοι μαθητές ακολουθούν ένα δομημένο μοτίβο παρακολούθησης, ενώ οι αποτυχημένοι μαθητές είναι πιο απρόβλεπτοι και παρουσιάζουν μη-δομημένη συμπεριφορά.

#### Συνεργατική μάθηση υποστηριζόμενη από υπολογιστή (Computer-Supported Collaborative Learning - CSCL)

Η συνεργατική μάθηση που υποστηρίζεται από υπολογιστή (CSCL) χαρακτηρίζεται από την ανταλλαγή και την κατασκευή γνώσης μεταξύ των συμμετεχόντων χρησιμοποιώντας την τεχνολογία ως πρωταρχικό μέσο επικοινωνίας τους ή ως κοινό πόρο. Στον τομέα αυτό ωστόσο οι τεχνικές εξόρυξης διαδικασιών επιδέχονται περαιτέρω μελέτη.

Η εξόρυξη διαδικασιών έχει εφαρμοστεί στην CSCL προκειμένου να παρέχει ανατροφοδότηση στους μαθητές σχετικά με τις διαδικασίες λήψης αποφάσεων (Reimann P., 2009). Ο στόχος ήταν να χρησιμοποιηθούν τεχνικές process mining για τον προσδιορισμό μοντέλων ομάδων λήψης αποφάσεων που πραγματοποιήθηκαν σε μια αίθουσα συνομιλίας (chat room). Το συμπέρασμα της συγκεκριμένης μελέτης ήταν ότι τα παραγόμενα μοντέλα ήταν μη δομημένα, περίπλοκα και κυκλικά. Η διαδικασία

λήψης αποφάσεων έπαιρνε διαφορετική διαδρομή κάθε φορά που εκτελούνταν κι επιπλέον παρουσιάζονταν συχνά βρόχοι που οδηγούσαν σε προηγούμενα γεγονότα.

Μια άλλη εφαρμογή της εξόρυξης εκπαιδευτικών διαδικασιών σ' αυτόν τον τομέα είναι η Συνεργατική Συγγραφή (Collaborative Writing - CW), η οποία χρησιμοποιείται ευρέως σε εκπαιδευτικά περιβάλλοντα όπου οι μαθητές χρησιμοποιούν συχνά υπολογιστές προκειμένου να σημειώνουν κατά τη διάρκεια διαλέξεων και να γράφουν δοκίμια για τις εργασίες τους. Χάρη στο διαδίκτυο, οι μαθητές μπορούν επίσης να γράφουν συνεργατικά κοινοποιώντας τα έγγραφά τους με διάφορους τρόπους. Τεχνικές εξόρυξης διαδικασιών χρησιμοποιήθηκαν από τους Southavilay et al. (Southavilay V., 2010) για την ανάλυση των διαδικασιών γραφής των μαθητών και του τρόπου με τον οποίο αυτές οι διαδικασίες σχετίζονται με την ποιότητα και τα σημασιολογικά χαρακτηριστικά του τελικού προϊόντος, και για το σκοπό αυτό χρησιμοποίησαν έγγραφα που συλλέχθηκαν από διαφορετικές ομάδες προπτυχιακών φοιτητών που έγραφαν συνεργατικά. Το συγκεκριμένο έργο θα μελετηθεί επιπλέον στο μέλλον, καθώς τα αποτελέσματα ήταν συγκεχυμένα και δεν ήταν δυνατό να υπάρξει διάκριση μεταξύ καλύτερων και ασθενέστερων φοιτητών.

### Επαγγελματική εκπαίδευση

Τα κέντρα εκπαίδευσης και κατάρτισης έχουν καταστήσει τα μαθήματα επαγγελματικής κατάρτισης πιο ευέλικτα προκειμένου να ανταποκριθούν στις μεταβαλλόμενες ανάγκες της αγοράς εργασίας και να ικανοποιήσουν τις απαιτήσεις χρόνου σε συνάρτηση με τις δεξιότητες που αποκτώνται. Η εξόρυξη διαδικασιών έχει χρησιμοποιηθεί σε διαφορετικούς τύπους μαθημάτων επαγγελματικής κατάρτισης.

Οι Cairns et al. (Cairns AH, 2014) έδειξαν πώς το process mining μπορεί να χρησιμοποιηθεί για την παρακολούθηση και τη βελτίωση των εκπαιδευτικών διαδικασιών στον τομέα της επαγγελματικής κατάρτισης. Η έρευνά τους στόχευε στην ανάπτυξη μεθόδων που θα μπορούσαν να εφαρμοστούν σε θέματα γενικής εκπαίδευσης και πιο συγκεκριμένες εφαρμογές σχετικά με την επαγγελματική κατάρτιση ή τους τομείς της ηλεκτρονικής μάθησης για την εξαγωγή, ανάλυση, ενίσχυση και εξατομίκευση των εκπαιδευτικών διαδικασιών. Συγκεκριμένα έγινε χρήση τεχνικών κοινωνικής εξόρυξης για να εξεταστούν οι αλληλεπιδράσεις μεταξύ παρόχων κατάρτισης και εκπαιδευτικών μαθημάτων, που εμπλέκονται στις διαδρομές κατάρτισης των μαθητών.

### Εξόρυξη προγράμματος σπουδών

Ένα πρόγραμμα σπουδών σχεδιάζεται από ένα εκπαιδευτικό ίδρυμα προκειμένου να επιτύχει συγκεκριμένους στόχους. Τα προγράμματα σπουδών συνήθως υποδηλώνουν ότι οι μαθητές μπορούν να ακολουθήσουν διαφορετικές διαδρομές από την αρχή έως το τέλος λόγω μιας πιο φιλελεύθερης προσέγγισης που υιοθετείται στην επιλογή μαθημάτων

Μια προσέγγιση εκπαιδευτικής εξόρυξης διαδικασιών προτάθηκε από τους Trcka και Rechenizkiy (Trcka N., 2009) σχετικά με την εξόρυξη προγραμμάτων σπουδών. Πρότειναν ένα πλαίσιο που προϋποθέτει ότι ένα σύνολο προτύπων και κανόνων μπορεί να προκαθοριστεί για να εστιάσει στην εξόρυξη με επιθυμητό τρόπο και να την κάνει πιο αποτελεσματική και αποδοτική. Το πλαίσιο αυτό αποσκοπεί στο να βοηθήσει τους εκπαιδευτικούς να αναλύσουν τις εκπαιδευτικές διαδικασίες σχετικά με τα υφιστάμενα προγράμματα σπουδών. Σε άλλες σχετικές έρευνες, οι Wang και Zaïane (Wang R., 2015) ανακάλυψαν ένα μοντέλο μαθησιακής διαδικασίας φοιτητών που παρακολουθούν μαθήματα και συνέκριναν τα μονοπάτια που τείνουν να ακολουθούν επιτυχημένοι και λιγότερο επιτυχημένοι φοιτητές, επισημαίνοντας τις διαφορές μεταξύ τους. Το συμπέρασμα στο οποίο κατέληξαν είναι ότι η εξόρυξη διαδικασιών έχει πράγματι μεγάλες δυνατότητες να βοηθήσει τους εκπαιδευτικούς και τους διαχειριστές να κατανοήσουν τη συμπεριφορά των φοιτητών, να τους προτείνουν τη σωστή πορεία και εντέλει να βελτιώσουν το σχεδιασμό ενός προγράμματος σπουδών.

#### Αξιολόγηση βάσει υπολογιστή

Η αξιολόγηση με βάση τον υπολογιστή (Computer-Based Assessment - CBA) είναι, στην ουσία, η πρακτική του να εξετάζεται κανείς σε κουίζ και tests μπροστά σε ηλεκτρονικό υπολογιστή αντί να χρησιμοποιεί παραδοσιακές μορφές μολυβιού και χαρτιού. Η αξιολόγηση βάσει υπολογιστή χρησιμοποιείται ευρέως σε πολλά περιβάλλοντα ηλεκτρονικής μάθησης.

Οι Tóth et al. (Tóth K., 2017) περιέγραψαν τον τρόπο εξαγωγής πληροφοριών που σχετίζονται με τη συνολική διαδικασία της αξιολόγησης από αρχεία καταγραφής συμβάντων και το πώς να χρησιμοποιηθούν αυτά τα δεδομένα σε αξιολογήσεις επίλυσης προβλημάτων και να περιγραφούν μέθοδοι που βοηθούν στην ανακάλυψη νέων πληροφοριών με βάση την παρατηρούμενη συμπεριφορά στην επίλυση προβλημάτων. Με τη μελέτη αυτή εξετάστηκαν οι αλληλεπιδράσεις ανθρώπου-υπολογιστή μέσω ιχνών που παραμένουν στο αρχείο καταγραφής και ανακαλύφθηκαν σχέσεις μεταξύ των δραστηριοτήτων των μαθητών και της απόδοσης στις εργασίες.

#### Εγγραφές φοιτητών

Η εγγραφή φοιτητών περιλαμβάνει όλες τις προϋποθέσεις και τα βήματα της διαδικασίας εγγραφής μαθημάτων. Είναι ζωτικής σημασίας να ελέγχονται οι διαδικασίες που παρέχονται από αυτοματοποιημένα συστήματα διαχείρισης στον εκπαιδευτικό τομέα, προκειμένου να παράγονται τα αναμενόμενα αποτελέσματα όσον αφορά την ποιότητα και την έγκαιρη διαδικασία εγγραφής των φοιτητών στα μαθήματα.

Οι Anuwatvisit et al. (Anuwatvisit S., 2012) χρησιμοποίησαν τον έλεγχο συμμόρφωσης προκειμένου να εντοπίσουν ασυμφωνίες μεταξύ των ροών που καθορίζονται σε ένα

μοντέλο εγγραφής φοιτητή σε μάθημα και των πραγματικών περιπτώσεων που αντλήθηκαν από τα αρχεία καταγραφής συμβάντων. Επιπλέον, επέκτειναν τα μοντέλα με χαρακτηριστικά απόδοσης και επιχειρηματικούς κανόνες.

### Τρισδιάστατοι Εκπαιδευτικοί Εικονικοί Κόσμοι (3D Educational Virtual Worlds)

Οι Τρισδιάστατοι Εκπαιδευτικοί Εικονικοί Κόσμοι είναι περιβάλλοντα που ενθαρρύνουν την αλληλεπίδραση μεταξύ μαθητών και εκπαιδευτικών. Αυτά τα περιβάλλοντα ενθαρρύνουν τους μαθητές (ως avatars) να εκτελούν μαθησιακές δραστηριότητες που δεν είχαν προγραμματιστεί αρχικά από τους εκπαιδευτικούς, όπως η αλληλεπίδραση με άλλους μαθητές μέσω συνομιλιών, η αναζήτηση νέων περιεχομένων πολυμέσων ή η χρήση τρισδιάστατων φυσικών αντικειμένων προκειμένου να μάθουν περισσότερα για ένα συγκεκριμένο θέμα. Η εξόρυξη διαδικασιών χρησιμοποιήθηκε για να ανακαλύψει τι συμβαίνει στις τρισδιάστατες μαθησιακές διαδικασίες από τους Fernández-Gallego et al. (Fernández-Gallego B., 2013), οι οποίοι παρουσίασαν ένα πλαίσιο μαθησιακών αναλυτικών στοιχείων για τρισδιάστατους εκπαιδευτικούς εικονικούς κόσμους που επικεντρώνεται στην ανακάλυψη των ροών μάθησης και στον έλεγχο της συμμόρφωσης μέσω τεχνικών εξόρυξης. Στόχος είναι οι εκπαιδευτικοί να γνωρίζουν τι συμβαίνει στη μαθησιακή διαδικασία επειδή στα τρισδιάστατα εκπαιδευτικά περιβάλλοντα μάθησης οι αλληλεπιδράσεις μεταξύ των μαθητών είναι συνεχείς και επομένως υπάρχει πολύς θόρυβος, δηλαδή μεγάλος αριθμός δραστηριοτήτων που δεν είναι σημαντικές από παιδαγωγική άποψη. Ωστόσο, για να επιτευχθούν καλύτερα αποτελέσματα, οι αλγόριθμοι ανακάλυψης διαδικασιών πρέπει να είναι πιο ισχυροί σε ότι αφορά στον θόρυβο από τους τρέχοντες αλγόριθμους της τελευταίας τεχνολογίας.

### **3.4 Μελλοντικές προκλήσεις στην εξόρυξη εκπαιδευτικών διαδικασιών – Εκπαιδευτική Διοίκηση**

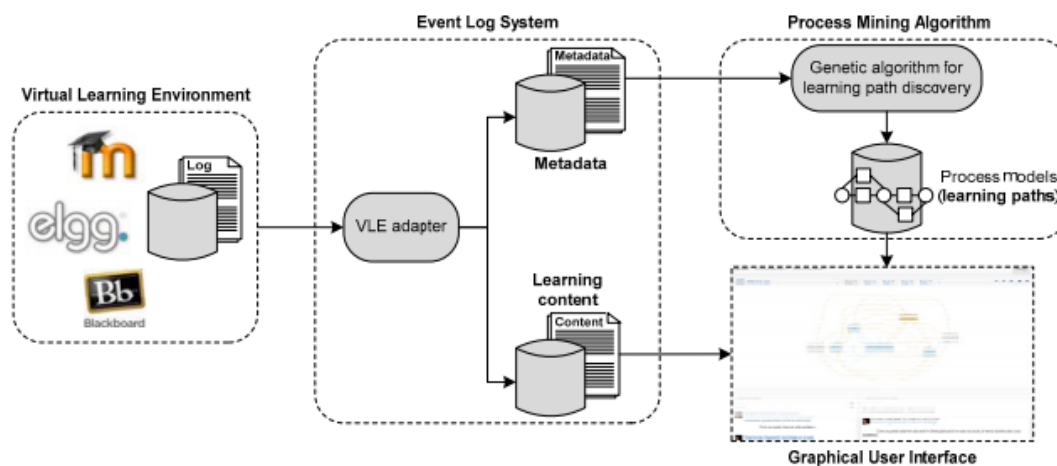
Η εξόρυξη εκπαιδευτικών διαδικασιών (EPM) αποτελεί ένα νέο, αναπτυσσόμενο πεδίο που σχετίζεται στενά με άλλα ερευνητικά πεδία όπως η Εξόρυξη Προθέσεων (Intention Mining – IM) και η Εξόρυξη Διαδοχικών Προτύπων (Sequential Pattern Mining - SPM). Τα αποτελέσματα της εξόρυξης εκπαιδευτικών διαδικασιών μπορούν να χρησιμοποιηθούν για την καλύτερη κατανόηση των υποκείμενων εκπαιδευτικών διαδικασιών, για την παροχή ανατροφοδότησης σε μαθητές, δασκάλους και ερευνητές, για τον εντοπισμό μαθησιακών δυσκολιών και για να βοηθήσουν τους μαθητές με συγκεκριμένες μαθησιακές δυσκολίες, για τη βελτίωση της διαχείρισης των εκπαιδευτικών αντικειμένων και για τη δημιουργία συμβουλών για τους μαθητές, μεταξύ πολλών άλλων.

Ωστόσο, η εξόρυξη εκπαιδευτικών διαδικασιών, ως αναδυόμενος τομέας, αντιμετωπίζει αρκετές προκλήσεις και έχει πολλές προοπτικές για το μέλλον, όπως (Bogarín A., 2017):



- **Ανάπτυξη πιο εξειδικευμένων εργαλείων** για την εξόρυξη εκπαιδευτικών διαδικασιών, προκειμένου να έρθει το process mining και οι τεχνικές του πιο κοντά στους ειδικούς του τομέα (δηλαδή στους εκπαιδευτικούς και στους ερευνητές, οι οποίοι δεν έχουν απαραίτητα όλο το τεχνικό υπόβαθρο), βοηθώντας τους έτσι να αναλύσουν τις εκπαιδευτικές διαδικασίες. Ένα εξειδικευμένο εργαλείο EPM είναι το SoftLearn, μια πλατφόρμα που χρησιμοποιεί έναν γενετικό αλγόριθμο για να ανακαλύψει ολοκληρωμένες διαδρομές μάθησης, διασφαλίζοντας ότι δεν υπάρχουν δραστηριότητες που λείπουν και επομένως επιτρέπει στους εκπαιδευτικούς να χρησιμοποιούν αυτήν την πλατφόρμα για την αξιολόγηση των μαθησιακών δραστηριοτήτων (Barreiros BV, 2014). Η πλατφόρμα SoftLearn έχει μια γραφική διεπαφή που έχει αναπτυχθεί ειδικά για να απεικονίσει τόσο τις μαθησιακές διαδρομές που ανακαλύφθηκαν από τον γενετικό αλγόριθμο όσο και τα δεδομένα που δημιουργούνται κατά τη διάρκεια των μαθησιακών δραστηριοτήτων.

Η αρχιτεκτονική της πλατφόρμας SoftLearn απεικονίζεται παρακάτω:



Εικόνα 7: Η αρχιτεκτονική της πλατφόρμας SoftLearn (Barreiros BV, Lama M, Mucientes M, Vidal JC. Softlearn: a process mining platform for the discovery of learning paths, 2014)

Το πρώτο συστατικό αυτής της αρχιτεκτονικής είναι το εικονικό περιβάλλον μάθησης (VLE) όπου οι μαθητές και οι εκπαιδευτικοί πραγματοποιούν τις δραστηριότητες μάθησης και υποστήριξης που προγραμματίζονται για ένα μάθημα, έχουν πρόσβαση στο περιεχόμενο της μάθησης και εκτελούν τις διαδικασίες που απαιτούνται για την ανάληψη αυτών των δραστηριοτήτων. Όταν ένας εκπαιδευόμενος εκτελεί μια μαθησιακή δραστηριότητα, όπως π.χ. η απάντηση σε ένα τεστ, το VLE αποθηκεύει σε μια βάση δεδομένων τις πληροφορίες που δημιουργούνται (Event Log System). Στη συνέχεια, οι πληροφορίες αυτές επεξεργάζονται από αλγορίθμους εξόρυξης διαδικασιών (Process Mining Algorithms - PMA), οι οποίοι είναι το βασικό συστατικό της πλατφόρμας SoftLearn και στόχος τους είναι να ανακαλύψουν τη ροή εργασιών που αντιπροσωπεύει τη μαθησιακή πορεία που ακολουθούν οι εκπαιδευόμενοι. Για την

επίτευξη αυτού του στόχου, οι PMA χρειάζεται μόνο να επεξεργαστούν τις πληροφορίες μεταδεδομένων που είναι διαθέσιμες στο αρχείο καταγραφής συμβάντων, οι οποίες είναι ανεξάρτητες από το είδος των μαθησιακών δραστηριοτήτων που αναλαμβάνουν οι εκπαιδευόμενοι. Τέλος, η γραφική διεπαφή χρήστη (Graphical User Interface) του SoftLearn είναι αυτή που επιτρέπει στους εκπαιδευτικούς να κατανοήσουν τη συμπεριφορά του μαθητή μέσω της οπτικοποίησης των μαθησιακών διαδρομών που ακολουθεί ο κάθε εκπαιδευόμενος κατά τη διάρκεια ενός μαθήματος.

- **Χρήση συστάσεων στην εξόρυξη εκπαιδευτικών διαδικασιών:** Οι πληροφορίες που ανακαλύπτονται κατά την εξόρυξη πρέπει να είναι όχι μόνο κατανοητές αλλά και χρήσιμες για τη λήψη αποφάσεων των τελικών χρηστών. Για παράδειγμα, αντί να δείξουμε το συνολικό μοντέλο διαδικασίας που προέκυψε, είναι καλύτερο να δείξουμε μια αφαιρετική αναπαράστασή του χρησιμοποιώντας την ακαδημαϊκή σημειογραφία που θα είναι πιο κατανοητή στους τελικούς χρήστες ή, ακόμη καλύτερα, με τη μορφή μιας λίστας προτάσεων, συστάσεων και συμπερασμάτων σχετικά με τα αποτελέσματα. Οι προτάσεις αυτές μπορεί να θεωρηθούν ως επέκταση και βελτίωση του μοντέλου.
- **Εφαρμογή της εξόρυξης εκπαιδευτικών διαδικασιών σε άλλους αναδύομενους εκπαιδευτικούς τομείς** όπως παιχνίδια και κινητά. Τα ίχνη που προκύπτουν από τη δραστηριότητα ενός μαθητή στα παιχνίδια μάθησης είναι δύσκολο να αναλυθούν και να ερμηνευτούν από τους εκπαιδευτικούς. Οι τεχνικές EPM έχουν αρχίσει να εφαρμόζονται στις εφαρμογές Learning Games προκειμένου να αναλυθούν οι συμπεριφορές των μαθητών. Επιπλέον, ψηφιακές φορητές συσκευές όπως tablet, PDA και έξυπνα τηλέφωνα χρησιμοποιούνται όλο και πιο συχνά για εκπαιδευτικούς σκοπούς. Για το λόγο αυτό, η εξόρυξη διαδικασιών πρέπει να συνδυαστεί με τεχνικές ανάλυσης συμπεριφοράς εκπαιδευόμενων σε ένα συνεργατικό και πολύπλευρο περιβάλλον.
- **Δημοσίευση περισσότερων δωρεάν συνόλων εκπαιδευτικών δεδομένων,** προκειμένου να δοκιμαστούν καλύτερα τα θεωρητικά μοντέλα με μεθόδους και τεχνικές εξόρυξης. Τα δωρεάν σύνολα εκπαιδευτικών δεδομένων θα μπορούσαν να είναι πολύ χρήσιμα για τη δοκιμή ορισμένων ad-hoc μοντέλων κι έτσι η έρευνα πάνω στην εξόρυξη εκπαιδευτικών διαδικασιών θα μπορούσε να αναπτυχθεί πολύ πιο γρήγορα. Επομένως, ένα από τα πιο σημαντικά επόμενα βήματα είναι η προώθηση της ανταλλαγής διαδικτυακών συλλογών δεδομένων, ανωνυμοποιημένων και διαθέσιμων προς επεξεργασία, που μπορούν να αναλυθούν με διάφορες μεθόδους και εργαλεία.

Οι παραπάνω προκλήσεις που αναφέρονται στην εξόρυξη εκπαιδευτικών διαδικασιών αφορούν κυρίως διαδικασίες που έχουν να κάνουν με το μαθησιακό κομμάτι της εκπαίδευσης και κυρίως τις πλατφόρμες online μάθησης. Είναι διαδικασίες που στηρίζονται σε σύγχρονες και αποτελεσματικές στρατηγικές διδασκαλίας (διερευνητική μάθηση, ομαδοσυνεργατική διδασκαλία, μάθηση μέσω της τέχνης, συστημική προσέγγιση, διαφοροποιημένη διδασκαλία, βιωματική μάθηση, project) και

επικεντρώνονται στη διαδικασία της μάθησης. Η παρούσα διπλωματική εργασία ωστόσο αφορά διαδικασίες που αναφέρονται στο διοικητικό (administrative) κομμάτι εκπαιδευτικών διεργασιών. Η Εκπαιδευτική Διοίκηση θεωρείται ως η διαδικασία ενσωμάτωσης των κατάλληλων ανθρώπινων και υλικών πόρων που διατίθενται για την επίτευξη των σκοπών μιας εκπαιδευτικής διαδικασίας. Ο όρος «Διοίκηση» εν προκειμένω δεν αναφέρεται σε καμία μεμονωμένη διαδικασία ή πράξη αλλά περιλαμβάνει μια σειρά από διαδικασίες όπως: σχεδιασμός, οργάνωση, διεύθυνση, συντονισμός, έλεγχος και αξιολόγηση της απόδοσης. Η έννοια της εκπαιδευτικής διοίκησης υφίσταται στην περίπτωση ενός εκπαιδευτικού οργανισμού που έχει συγκεκριμένους σκοπούς ή στόχους να εκπληρώσει. Για την επίτευξη αυτών των σκοπών ή στόχων ο επικεφαλής του εκπαιδευτικού οργανισμού σχεδιάζει προσεκτικά διάφορα προγράμματα και δραστηριότητες και τα οργανώνει με τη συνεργασία όλων των εμπλεκόμενων στην εκάστοτε εκπαιδευτική διαδικασία. Επίσης, αξιολογεί την απόδοση και την πρόοδο στην επίτευξη των σκοπών του εκπαιδευτικού προγράμματος και εντέλει παρέχει ανατροφοδότηση και φέρνει τροποποιήσεις εάν και όπου απαιτείται (Kashyap D., 2020).

### **3.5 Εξόρυξη Εκπαιδευτικών Δεδομένων για την Ανάλυση της Συμπεριφοράς των Εκπαιδευόμενων**

Ένας τρόπος για να επιτευχθεί υψηλότερο επίπεδο ποιότητας υπηρεσιών στα εκπαιδευτικά συστήματα είναι να ανακαλυφθούν γνώσεις και να γίνουν προβλέψεις σχετικά με την εγγραφή εκπαιδευόμενων σε ένα συγκεκριμένο μάθημα, την ανίχνευση αθέμιτων μέσων που χρησιμοποιούνται στην ηλεκτρονική εξέταση, την ανίχνευση μη φυσιολογικών τιμών στα φύλλα αποτελεσμάτων των εκπαιδευόμενων, την πρόβλεψη για την απόδοση των εκπαιδευόμενων και ούτω καθεξής. Η γνώση κρύβεται μέσα στα σύνολα εκπαιδευτικών δεδομένων και μπορεί να εξαχθεί μέσω τεχνικών εξόρυξης (Baradwaj B., 2011).

Τα δεδομένα που συλλέγονται από διαφορετικές εφαρμογές απαιτούν κατάλληλη μέθοδο εξαγωγής γνώσεων από μεγάλα αποθετήρια για καλύτερη λήψη αποφάσεων. Οι τεχνικές εξόρυξης χρησιμοποιούνται για να λειτουργήσουν σε μεγάλους όγκους δεδομένων για να ανακαλύψουν κρυμμένα μοτίβα και σχέσεις χρήσιμες στη λήψη αποφάσεων. Στη σύγχρονη εκπαίδευση χρησιμοποιούνται διάφορα συστήματα πληροφοριών για την υποστήριξη εκπαιδευτικών διαδικασιών. Στις περισσότερες περιπτώσεις αυτά τα συστήματα έχουν δυνατότητες καταγραφής για έλεγχο και παρακολούθηση των διαδικασιών που υποστηρίζουν. Π.χ. σε επίπεδο πανεπιστημίου, τα διοικητικά συστήματα πληροφοριών συλλέγουν πληροφορίες σχετικά με τους φοιτητές, την εγγραφή τους σε συγκεκριμένα προγράμματα και μαθήματα, καθώς και τους βαθμούς εξέτασης στο τέλος του κάθε εξαμήνου. Αυτά τα δεδομένα μπορούν να αναλυθούν από διάφορα επίπεδα και προοπτικές, δείχνοντας διαφορετικές πτυχές της εκπαιδευτικής διαδικασίας (Trcka N., 2009).

Ωστόσο, έχει παρατηρηθεί ότι οι υπάρχουσες προσεγγίσεις εξόρυξης διαδικασιών έχουν μερικές φορές σοβαρές δυσκολίες στην εξόρυξη ολοκληρωμένων και εύληπτων γραφικών αναπαραστάσεων των εκπαιδευτικών δεδομένων. Ο μεγάλος όγκος και διασκορπισμός των δεδομένων σε βάσεις από διαφορετικές πλατφόρμες και οι εξαρτήσεις που έχουν αυτά μεταξύ τους οδηγούν πολλές φορές σε ισχυρά κυριαρχούντα αλλά χωρίς νόημα μοντέλα διεργασιών, πράγμα που καθιστά τη διαδικασία της εξόρυξης ως μία πρόκληση.

Επιπλέον, η αύξηση των πόρων στις πλατφόρμες ηλεκτρονικής μάθησης, τα υπάρχοντα εκπαιδευτικά λογισμικά, η χρήση του διαδικτύου στην εκπαίδευση και η δημιουργία από τους εκάστοτε εκπαιδευτικούς φορείς βάσεων δεδομένων με πληροφορίες για τους εκπαιδευόμενους, δημιούργησαν μεγάλα αποθετήρια δεδομένων. Όλες αυτές οι πληροφορίες παρέχουν ένα χρυσωρυχείο εκπαιδευτικών δεδομένων που μπορούν να διερευνηθούν και να αξιοποιηθούν για να γίνει κατανοητό το «πώς μαθαίνουν οι μαθητές» και μία από τις μεγαλύτερες προκλήσεις που αντιμετωπίζουν σήμερα οι φορείς της Εκπαίδευσης είναι η εκθετική αύξηση των εκπαιδευτικών δεδομένων και η χρήση αυτών των δεδομένων για τη βελτίωση της ποιότητας των διαχειριστικών αποφάσεων (Romero C., 2013).

Στην παρούσα εργασία επικεντρωνόμαστε στη συλλογή και επεξεργασία των δεδομένων που αφορούν στο εκπαιδευτικό σύστημα των Πανελλαδικών Εξετάσεων. Με τη διαδικασία των Πανελλαδικών Εξετάσεων, η επιτυχία των υποψηφίων καθορίζεται από τις βαθμολογίες στην εξέταση σε τέσσερα γενικά μαθήματα της ομάδας προσανατολισμού που έχουν επιλέξει, βαθμολογίες σε τυχόν εξέταση σε ειδικά μαθήματα εάν αυτή απαιτείται για την εισαγωγή σε συγκεκριμένες σχολές και τέλος την επιλογή και τη σειρά επιλογής των σχολών στο μηχανογραφικό, η οποία πραγματοποιείται συνήθως μετά από την ανακοίνωση των βαθμολογιών.

## **4. ΕΞΟΥΞΗ ΔΙΑΔΙΚΑΣΙΩΝ ΠΑΝΕΛΛΑΔΙΚΩΝ ΕΞΕΤΑΣΕΩΝ**

### **4.1 Διαδικασία διεξαγωγής Πανελλαδικών Εξετάσεων για την πρόσβαση στην Τριτοβάθμια Εκπαίδευση**

Οι Πανελλαδικές εξετάσεις είναι θεσμός του ελληνικού κράτους για την εισαγωγή μαθητών και αποφοίτων των Λυκείων στις σχολές της τριτοβάθμιας εκπαίδευσης, που ισχύει από το 1960 μέχρι σήμερα. Λαμβάνουν χώρα στο τέλος της σχολικής χρονιάς, τον Ιούνιο, κάθε χρόνο για τους μαθητές της τρίτης Λυκείου και τους αποφοίτους Λυκείων και οργανώνονται από το Υπουργείο Παιδείας και Θρησκευμάτων. Τα θέματα των εξετάσεων είναι κοινά για όλους τους υποψηφίους που λαμβάνουν μέρος στις εξετάσεις αυτές και μεταδίδονται κεντρικά από το υπουργείο Παιδείας προς όλα τα σχολεία. Τα αποτελέσματα των εξετάσεων ανακοινώνονται την ίδια ημερομηνία για όλους τους υποψηφίους, μέσω Διαδικτύου, αλλά και μέσω των Λυκείων τους (<https://el.wikipedia.org/>).

Η μελέτη περίπτωσης (case study) στην παρούσα εργασία στοχεύει στις διεργασίες που λαμβάνουν χώρα κατά τη συνολική διενέργεια και οργάνωση των πανελλαδικών εξετάσεων. Συγκεκριμένα, την πορεία ενός υποψηφίου από τη στιγμή που θα υποβάλει αίτηση-δήλωση συμμετοχής στις πανελλαδικές εξετάσεις, τη βαθμολόγηση των γραπτών του και την καταχώρησή της στο πληροφοριακό σύστημα, την υποβολή μηχανογραφικού δελτίου με τη δήλωση των προτιμήσεών του και τέλος την ανακοίνωση της σχολής επιτυχίας και την εγγραφή του σε αυτή. Η εξόρυξη διαδικασιών στις διεργασίες των πανελλαδικών εξετάσεων αποσκοπεί στο να μετασηματιστούν τα δεδομένα που είναι αποθηκευμένα στα αρχεία καταγραφής συμβάντων των πληροφοριακών συστημάτων σε πληροφορία και γνώση. Στόχος είναι τα αποτελέσματα της ανάλυσης που θα προκύψει από την εξόρυξη διαδικασιών να χρησιμοποιηθούν για να βελτιώσουν την όλη διαδικασία καθώς και για τον έλεγχο συμβατότητας με τους υφιστάμενους κανόνες.

### **4.2 Καταγραφή διαδικασίας των Πανελλαδικών εξετάσεων**

Οι διαδικασίες των Πανελλαδικών Εξετάσεων που ακολουθούνται από τις Διευθύνσεις Εξετάσεων και Πιστοποιήσεων, Ανάπτυξης Πληροφοριακών Συστημάτων και Εκπαιδευτικής Τεχνολογίας και Καινοτομίας του ΥΠΑΙΘ , σε συνεργασία με τις Διευθύνσεις Δευτεροβάθμιας Εκπαίδευσης και τις Σχολικές Μονάδες της χώρας, είναι οι παρακάτω:

Αρχικά οι υποψήφιοι υποβάλλουν στις σχολικές μονάδες Αίτηση-Δήλωση συμμετοχής στις Πανελλαδικές εξετάσεις του τρέχοντος έτους, όπου επιλέγουν την ομάδα προσανατολισμού που επιθυμούν (καθεμία από τις οποίες περιλαμβάνει συγκεκριμένα μαθήματα) καθώς και ειδικά μαθήματα που τυχόν απαιτούνται για την εισαγωγή τους σε ορισμένες σχολές της τριτοβάθμιας εκπαίδευσης, μουσικά μαθήματα για την εισαγωγή σε τμήματα μουσικών σπουδών ή πρακτικές δοκιμασίες για την εισαγωγή σε

τμήματα ΤΕΦΑΑ. Αίτηση-δήλωση έχουν το δικαίωμα να υποβάλουν οι μαθητές της Γ' τάξης του Λυκείου καθώς και παλαιοί απόφοιτοι οι οποίοι είναι κάτοχοι απολυτηρίου οποιουδήποτε τύπου Λυκείου. Η διαδικασία της Αίτησης-Δήλωσης γίνεται ηλεκτρονικά και υποστηρίζεται από το ΠΣ-MySchool, ένα ενιαίο πληροφοριακό σύστημα που υποστηρίζεται από τη Διεύθυνση Εκπαιδευτικής Τεχνολογίας και Καινοτομίας του ΥΠΑΙΘ και το οποίο έχει στόχο τη μηχανογραφική υποστήριξη των σχολικών μονάδων και των διοικητικών δομών της εκπαίδευσης στην ελληνική επικράτεια. Πιο συγκεκριμένα, ο υποψήφιος καταθέτει την αίτηση-δήλωσή του στο σχολείο και ο υπεύθυνος του ΠΣ-MySchool την καταχωρεί στο πληροφοριακό σύστημα. Εάν ο υποψήφιος είναι τελειόφοιτος αντλούνται τα μαθήματα στα οποία θα εξεταστεί βάσει της ομάδας προσανατολισμού που έχει ήδη επιλέξει από την αρχή της χρονιάς, αλλιώς επιλέγει στην αίτηση-δήλωση την ομάδα προσανατολισμού που επιθυμεί. Κατά τη διαδικασία αυτή, ο υποψήφιος επιλέγει αν επιθυμεί να εξεταστεί σε ειδικά μαθήματα, σε στρατιωτικές σχολές, σε αγωνίσματα για εισαγωγή στα ΤΕΦΑΑ, εάν επιθυμεί εξέταση μόνο για τμήματα μουσικών σπουδών και εάν τέλος έχει το δικαίωμα να εξεταστεί προφορικά. Αφού ολοκληρωθεί η διαδικασία καταχώρησης των στοιχείων, ο υπεύθυνος κάνει αποθήκευση και οριστική παραλαβή της αίτησης-δήλωσης στο πληροφοριακό σύστημα και στη συνέχεια την εκτυπώνει, ο υποψήφιος την ελέγχει και την υπογράφει.

Στη συνέχεια, αφού οριστικοποιηθεί το αρχείο με τις αιτήσεις-δηλώσεις, εξάγεται από το ΠΣ Myschool και παραδίδεται στη Διεύθυνση Ανάπτυξης Πληροφοριακών Συστημάτων του ΥΠΑΙΘ όπου και αποθηκεύεται σε διαφορετικό πληροφοριακό σύστημα. Το αρχείο των αιτήσεων αποστέλλεται σε κάθε σχολείο για επιβεβαίωση των αιτήσεων-δηλώσεων και επίσης επικαιροποιείται μετά από κάθε τροποποίηση που θα γίνει στο ΠΣ Myschool. Πριν από τις εξετάσεις, με την έκδοση υπουργικών αποφάσεων ορίζονται τα εξεταστικά κέντρα των μαθημάτων και των πρακτικών δοκιμασιών στα οποία θα εξεταστούν οι υποψήφιοι. Μετά από την εξέταση, με νέες υπουργικές αποφάσεις ορίζονται τα βαθμολογικά κέντρα στα οποία κατανέμονται τα γραπτά δοκίμια των υποψηφίων και όπου πραγματοποιείται η βαθμολόγηση του κάθε μαθήματος για τον κάθε υποψήφιο. Η διαδικασία της καταχώρισης της βαθμολογίας πραγματοποιείται μέσω ηλεκτρονικής εφαρμογής που διαθέτουν τα βαθμολογικά κέντρα, στην οποία καταχωρούνται, για κάθε μάθημα και για κάθε υποψήφιο, οι βαθμοί του 1<sup>ου</sup> και του 2<sup>ου</sup> βαθμολογητή και αν υπάρχει και ο βαθμός του αναβαθμολογητή. Οι βαθμολογίες αυτές συγκεντρώνονται στη Διεύθυνση Ανάπτυξης Πληροφοριακών Συστημάτων και πραγματοποιείται ο υπολογισμός των τελικών μορίων πρόσβασης για τον κάθε υποψήφιο βάσει της ισχύουσας νομοθεσίας.

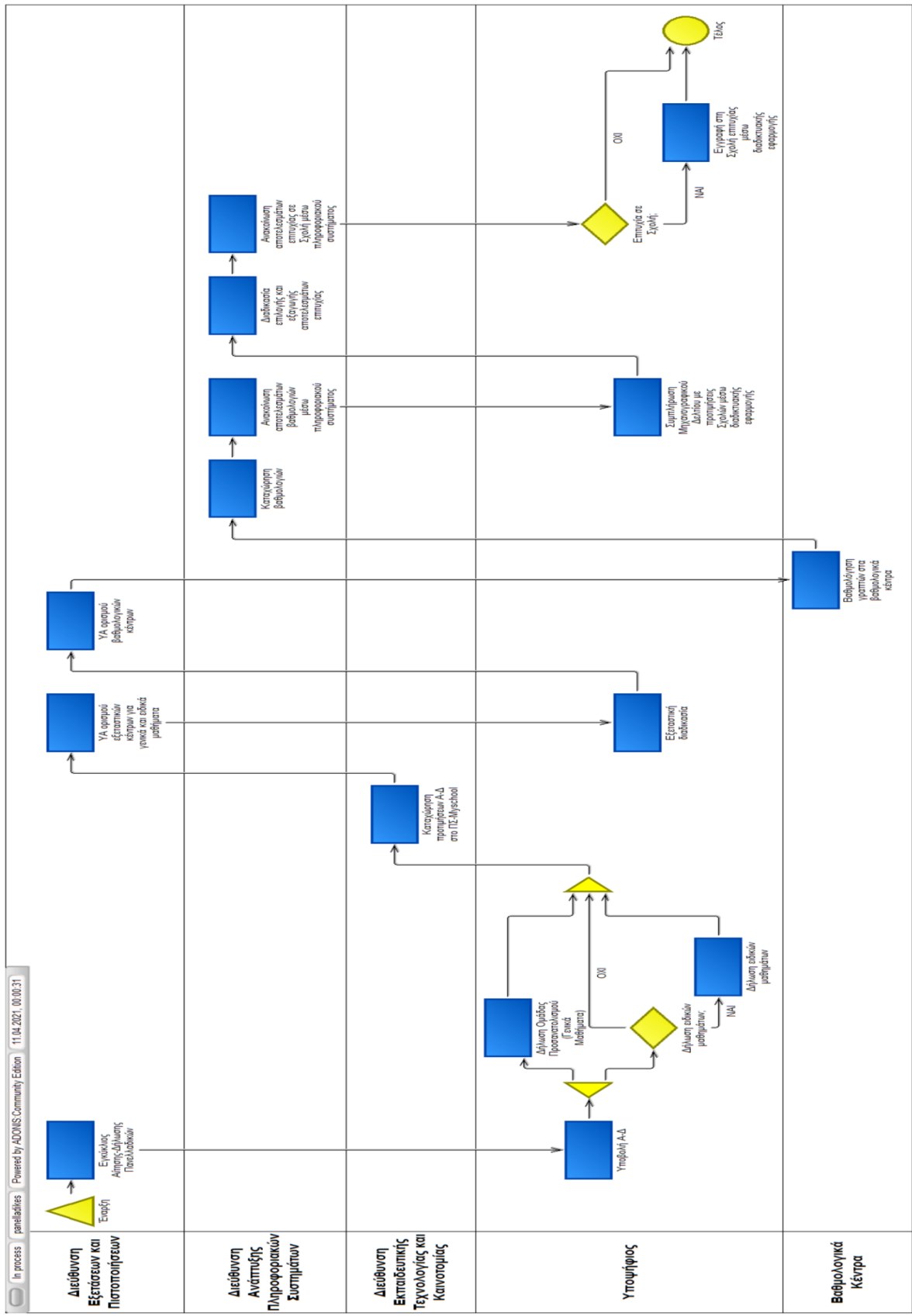
Μετά από την ανακοίνωση των βαθμολογιών, η οποία γίνεται μέσω συγκεκριμένης ιστοσελίδας όπου έχει πρόσβαση ο κάθε υποψήφιος με κωδικούς που διαθέτει, οι υποψήφιοι καλούνται να δημιουργήσουν και να συμπληρώσουν το μηχανογραφικό τους δελτίο χρησιμοποιώντας τη διαδικτυακή εφαρμογή «Σύστημα Ηλεκτρονικού Μηχανογραφικού» για την ηλεκτρονική καταχώριση του Μηχανογραφικού τους

Δελτίου στην Κεντρική Βάση του Υπουργείου Παιδείας και Θρησκευμάτων. Στόχος του εν λόγω πληροφοριακού συστήματος είναι η καταγραφή των επιλογών των υποψηφίων με τη σειρά προτεραιότητας που ο ίδιος επιθυμεί, για εισαγωγή στα Ανώτερα και Ανώτατα Εκπαιδευτικά Ιδρύματα μέσω του θεσμού των Πανελλαδικών Εξετάσεων.

Τέλος, μετά από τις διαδικασίες επιλογής που εφαρμόζονται ηλεκτρονικά από τη Διεύθυνση Ανάπτυξης Πληροφοριακών Συστημάτων του ΥΠΑΙΘ, εκδίδονται τα αποτελέσματα της επιλογής, δηλαδή η σχολή επιτυχίας του κάθε υποψηφίου. Σε μεταγενέστερο χρόνο, από το Υπουργείο Παιδείας και Θρησκευμάτων ανακοινώνεται ότι οι εγγραφές των επιτυχόντων από τα Γενικά Λύκεια (ΓΕΛ) ή από τα Επαγγελματικά Λύκεια (ΕΠΑΛ), στα Τμήματα και τις Σχολές της Τριτοβάθμιας Εκπαίδευσης επιτυχίας τους, γίνονται ηλεκτρονικά μέσω της διαδικτυακής εφαρμογής Ηλεκτρονικής Εγγραφής στην Τριτοβάθμια Εκπαίδευση «Ηλεκτρονικές Εγγραφές Επιτυχόντων».

Από τα παραπάνω είναι εμφανές ότι πλέον γίνεται εκτεταμένη χρήση πληροφοριακών συστημάτων προκειμένου να διεκπεραιωθεί η διαδικασία των πανελλαδικών εξετάσεων με ασφάλεια και με αξιοπιστία κι αυτό είναι κάτι που επιτρέπει την άντληση δεδομένων και την περαιτέρω επεξεργασία τους με τη βοήθεια διαφόρων εφαρμογών για την εξαγωγή χρήσιμων συμπερασμάτων.

Προκειμένου να υπάρχει μια ολοκληρωμένη επισκόπηση της πλήρους διαδικασίας των Πανελλαδικών Εξετάσεων, δημιουργήθηκε μια διαγραμματική αναπαράσταση με τη μορφή ενός BPMN διαγράμματος από το λογισμικό ADONIS (Community Edition Business Process Management Toolkit), η οποία απεικονίζεται στην παρακάτω εικόνα:



Εικόνα 8: Διάγραμμα διαδικασιών Πανελλαδικών Εξετάσεων στο λογισμικό ADONIS



### **4.3 Βασικοί στόχοι της ανάλυσης και εξόρυξης της διαδικασίας Πανελλαδικών Εξετάσεων**

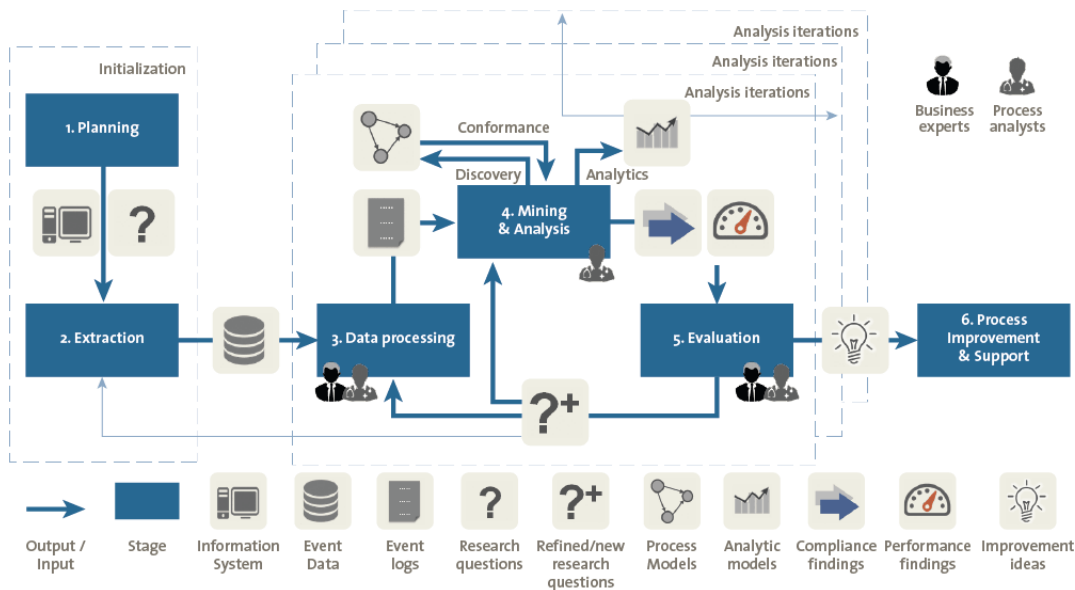
Σε γενικές γραμμές η διαδικασία των Πανελλαδικών Εξετάσεων πρόκειται για μια διαδικασία που διέπεται από αυστηρούς κανόνες και νομοθεσία που ορίζουν τις επιμέρους διεργασίες της. Θεωρητικά δεν υπάρχουν ιδιαίτερα σύνθετα βήματα και δεν προβλέπονται πολλές εναλλακτικές διαδρομές κατά την εκτέλεσή της. Ωστόσο, επειδή ο όγκος των δεδομένων είναι ιδιαίτερα μεγάλος, καθώς εξετάζονται περίπου 90.000 υποψήφιοι σε ετήσια βάση, κι επειδή οι χειριστές του πληροφοριακού συστήματος που αποθηκεύει τα δεδομένα των υποψηφίων αυτών είναι εκπαιδευτικοί που ορίζονται ως υπεύθυνοι από κάθε Λύκειο της χώρας, αυξάνουν οι πιθανότητες για αποτύπωση στα αρχεία καταγραφής συμβάντων εναλλακτικών διαδρομών και αποκλίσεων από τη διαδικασία όπως αυτή είναι καταγεγραμμένη. Επιπροσθέτως, πρόκειται για μια διαδικασία όπου ο χρονισμός παίζει σημαντικό ρόλο καθώς οι επιμέρους διεργασίες πρέπει να διαδέχονται η μία την άλλη ακολουθώντας τις οριζόμενες προθεσμίες.

Για τους παραπάνω λόγους, η εφαρμογή τεχνικών εξόρυξης διαδικασιών θα μπορούσε να συμβάλει στη δημιουργική ανάλυση της διαδικασίας των Πανελλαδικών εξετάσεων. Κατά την ανακάλυψη διεργασιών και τη δημιουργία μοντέλων που θα προκύψουν από την εξόρυξη της ανωτέρω διαδικασίας, βασικός στόχος είναι να διερευνήσουμε τι πραγματικά συμβαίνει στο σύνολο των διαδρομών της διαδικασίας και αν αυτό ταυτίζεται με αυτό που νομίζουμε ότι ισχύει ή αν υπάρχουν αποκλίσεις από τους κανόνες και τα καθορισμένα βήματα. Επιπλέον, στόχος είναι να δούμε εάν υπάρχει κάποια συμφόρηση στην εκτέλεση των διεργασιών ή στη μετάβαση από τη μία διεργασία στην άλλη. Θα βοηθήσει επομένως η εξόρυξη στο να προκύψει μια αντικειμενική άποψη της πραγματικότητας, η οποία θα μελετηθεί στη συνέχεια προκειμένου να βγουν συμπεράσματα που θα οδηγήσουν στη βελτιστοποίηση και της συνολικής διαδικασίας αλλά και του πληροφοριακού συστήματος που την υποστηρίζει. Τέλος, επειδή τα δεδομένα αποθηκεύονται σε διαφορετικά πληροφοριακά συστήματα, ένας ακόμη στόχος είναι να παρατηρήσουμε, εάν αυτό είναι εφικτό, αν υπάρχει πρόβλημα επικοινωνίας μεταξύ των πληροφοριακών συστημάτων που συμμετέχουν στην καταγραφή των δεδομένων.

Επειδή ο όγκος των δεδομένων είναι πολύ μεγάλος, η ανάλυση των στοιχείων που βρίσκονται καταχωρημένα στα πληροφοριακά συστήματα αναμένεται να οδηγήσει σε ένα αξιόπιστο αποτέλεσμα και συμπέρασμα.

### **4.4 Γενική μεθοδολογία εξόρυξης διαδικασιών**

Μία επισκόπηση της γενικότερης μεθοδολογίας που ακολουθείται κατά την εξόρυξη διαδικασιών περιγράφεται στην παρακάτω εικόνα:



Εικόνα 9: Επισκόπηση της μεθοδολογίας για την εξόρυξη διαδικασιών (van Eck M.L., 2015, σ. 3)

Η παραπάνω μεθοδολογία αποτελείται από έξι στάδια που σχετίζονται με πολλά διαφορετικά αντικείμενα εισόδου και εξόδου των ακόλουθων τύπων: αντικείμενα που σχετίζονται με στόχους, αντικείμενα δεδομένων και μοντέλα.

Τα αντικείμενα που σχετίζονται με τους στόχους είναι ερευνητικά ερωτήματα που προέρχονται από στόχους έργου, τα οποία απαντώνται από ευρήματα απόδοσης και ευρήματα συμμόρφωσης που προκύπτουν από την εφαρμογή τεχνικών εξόρυξης διαδικασιών, οδηγώντας σε ιδέες βελτίωσης για την επίτευξη των στόχων.

Τα αντικείμενα δεδομένων υποδηλώνουν τις αναπαραστάσεις των δεδομένων που σχετίζονται με τη διαδικασία: τα συστήματα πληροφοριών περιέχουν δεδομένα ζωντανών διεργασιών σε διάφορες μορφές, τα οποία μπορούν να εξαχθούν και να συνδεθούν με διακριτά συμβάντα για να σχηματίσουν δεδομένα συμβάντων. Τα δεδομένα συμβάντων μπορούν να μετατραπούν σε αρχεία καταγραφής.

Τέλος, υπάρχουν δύο τύποι μοντέλων: (1) μοντέλα διεργασίας και (2) αναλυτικά μοντέλα. Τα μοντέλα διεργασίας περιγράφουν την ταξινόμηση των δραστηριοτήτων σε μια διαδικασία, πιθανώς ενισχυμένη με πρόσθετες πληροφορίες π.χ. χρονικοί περιορισμοί, χρήση πόρων ή χρήση δεδομένων. Τα αναλυτικά μοντέλα είναι οποιοσδήποτε άλλος τύπος μοντέλων που δίνουν πληροφορίες για τη διαδικασία, π.χ. δέντρα αποφάσεων (van Eck M.L., 2015).

#### 4.5 Λήψη δεδομένων των Πανελλαδικών Εξετάσεων από το Υπουργείο Παιδείας και προετοιμασία τους για αξιοποίηση με τεχνικές εξόρυξης

Οι υπηρεσίες του Υπουργείου Παιδείας και Θρησκευμάτων, που είναι ο φορέας διοργάνωσης των Πανελλαδικών Εξετάσεων, έχουν αναπτύξει διάφορες ηλεκτρονικές

πλατφόρμες προκειμένου να υποστηρίζεται και να καταγράφεται η πορεία ενός υποψηφίου από την εκδήλωση της επιθυμίας του για συμμετοχή στις Πανελλαδικές Εξετάσεις μέχρι και την εγγραφή του σε κάποια Σχολή που έχει επιτύχει. Τα δεδομένα που αποθηκεύονται στις πλατφόρμες αυτές μπορούν να συλλεχθούν και να αξιοποιηθούν, αφού προηγηθεί μία προεπεξεργασία αυτών.

Σε εκπαιδευτικά πλαίσια, είναι φυσικό η προεπεξεργασία των δεδομένων να είναι μια πολύ σημαντική και περίπλοκη εργασία και μερικές φορές ακόμη να καταλαμβάνει περισσότερο από το ήμισυ του συνολικού χρόνου που δαπανάται για την εφαρμογή τεχνικών αξιοποίησής τους. Οι λόγοι για τους οποίους συμβαίνει αυτό είναι, καταρχάς, ότι τα διαθέσιμα εκπαιδευτικά δεδομένα για την επίλυση ενός προβλήματος συνήθως δεν είναι στην κατάλληλη μορφή που απαιτεί η εκάστοτε επεξεργασία τους από εφαρμογές όπως π.χ. αυτές της εξόρυξης διαδικασιών, και κατά δεύτερον, δεδομένης της ετερογενούς φύσης των εκπαιδευτικών δεδομένων, ο καθορισμός των δομών που αντιπροσωπεύουν ένα υπό εξέταση γεγονός αποτελεί κλειδί στην όλη διαδικασία. Επομένως, είναι απαραίτητη η μετατροπή των δεδομένων σε κατάλληλη μορφή για την επίλυση ενός συγκεκριμένου εκπαιδευτικού προβλήματος. Αυτό περιλαμβάνει την επιλογή των δεδομένων που θα συλλεχθούν, την εστίαση στις ερωτήσεις που πρέπει να απαντηθούν και τη διασφάλιση της ευθυγράμμισης των δεδομένων με τις ερωτήσεις. Επίσης, είναι σημαντικό να διατηρηθεί και να προστατευθεί η εμπιστευτικότητα των πληροφοριών των εκπαιδευόμενων κατά την ενσωμάτωση όλων των συλλεγόμενων δεδομένων διαγράφοντας ορισμένα προσωπικά στοιχεία (τα οποία ούτως ή άλλως δεν είναι χρήσιμα για σκοπούς εξόρυξης) όπως όνομα, e-mail, αριθμός τηλεφώνου κ.ο.κ. (Romero C., 2013).

Το σημείο εκκίνησης για την εξόρυξη των διαδικασιών είναι, όπως έχει ήδη αναφερθεί, δεδομένα σε αρχεία καταγραφής συμβάντων (event logs). Υπάρχουν επομένως αρχεία καταγραφής στα οποία κάθε συμβάν αναφέρεται σε μια υπόθεση, μια δραστηριότητα και μια χρονική στιγμή. Ένα αρχείο καταγραφής συμβάντων μπορεί να θεωρηθεί ως μια συλλογή περιπτώσεων (που μερικές φορές αναφέρονται επίσης ως ίχνη), όπου κάθε περίπτωση αντιστοιχεί σε μια ακολουθία γεγονότων (Romero C., 2016).

Στην παρούσα εργασία, παρουσιάζεται πώς μπορεί να εισαχθεί ένα σύνολο εκπαιδευτικών δεδομένων σε μια πλατφόρμα που εφαρμόζει τεχνικές εξόρυξης διαδικασιών και πώς μπορούν να ληφθούν ορισμένα χρήσιμα αποτελέσματα και συμπεράσματα από αυτό το σύνολο. Τα δεδομένα που συλλέγονται από το εκπαιδευτικό μας σύστημα πληροφοριών και αφορούν στη διαδικασία των Πανελλαδικών Εξετάσεων θα πρέπει να αποθηκεύονται σε μια σχεσιακή βάση δεδομένων και να περιέχουν διάφορες πληροφορίες που θα ξεκινούν από την αίτηση-δήλωση επιθυμίας ενός υποψηφίου για συμμετοχή στις πανελλαδικές, θα περιέχουν δημογραφικά δεδομένα και δεδομένα προτιμήσεων των υποψηφίων, δεδομένα και χρονικές στιγμές βαθμολόγησης, υποβολής μηχανογραφικού δελτίου με τις προτιμήσεις σε Σχολές και τέλος δεδομένα και χρονικές στιγμές εγγραφής στις Σχολές επιτυχίας για τον κάθε υποψήφιο.

Έχοντας στην κατοχή μας λοιπόν τα παραπάνω δεδομένα και τις συσχετίσεις τους θα υπάρχει στην ουσία καταγεγραμμένη όλη η πορεία ενός υποψηφίου κατά τη διαδικασία των Πανελλαδικών Εξετάσεων και την τελική εγγραφή στο Τμήμα επιτυχίας του. Παρακάτω αναλύονται τα βήματα που ακολουθήθηκαν προκειμένου να αποκτηθούν τα δεδομένα αυτά και η επεξεργασία που πραγματοποιήθηκε στα δεδομένα που τελικά είχαμε στη διάθεσή μας για να προχωρήσουμε στη συνέχεια στη διαδικασία εξόρυξής τους.

#### **4.5.1 Προγραμματισμός και εξαγωγή δεδομένων**

Κατά την van Eck M.L., 2015 μεθοδολογία εξόρυξης διαδικασιών (Εικόνα 9), τα δύο πρώτα στάδια είναι ο προγραμματισμός και η εξαγωγή των δεδομένων, κατά τη διάρκεια των οποίων ορίζονται οι αρχικές ερευνητικές ερωτήσεις και εξάγονται τα δεδομένα συμβάντων. Στην παρούσα μελέτη περίπτωσης ο βασικός στόχος είναι να ελεγχθεί κατά πόσο οι διεργασίες που λαμβάνουν χώρα στην πραγματικότητα ταυτίζονται με τις διεργασίες όπως αυτές ορίζονται από τις αρμόδιες Διευθύνσεις του Υπουργείου Παιδείας και Θρησκευμάτων.

Στο πλαίσιο αυτό, πραγματοποιήθηκαν αιτήματα προς τη Διεύθυνση Ανάπτυξης Πληροφοριακών Συστημάτων και προς τη Διεύθυνση Εκπαιδευτικής Τεχνολογίας και Καινοτομίας του Υπουργείου.

Από τη Διεύθυνση Ανάπτυξης Πληροφοριακών Συστημάτων αιτήθηκε η παροχή δεδομένων που διαθέτουν πληροφορία για χρονικές στιγμές στις οποίες πραγματοποιήθηκαν γεγονότα σχετικά με τη διαδικασία των πανελλαδικών εξετάσεων, όπως α) εξέταση των υποψηφίων στα εξεταστικά κέντρα, β) βαθμολόγηση του κάθε μαθήματος για τον κάθε υποψήφιο, γ) υποβολή μηχανογραφικού δελτίου, δ) ανακοίνωση αποτελεσμάτων, ε) εγγραφή επιτυχόντων στα τμήματα επιτυχίας τους και συγκεκριμένα αιτήθηκαν οι παρακάτω πληροφορίες:

1. Η χρονική στιγμή στην οποία καταχωρήθηκαν οι βαθμολογίες του κάθε υποψηφίου σε κάθε βαθμολογικό κέντρο (ή συνολικά των υποψηφίων σε κάθε βαθμολογικό κέντρο) και από ποιον βαθμολογητή. Επίσης, πότε ολοκληρώθηκε η βαθμολόγηση σε κάθε βαθμολογικό κέντρο και για κάθε ΒΚ στοιχεία όπως όνομα, συντεταγμένες κ.α.

Επιπλέον, στοιχεία για οποιαδήποτε ενέργεια γίνεται από εκεί και πέρα και υπάρχει η πληροφορία της χρονικής στιγμής στην οποία έγινε (date ή timestamp).

2. Σε ποιο βαθμολογικό κέντρο βαθμολογείται το κάθε εξεταστικό κέντρο και πότε το κάθε ΒΚ παραλαμβάνει τα γραπτά από το ΕΚ.

3. Σε ποιο εξεταστικό κέντρο εξετάζεται ο κάθε υποψήφιος.

4. Η χρονική στιγμή όπου ο κάθε υποψήφιος οριστικοποίησε το Μηχανογραφικό του Δελτίο. Να υπάρχει επιπλέον η πληροφορία του φύλου, της βαθμολογίας/μορίων που συγκέντρωσε ο υποψήφιος, τα επιστημονικά πεδία για τα οποία συμπληρώνει ΜΔ, η πρώτη προτίμηση του υποψηφίου στο ΜΔ.

Επιπλέον, αν υπάρχει και πληροφορία χρονικών στιγμών στις οποίες έγιναν άλλες ενέργειες σχετικές με την υποβολή μηχανογραφικού.

5. Η χρονική στιγμή όπου ο κάθε υποψήφιος πραγματοποίησε την εγγραφή του στο Τμήμα επιτυχίας του. Να υπάρχει επιπλέον η πληροφορία του φύλου, του τμήματος επιτυχίας, του Ιδρύματος, η σειρά προτίμησης του τμήματος επιτυχίας.

Αντίστοιχα, από τη Διεύθυνση Εκπαιδευτικής Τεχνολογίας και Καινοτομίας αιτήθηκε η παροχή δεδομένων που αφορούν τη διαδικασία αίτησης-δήλωσης των υποψηφίων στο ΠΣ Myschool και συγκεκριμένα οι παρακάτω πληροφορίες:

1. Η χρονική στιγμή στην οποία δημιουργήθηκε η αίτηση-δήλωση για τον κάθε υποψήφιο, τότε τροποποιήθηκε τελευταία φορά και τότε έγινε η οριστική καταχώρηση. Να υπάρχει επιπλέον η πληροφορία του φύλου του υποψηφίου, σχολικής μονάδας, ομάδας προσανατολισμού, μαθημάτων που δήλωσε και τυχόν άλλων προτιμήσεων (ειδικά μαθήματα, ΤΕΦΑΑ, στρατιωτικές σχολές κτλ).

2. Για τη δήλωση μαθημάτων, τη χρονική στιγμή στην οποία δηλώθηκε το κάθε μάθημα και τότε τροποποιήθηκε τελευταία φορά.

Επιπλέον, στοιχεία για οποιαδήποτε ενέργεια γίνεται από εκεί και πέρα και αφορά την αίτηση-δήλωση και για την οποία υπάρχει η πληροφορία της χρονικής στιγμής στην οποία έγινε (date ή timestamp).

#### **4.5.2 Συλλογή δεδομένων**

Η Διεύθυνση Εκπαιδευτικής Τεχνολογίας και Καινοτομίας ανταποκρίθηκε αποστέλλοντας πολλά και διαφορετικά αρχεία για τις αιτήσεις-δηλώσεις:

Συγκεκριμένα, εστάλησαν 4 διαφορετικοί πίνακες για α) τη δημιουργία, β) την αποδοχή, γ) την ακύρωση και δ) την ολοκλήρωση των αιτήσεων-δηλώσεων των υποψηφίων, με στοιχεία όπως το ID της αίτησης-δήλωσης, την ημερομηνία της εκάστοτε ενέργειας, το ID του υποψηφίου, τη διεύθυνσή του, το έτος γέννησης, την κατηγορία του υποψηφίου (ΓΕΛ ή ΕΠΑΛ), το φύλο, την κατεύθυνση που έχει επιλέξει, το διδακτικό έτος.

Επιπλέον, εστάλησαν 2 πίνακες με δεδομένα για τις δηλώσεις α) των γενικών μαθημάτων που ανήκουν στις ομάδες προσανατολισμού των ΓΕΛ ή στους τομείς των ΕΠΑΛ και β) των ειδικών μαθημάτων που τυχόν έχει επιλέξει κάποιος υποψήφιος, με στοιχεία όπως το ID της αίτησης-δήλωσης, την ημερομηνία δήλωσης του μαθήματος, τον κωδικό και την ονομασία του μαθήματος, την ομάδα προσανατολισμού ή τομέα για τα γενικά μαθήματα, το ID του υποψηφίου, το διδακτικό έτος.

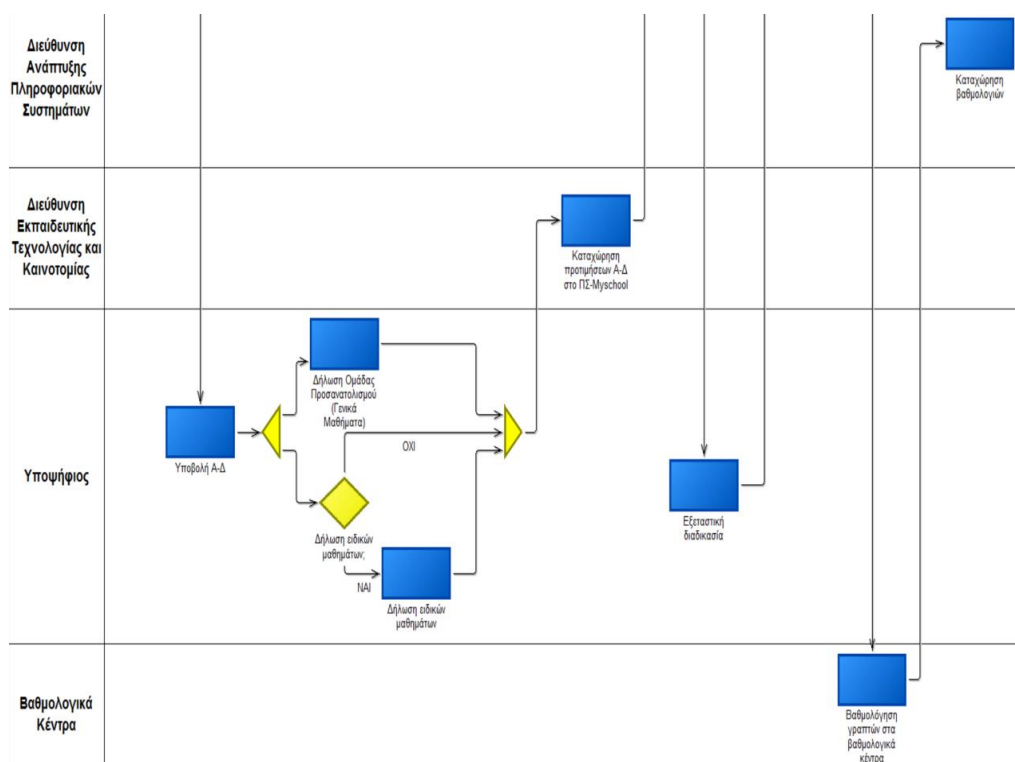
Τέλος, εστάλησαν 2 πίνακες με δεδομένα για την επεξεργασία στη δήλωση α) των γενικών μαθημάτων που ανήκουν στις ομάδες προσανατολισμού των ΓΕΛ ή στους τομείς των ΕΠΑΛ και β) των ειδικών μαθημάτων που τυχόν έχει επιλέξει κάποιος

υποψήφιος, με στοιχεία όπως το ID της αίτησης-δήλωσης, την ημερομηνία επεξεργασίας του μαθήματος, το ID της δήλωσης, τον κωδικό και την ονομασία του μαθήματος, την ομάδα προσανατολισμού ή τομέα για τα γενικά μαθήματα, το ID του υποψηφίου, το διδακτικό έτος.

Επομένως, σε ότι αφορά στη διαδικασία υποβολής της Αίτησης-Δήλωσης, η οποία είναι και η πρώτη που λαμβάνει χώρα κατά την συνολική διαδικασία των Πανελλαδικών Εξετάσεων και της εισαγωγής σε Σχολή επιτυχίας, υπάρχουν αρκετά δεδομένα που θα μπορούσαν να υποστούν κάποια περαιτέρω επεξεργασία και να εισαχθούν σε μία εφαρμογή εξόρυξης διαδικασιών. Στο σημείο αυτό πρέπει να τονιστεί πως τα παραπάνω στοιχεία εξήχθησαν με τη χρήση κατάλληλων ερωτημάτων από τη βάση δεδομένων του ΠΣ Myschool, καθώς δεν υπήρχαν αρχεία καταγραφής συμβάντων σε μορφή συμβατή με αυτή που απαιτείται από τις εφαρμογές εξόρυξης διαδικασιών. Συγκεκριμένα, επειδή το ΠΣ Myschool δεν αποθηκεύει πληροφορίες για το πότε έγινε η κάθε ενέργεια κατά τη διαδικασία της αίτησης-δήλωσης (αποθήκευση, παραλαβή, ακύρωση και επαναπαραλαβή κοκ), για να εξαχθεί η επιθυμητή πληροφορία προκειμένου να πραγματοποιηθεί η εξόρυξη διαδικασιών χρησιμοποιήθηκαν ερωτήματα προς τη βάση δεδομένων αξιοποιώντας τα πεδία που καθορίζουν αν η αίτηση-δήλωση έχει ακυρωθεί ή ολοκληρωθεί σε συνδυασμό με το πεδίο τελευταίας τροποποίησης της εγγραφής στη βάση.

Η Διεύθυνση Ανάπτυξης Πληροφοριακών Συστημάτων, από την οποία ζητήθηκαν δεδομένα που αφορούν στις υπόλοιπες διαδικασίες των Πανελλαδικών Εξετάσεων (βαθμολόγηση, υποβολή μηχανογραφικού δελτίου, εγγραφή σε σχολή επιτυχίας), απέστειλε 2 πίνακες για τη διαδικασία βαθμολόγησης στα βαθμολογικά κέντρα για τα έτη 2019 και 2020, με στοιχεία όπως το ID του υποψηφίου, ο κωδικός του εξεταστικού κέντρου, ο κωδικός του βαθμολογικού κέντρου (BK), η ημερομηνία καταχώρησης βαθμολογίας στο BK και η ημερομηνία ολοκλήρωσης βαθμολογίας μαθήματος στο BK.

Συνεπώς, με τα δεδομένα που έχουν εξαχθεί και αποσταλεί, δεν είναι δυνατό να πραγματοποιηθεί εξόρυξη στη συνολική διαδικασία των Πανελλαδικών Εξετάσεων παρά μόνο σε τμήμα αυτής, καθώς δεν είναι καταγεγραμμένη σε αρχεία καταγραφής συμβάντων (event logs) όλη η πορεία ενός υποψηφίου κατά τη διαδικασία των Πανελλαδικών Εξετάσεων και την τελική εγγραφή στο Τμήμα επιτυχίας του. Συγκεκριμένα, θα πραγματοποιηθεί εξόρυξη στο κομμάτι που αφορά τη διαδικασία των αιτήσεων-δηλώσεων και καταχώρησης βαθμολογίας ενός υποψηφίου, όπως απεικονίζεται παρακάτω:



Εικόνα 10: Διαδικασία αίτησης-δήλωσης Πανελλαδικών Εξετάσεων και καταχώρησης βαθμολογίας στο ΠΣ-MySchool

#### 4.6 Προετοιμασία δεδομένων για εισαγωγή στα εργαλεία εξόρυξης

Τα σύνολα δεδομένων τα οποία ήταν διαθέσιμα προς εξόρυξη, αφορούσαν κυρίως τη διαδικασία των Αιτήσεων-Δηλώσεων για το σχολικό έτος 2019-2020. Αξίζει να σημειωθεί ότι επελέγησαν προς ανάλυση μόνο τα δεδομένα του σχολικού έτους 2019-2020, παρόλο που πραγματοποιήθηκαν δοκιμές για το σύνολο των δεδομένων των σχολικών ετών 2013-2020 τα οποία και δόθηκαν από τη Διεύθυνση Εκπαιδευτικής Τεχνολογίας και Καινοτομίας, λόγω του μεγάλου όγκου των δεδομένων τα οποία δεν ήταν δυνατό να τα διαχειριστεί το λογισμικό ProM σε υπολογιστή με επεξεργαστή Intel i3-3220 3.30GHz και μνήμη RAM 4GB, με αποτέλεσμα να «παγώνουν» οι διαδικασίες μετατροπής σε αρχεία XES και να μην είναι δυνατό έτσι να πραγματοποιηθεί περαιτέρω ανάλυση.

Από τη Διεύθυνση Εκπαιδευτικής Τεχνολογίας και Καινοτομίας του Υπουργείου Παιδείας εστάλησαν με μορφή .csv αρχείων δεδομένα τα οποία αφορούσαν στην Καταχώρηση της Αίτησης-Δήλωσης, στην Αποδοχή της, στην Ολοκλήρωσή της (η οποία ήταν η οριστική παραλαβή και αποθήκευση στο σύστημα), στην πιθανή Ακύρωσή της εάν κάποιος υποψήφιος δεν επιθυμούσε τελικά να συμμετέχει στη διαδικασία των Πανελλαδικών Εξετάσεων, στην Καταχώρηση των Βασικών μαθημάτων στην Αίτηση-Δήλωση, στην πιθανή Καταχώρηση Ειδικών μαθημάτων στην Αίτηση-Δήλωση και τέλος στην Καταχώρηση της Βαθμολογίας των Βασικών μαθημάτων στο σύστημα. Οι διαδικασίες αυτές κατά την εκτέλεσή τους δημιουργούν ψηφιακά ίχνη στα αντίστοιχα

πληροφοριακά συστήματα που τις υποστηρίζουν, τα οποία με τη μορφή με την οποία εστάλησαν αποτελούν δεδομένα που μπορούν με τη χρήση κατάλληλων εφαρμογών να μετατραπούν σε αρχεία καταγραφής συμβάντων και να αναλυθούν. Τα παραπάνω .csv αρχεία ενοποιήθηκαν, συνδυάζοντας τις κοινές στήλες που είχαν και αφαιρώντας ορισμένες στήλες που έκαναν αργή την επεξεργασία τους, και δημιουργήθηκε το αρχείο *Aitiseis\_Diloseis.csv* με τις παρακάτω κεφαλίδες:

- ApplicationID: Ο κωδικός αριθμός της κάθε αίτησης-δήλωσης
- Activity: Η περιγραφή της κάθε δραστηριότητας (ΚΑΤΑΧΩΡΗΣΗ ΑΙΤΗΣΗΣ, ΚΑΤΑΧΩΡΗΣΗ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ, ΚΑΤΑΧΩΡΗΣΗ ΕΙΔΙΚΟΥ ΜΑΘΗΜΑΤΟΣ, ΑΚΥΡΩΣΗ ΑΙΤΗΣΗΣ, ΑΠΟΔΟΧΗ ΑΙΤΗΣΗΣ, ΟΛΟΚΛΗΡΩΣΗ ΑΙΤΗΣΗΣ, ΚΑΤΑΧΩΡΗΣΗ ΒΑΘΜΟΛΟΓΙΑΣ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ)
- ActivityDate: Η ημερομηνία στην οποία ολοκληρώθηκε η κάθε διεργασία/δραστηριότητα.
- BirthYear: Το έτος γέννησης του κάθε υποψηφίου
- PseudoCandidateCode: Ο κωδικός αριθμός του κάθε υποψηφίου, ο οποίος κατά την εξαγωγή των δεδομένων έχει υποστεί επεξεργασία προκειμένου τα τελικά δεδομένα να είναι ανωνυμοποιημένα
- DescriptionSchoolType: Ο διαχωρισμός του σχολείου του κάθε υποψηφίου σε Γενικό Λύκειο (ΓΕΛ) ή Επαγγελματικό Λύκειο (ΕΠΑΛ)
- ExamLessonSchoolType: Ο διαχωρισμός του σχολείου του κάθε υποψηφίου σε Ημερήσιο (Η) ή Εσπερινό (Ε)
- Sex: Το φύλο του κάθε υποψηφίου

Επιπλέον, συνδυάζοντας τα παραπάνω σύνολα δεδομένων με το αρχείο που εστάλη από τη Διεύθυνση Ανάπτυξης Πληροφοριακών Συστημάτων και αφορά στην καταχώρηση βαθμολογιών των μαθημάτων στα βαθμολογικά κέντρα (ΒΚ), προέκυψε ένα ακόμη αρχείο *Aitiseis\_Diloseis\_Kai\_BK.csv* με τις παρακάτω κεφαλίδες:

- PseudoCandidateCode: Ο κωδικός αριθμός του κάθε υποψηφίου, ο οποίος κατά την εξαγωγή των δεδομένων έχει υποστεί επεξεργασία προκειμένου τα τελικά δεδομένα να είναι ανωνυμοποιημένα
- Activity: Η περιγραφή της κάθε δραστηριότητας (ΚΑΤΑΧΩΡΗΣΗ ΑΙΤΗΣΗΣ, ΚΑΤΑΧΩΡΗΣΗ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ, ΚΑΤΑΧΩΡΗΣΗ ΕΙΔΙΚΟΥ ΜΑΘΗΜΑΤΟΣ, ΑΚΥΡΩΣΗ ΑΙΤΗΣΗΣ, ΑΠΟΔΟΧΗ ΑΙΤΗΣΗΣ, ΟΛΟΚΛΗΡΩΣΗ ΑΙΤΗΣΗΣ, ΟΛΟΚΛΗΡΩΣΗ ΒΑΘΜΟΛΟΓΙΑΣ ΜΑΘΗΜΑΤΟΣ ΣΤΟ ΒΚ, ΚΑΤΑΧΩΡΗΣΗ ΒΑΘΜΟΛΟΓΙΑΣ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ)
- ActivityDate: Η ημερομηνία στην οποία ολοκληρώθηκε η κάθε διεργασία/δραστηριότητα.

Τα παραπάνω δύο ενοποιημένα αρχεία αποτελούν τα αρχεία που εισήχθησαν στα λογισμικά εξόρυξης διαδικασιών ProM και Disco (Academic Lisence).

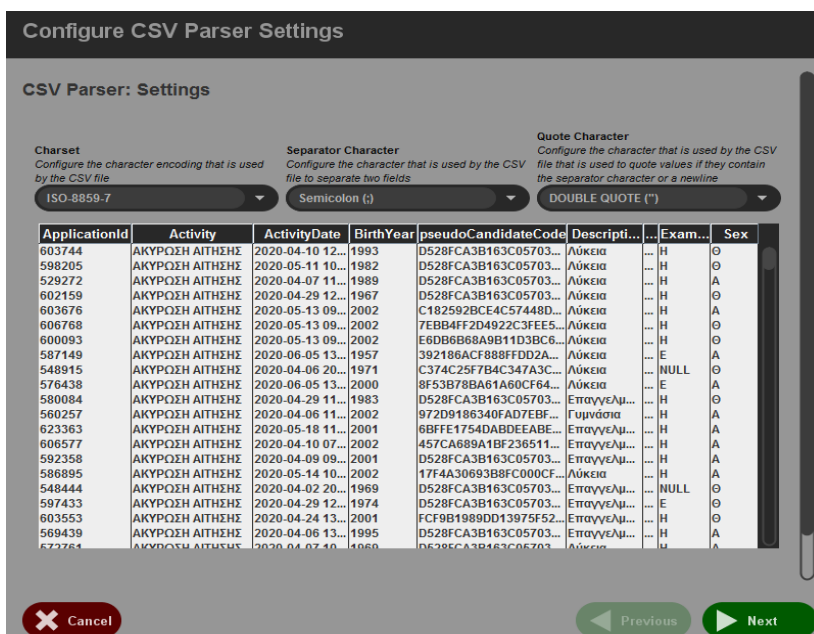


## 5. ΥΛΟΠΟΙΗΣΗ ΚΑΙ ΕΦΑΡΜΟΓΗ

### 5.1 Εισαγωγή των αρχείων στο ProM και επιλογές οπτικοποίησής τους

Τα παραπάνω αρχεία *Aitiseis\_Diloseis.csv* και *Aitiseis\_Diloseis\_Kai\_BK.csv* εισήχθησαν στο λογισμικό ProM PM 6.10 και στη συνέχεια με τη χρήση του πρόσθετου “Convert CSV to XES” δημιουργήθηκαν τα αντίστοιχα αρχεία καταγραφής συμβάντων (event logs) με επέκταση XES, προκειμένου να χρησιμοποιηθούν στην περαιτέρω ανάλυση των διαδικασιών με το ProM.

Κατά τη διαδικασία της μετατροπής, εμφανίζονται αρχικά όλες οι κεφαλίδες του αρχείου οι οποίες εισάγονται:



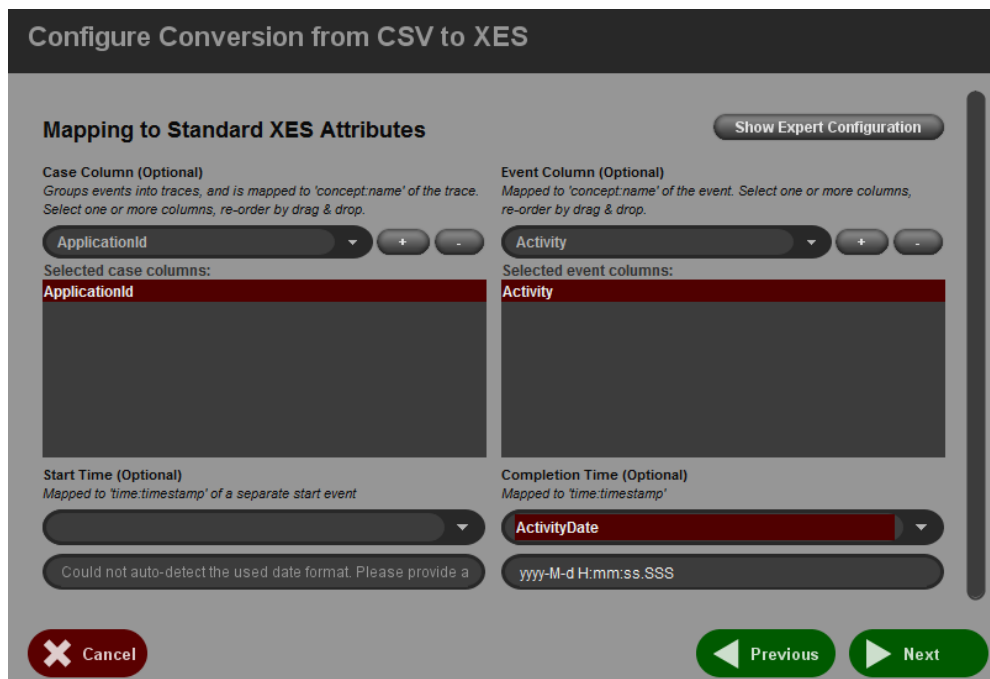
Εικόνα 11: Κεφαλίδες αρχείου καταγραφής συμβάντων

Στη συνέχεια ζητείται η επιλογή των βασικών παραμέτρων, οι οποίες είναι το CaseID, το Activity και το Timestamp, προκειμένου να δημιουργηθεί το αρχείο καταγραφής συμβάντων. Λαμβάνοντας υπόψη ότι ο στόχος της ανάλυσης που θα πραγματοποιηθεί είναι να διερευνηθεί η πορεία κάθε αίτησης-δήλωσης που υποβάλλει ο κάθε υποψήφιος, επελέγησαν οι παρακάτω παράμετροι, όπως απεικονίζονται παρακάτω:

**CaseID:** ApplicationID

**Activity:** Activity

**Timestamp:** ActivityDate



Εικόνα 12: Επιλογή βασικών παραμέτρων

Για το δεύτερο αρχείο, το οποίο περιέχει επιπλέον και στοιχεία από τα βαθμολογικά κέντρα (BK), υπάρχουν μόνο τρεις κεφαλίδες, επομένως οι βασικές παράμετροι που επελέγησαν ήταν αντίστοιχα:

**CaseID:** PseudoCandidateCode

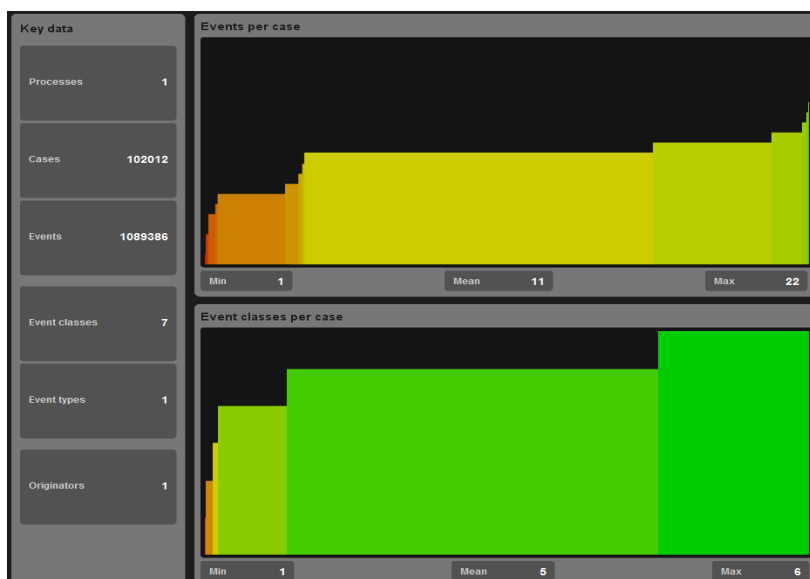
**Activity:** Activity

**Timestamp:** ActivityDate

Στην συγκεκριμένη περίπτωση θα παρακολουθήσουμε την πορεία του κάθε υποψηφίου, οπότε το CaseID είναι ο κωδικός του υποψηφίου.

### 5.1.1 Οπτικοποίηση (Visualization) πρώτου αρχείου καταγραφής συμβάντων

Όπως φαίνεται και στην παρακάτω εικόνα, όπου γίνεται μια οπτικοποίηση του πρώτου αρχείου καταγραφής συμβάντων το οποίο προκύπτει στο ProM, αυτό περιέχει 1.089.386 συμβάντα/γεγονότα τα οποία αφορούν 102.012 περιπτώσεις (αιτήσεις-δηλώσεις). Στα γεγονότα αυτά καταγράφονται 7 διαφορετικές δραστηριότητες που αναφέρθηκαν και παραπάνω (event classes).



Εικόνα 13: Οπτικοποίηση πρώτου αρχείου καταγραφής συμβάντων

Επιλέγοντας στο ProM τη λειτουργία παρουσίασης της σύνοψης των γεγονότων του πρώτου αρχείου καταγραφής συμβάντων (**Log Summary**) παρατηρούμε τα παρακάτω:

Log Summary		
Total number of process instances: <b>102012</b>		
Total number of events: <b>1089386</b>		
Event Name		
Event classes defined by Event Name		
<b>All events</b>		
Total number of classes: 7		
Class	Occurrences (absolute)	Occurrences (relative)
ΚΑΤΑΧΩΡΗΣΗ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ	407405	37,398%
ΚΑΤΑΧΩΡΗΣΗ ΒΑΘΜΟΛΟΓΙΑΣ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ	343141	31,499%
ΚΑΤΑΧΩΡΗΣΗ ΑΙΤΗΣΗΣ	102012	9,364%
ΑΠΟΔΟΧΗ ΑΙΤΗΣΗΣ	100265	9,204%
ΟΛΟΚΛΗΡΩΣΗ ΑΙΤΗΣΗΣ	100265	9,204%
ΚΑΤΑΧΩΡΗΣΗ ΕΙΔΙΚΟΥ ΜΑΘΗΜΑΤΟΣ	36160	3,319%
ΑΚΥΡΩΣΗ ΑΙΤΗΣΗΣ	138	0,013%

Εικόνα 14: Σύνοψη των γεγονότων του πρώτου αρχείου καταγραφής συμβάντων

- Για όλες τις περιπτώσεις (102.012) έχει πραγματοποιηθεί η ΚΑΤΑΧΩΡΗΣΗ ΑΙΤΗΣΗΣ στο σύστημα.
- Οι περιπτώσεις ΚΑΤΑΧΩΡΗΣΗΣ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ στην αίτηση-δήλωση και ΚΑΤΑΧΩΡΗΣΗΣ ΒΑΘΜΟΛΟΓΙΑΣ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ είναι περίπου τετραπλάσιες από αυτές της ΚΑΤΑΧΩΡΗΣΗΣ και ΟΛΟΚΛΗΡΩΣΗΣ ΑΙΤΗΣΗΣ, πράγμα το οποίο είναι λογικό καθώς για κάθε αίτηση-δήλωση αντιστοιχούν τέσσερα (4) βασικά μαθήματα της ομάδας προσανατολισμού που έχει επιλέξει ο κάθε υποψήφιος.

Start events		
Total number of classes: 1		
Class	Occurrences (absolute)	Occurrences (relative)
ΚΑΤΑΧΩΡΗΣΗ ΑΙΤΗΣΗΣ	102012	100,0%
End events		
Total number of classes: 7		
Class	Occurrences (absolute)	Occurrences (relative)
ΚΑΤΑΧΩΡΗΣΗ ΒΑΘΜΟΛΟΓΙΑΣ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ	85776	84,084%
ΟΛΟΚΛΗΡΩΣΗ ΑΙΤΗΣΗΣ	13372	13,108%
ΚΑΤΑΧΩΡΗΣΗ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ	1568	1,537%
ΑΠΟΔΟΧΗ ΑΙΤΗΣΗΣ	617	0,605%
ΚΑΤΑΧΩΡΗΣΗ ΕΙΔΙΚΟΥ ΜΑΘΗΜΑΤΟΣ	482	0,472%
ΑΚΥΡΩΣΗ ΑΙΤΗΣΗΣ	138	0,135%
ΚΑΤΑΧΩΡΗΣΗ ΑΙΤΗΣΗΣ	59	0,058%

Εικόνα 15: Σύνοψη των γεγονότων έναρξης-λήξης στο πρώτο αρχείο καταγραφής συμβάντων

- Το 84,084% των περιπτώσεων (85.776 αιτήσεις-δηλώσεις υποψηφίων) ολοκληρώθηκε με την ΚΑΤΑΧΩΡΗΣΗ ΒΑΘΜΟΛΟΓΙΑΣ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ, που σημαίνει ότι οι υποψήφιοι αυτοί εξετάστηκαν και στη συνέχεια βαθμολογήθηκαν στα μαθήματα τα οποία δήλωσαν.
- Το υπόλοιπο ποσοστό (15,916%) είναι περιπτώσεις στις οποίες οι υποψήφιοι είτε ακύρωσαν την αίτηση-δήλωση που είχαν υποβάλει είτε δεν εξετάστηκαν στα μαθήματα τα οποία είχαν δηλώσει.
- Είναι αξιοσημείωτο ότι 13.372 αιτήσεις-δηλώσεις υποψηφίων ολοκληρώθηκαν μεν στο πληροφοριακό σύστημα αλλά οι υποψήφιοι αυτοί δεν παρουσιάστηκαν να εξεταστούν. Θα ήταν ενδιαφέρον σε κάποια μελλοντική μελέτη να διερευνηθεί το προφίλ των υποψηφίων αυτών.
- Επίσης, 617 αιτήσεις-δηλώσεις υποψηφίων παρέμειναν στο στάδιο της ΑΠΟΔΟΧΗΣ ΑΙΤΗΣΗΣ που σημαίνει ότι οι υποψήφιοι αυτοί κατέθεσαν μεν την αίτηση-δήλωση στη σχολική μονάδα, ο χειριστής του πληροφοριακού συστήματος την καταχώρησε και την αποθήκευσε προσωρινά, αλλά στη συνέχεια οι υποψήφιοι δεν παρουσιάστηκαν να επιβεβαιώσουν την ορθότητα των στοιχείων που καταχωρήθηκαν προκειμένου να πραγματοποιηθεί η οριστική παραλαβή, δηλ. η ΟΛΟΚΛΗΡΩΣΗ ΑΙΤΗΣΗΣ, κατά πώς προβλέπεται από τη διαδικασία.
- Τέλος, ένα πολύ μικρό ποσοστό αιτήσεων-δηλώσεων ούτε ακυρώθηκε, ούτε ολοκληρώθηκε, ούτε έμεινε προσωρινά καταχωρημένο, αλλά παρέμειναν «ελλιπείς» στο σύστημα σταματώντας είτε στην καταχώρηση μαθημάτων είτε στην αρχική καταχώρηση της αίτησης-δήλωσης. Θεωρητικά δεν υπάρχει κάποιος λόγος για να γίνει κάτι τέτοιο, καθώς όταν ένας υποψήφιος καταθέτει μία αίτηση-δήλωση ο χειριστής του πληροφοριακού συστήματος καταχωρώντας την στο σύστημα θα πρέπει να φτάσει τουλάχιστον μέχρι το στάδιο της ΑΠΟΔΟΧΗΣ ΑΙΤΗΣΗΣ, επομένως και σε αυτή την περίπτωση θα ήταν καλό να διερευνηθεί αφενός ο λόγος για τον οποίο γίνεται κάτι τέτοιο, αφετέρου γιατί το σύστημα το επιτρέπει.

Ορισμένες πληροφορίες για τη συνολική διαδικασία και τις διαφορετικές περιπτώσεις που εμφανίζονται σε αυτή, μπορούμε να δούμε και να εξετάσουμε χρησιμοποιώντας την απεικόνιση “Explore Event Log”, με τη βοήθεια της οποίας βλέπουμε όλες τις πιθανές διαδρομές των δραστηριοτήτων της διαδικασίας καθώς και τη συχνότητα με την οποία εμφανίζεται η καθεμία.

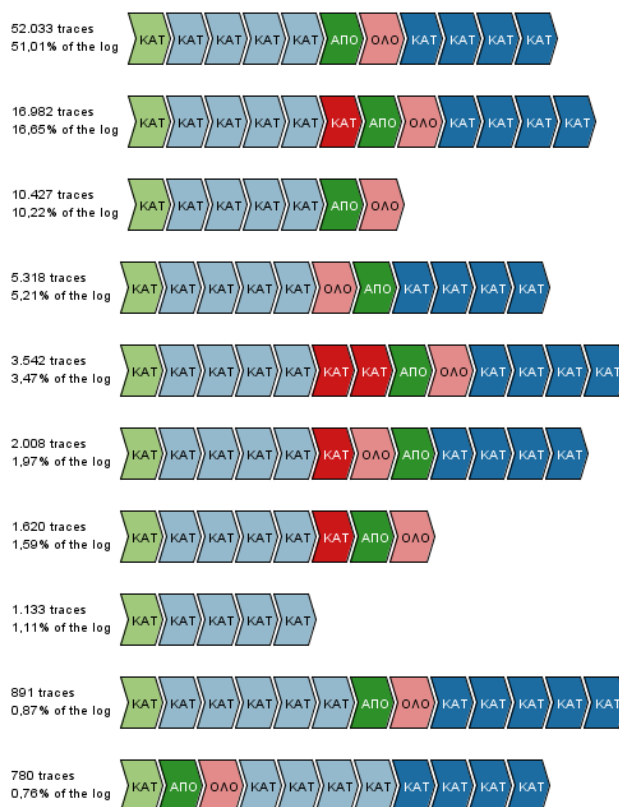
Τα στατιστικά στοιχεία που προκύπτουν επιλέγοντας την απεικόνιση “Explore Event Log” είναι τα παρακάτω:

Traces	102.012
Events	1.089.386
Event Classes	7
Attributes	9
Variants	168
Events per Trace	10,679
First Event	2020-03-30T14:45:02Z
Last Event	2022-03-15T08:25:34Z

Εικόνα 16: Στατιστικά στοιχεία απεικόνισης Explore Event Log

Φαίνεται ότι υπάρχουν 168 διαφορετικές παραλλαγές/διαδρομές στη διαδικασία που εξετάζεται, με κατά μέσο όρο 10,679 γεγονότα στην κάθε διαδρομή. Επίσης, φαίνεται ότι η ημερομηνία στην οποία έλαβε χώρα το τελευταίο γεγονός ήταν το Μάρτιο του 2022, δηλ. στο μεθεπόμενο σχολικό έτος από αυτό που εξετάζουμε, κάτι που κανονικά δεν θα έπρεπε να συμβαίνει.

Εάν επιλέξουμε να εμφανίσουμε τις διαδρομές με φθίνουσα ταξινόμηση της συχνότητας με την οποία εμφανίζονται, παίρνουμε την παρακάτω εικόνα:



Εικόνα 17: Διαδρομές δραστηριοτήτων με φθίνουσα συχνότητα (1<sup>ο</sup> event log)

Η αντιστοίχιση των παραπάνω χρωματικών απεικονίσεων στις δραστηριότητες που υπάρχουν στο αρχείο καταγραφής συμβάντων είναι η παρακάτω:



→ ΚΑΤΑΧΩΡΗΣΗ ΑΙΤΗΣΗΣ



→ ΚΑΤΑΧΩΡΗΣΗ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ



→ ΑΠΟΔΟΧΗ ΑΙΤΗΣΗΣ



→ ΟΛΟΚΛΗΡΩΣΗ ΑΙΤΗΣΗΣ



→ ΚΑΤΑΧΩΡΗΣΗ ΒΑΘΜΟΛΟΓΙΑΣ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ



→ ΚΑΤΑΧΩΡΗΣΗ ΕΙΔΙΚΟΥ ΜΑΘΗΜΑΤΟΣ



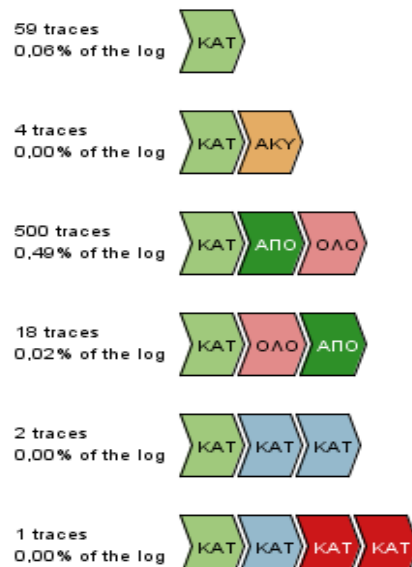
→ ΑΚΥΡΩΣΗ ΑΙΤΗΣΗΣ

Παρατηρούμε ότι το 71,13% των περιπτώσεων καταλαμβάνονται από τα δύο πρώτα μονοπάτια και από το πέμπτο μονοπάτι, τα οποία και απεικονίζουν την αναμενόμενη διαδικασία, δηλ. καταχώρηση της αίτησης, των τεσσάρων βασικών μαθημάτων (και επιπλέον ενός ειδικού στο δεύτερο μονοπάτι ή δύο ειδικών στο πέμπτο μονοπάτι), αποδοχή και ολοκλήρωση της αίτησης και τέλος καταχώρηση των βαθμολογιών των μαθημάτων.

Η τρίτη και η έβδομη σε συχνότητα διαδρομή (συνολικό ποσοστό 11,81% των περιπτώσεων) είναι οι αιτήσεις-δηλώσεις (με ή χωρίς ειδικό μάθημα) οι οποίες καταχωρήθηκαν, ολοκληρώθηκαν αλλά οι υποψήφιοι δεν προσήλθαν να εξεταστούν, ποσοστό αξιοσημείωτο όπως έχει αναφερθεί και παραπάνω.

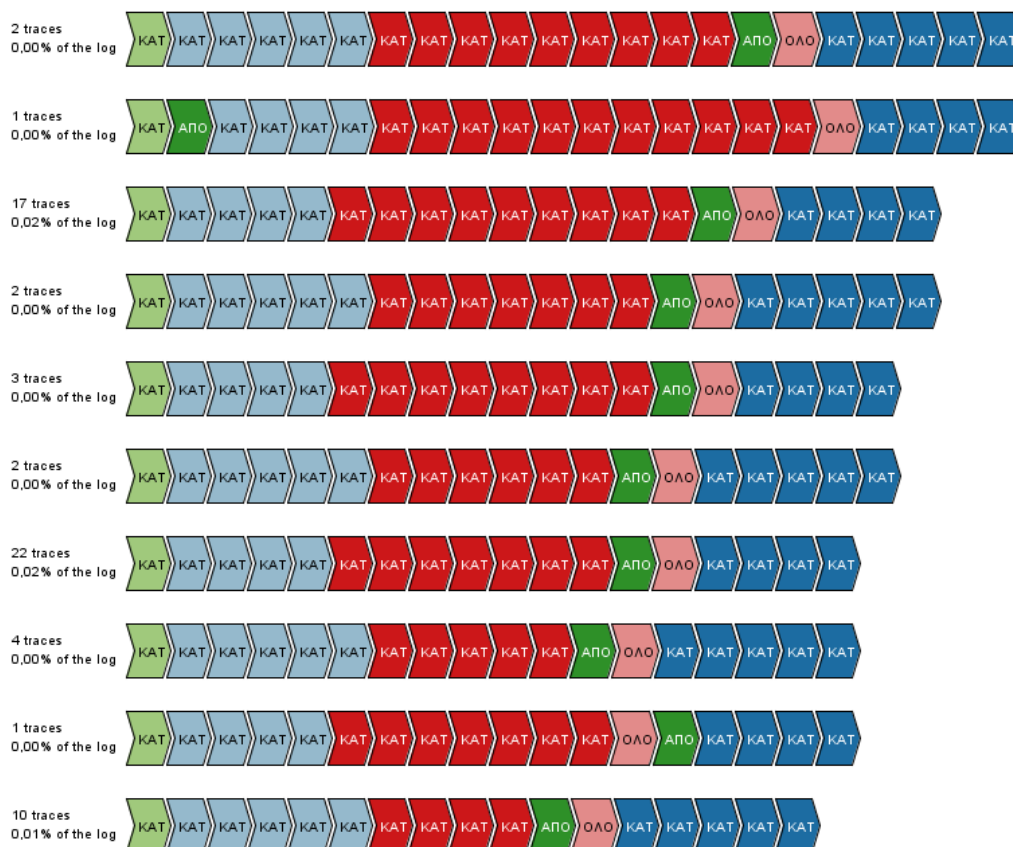
Αξιοσημείωτο είναι και το ποσοστό της τέταρτης σε συχνότητα διαδρομής (5,21% - 5.318 περιπτώσεις), στην οποία οι δραστηριότητες είναι αυτές που εμφανίζονται και στην πρώτη σε συχνότητα διαδρομή μόνο που έχει αντιστραφεί η σειρά της ΚΑΤΑΧΩΡΗΣΗΣ και της ΟΛΟΚΛΗΡΩΣΗΣ της αίτησης-δήλωσης. Παρόλο που το αποτέλεσμα δεν αλλάζει, θα ήταν καλό να διερευνηθεί και στην περίπτωση αυτή ο λόγος για τον οποίο το σύστημα επιτρέπει κάτι τέτοιο, καθώς με τον τρόπο αυτό δημιουργούνται περιθώρια για σφάλματα στη συνολική διαδικασία.

Εάν επιλέξουμε να εμφανίσουμε τις διαδρομές με αύξουσα και με φθίνουσα ταξινόμηση του μήκους τους, παίρνουμε αντίστοιχα τις δύο παρακάτω εικόνες:



Εικόνα 18: Διαδρομές δραστηριοτήτων με αύξον μήκος μονοπατιού (1<sup>ο</sup> event log)

Αξιοσημείωτη είναι η τρίτη μικρότερη διαδρομή, όπου για 500 υποψήφιους πραγματοποιήθηκε αίτηση-δήλωση χωρίς κανένα μάθημα μέσα. Οι υπόλοιπες μικρές διαδρομές αφορούν ελάχιστες περιπτώσεις στις οποίες δεν αξίζει να εστιάσουμε.



Εικόνα 19: Διαδρομές δραστηριοτήτων με φθίνον μήκος μονοπατιού (1<sup>ο</sup> event log)

Όσον αφορά στις μεγαλύτερες διαδρομές, όπως είναι αναμενόμενο αυτές εμφανίζονται όταν ορισμένοι υποψήφιοι κατά την αίτηση-δήλωσή τους επιλέγουν να δηλώσουν σχεδόν όλα τα ειδικά μαθήματα που είναι διαθέσιμα, κάτι που σπάνια ανταποκρίνεται στην πραγματική επιθυμία του υποψηφίου. Επειδή η οργάνωση των εξετάσεων των ειδικών μαθημάτων (εξεταστικά κέντρα, αίθουσες, επιτηρητές, σύσταση επιτροπών) στηρίζεται στα δεδομένα των αιτήσεων-δηλώσεων, θα άξιζε να γίνει κάποια διερεύνηση σε περίπτωση που οι περιπτώσεις αυτές ήταν περισσότερες. Επειδή πρόκειται για ελάχιστες και μεμονωμένες περιπτώσεις και εδώ, δεν έχει νόημα να εστιάσουμε την ανάλυσή μας σε αυτό το σημείο.

### 5.1.2 Οπτικοποίηση (Visualization) δεύτερου αρχείου καταγραφής συμβάντων

Αντίστοιχα, χρησιμοποιώντας τα εργαλεία απεικόνισης στο δεύτερο αρχείο καταγραφής συμβάντων που δημιουργήθηκε, το οποίο επιπλέον των υπολοίπων δεδομένων περιέχει και δεδομένα της καταχώρησης των βαθμολογιών των βασικών μαθημάτων στα βαθμολογικά κέντρα (BK), παρατηρούμε τα παρακάτω:

Log Summary		
Total number of process instances: 102747		
Total number of events: 1488554		
Event Name		
Event classes defined by Event Name		
All events		
Total number of classes: 8		
Class	Occurrences (absolute)	Occurrences (relative)
ΚΑΤΑΧΩΡΗΣΗ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ	407405	27,369%
ΟΛΟΚΛΗΡΩΣΗ ΒΑΘΜΟΛΟΓΙΑΣ ΜΑΘΗΜΑΤΟΣ ΣΤΟ ΒΚ	399168	26,816%
ΚΑΤΑΧΩΡΗΣΗ ΒΑΘΜΟΛΟΓΙΑΣ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ	343141	23,052%
ΚΑΤΑΧΩΡΗΣΗ ΑΙΤΗΣΗΣ	102012	6,853%
ΑΠΟΔΟΧΗ ΑΙΤΗΣΗΣ	100265	6,736%
ΟΛΟΚΛΗΡΩΣΗ ΑΙΤΗΣΗΣ	100265	6,736%
ΚΑΤΑΧΩΡΗΣΗ ΕΙΔΙΚΟΥ ΜΑΘΗΜΑΤΟΣ	36160	2,429%
ΑΚΥΡΩΣΗ ΑΙΤΗΣΗΣ	138	0,009%

Εικόνα 20: Σύνοψη των γεγονότων του δεύτερου αρχείου καταγραφής συμβάντων

Αρχικά, δεν παρατηρείται κάτι ιδιαίτερα διαφορετικό σε σχέση με τη σύνοψη του πρώτου αρχείου καταγραφής συμβάντων.

- Υπάρχει μία επιπλέον δραστηριότητα (event class), αυτή της ΟΛΟΚΛΗΡΩΣΗΣ ΒΑΘΜΟΛΟΓΙΑΣ ΜΑΘΗΜΑΤΟΣ ΣΤΟ ΒΚ, η οποία αφορά λιγότερες περιπτώσεις από την ΚΑΤΑΧΩΡΗΣΗ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ (οι υποψήφιοι που δηλώνουν το μάθημα αλλά δεν εξετάζονται), αλλά περισσότερες περιπτώσεις από την ΚΑΤΑΧΩΡΗΣΗ ΒΑΘΜΟΛΟΓΙΑΣ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ στο σύστημα, κάτι που πιθανό οφείλεται στις αναβαθμολογήσεις που γίνονται στα ΒΚ όταν υπάρχει μεγάλη διαφορά στους δύο αρχικούς βαθμούς.

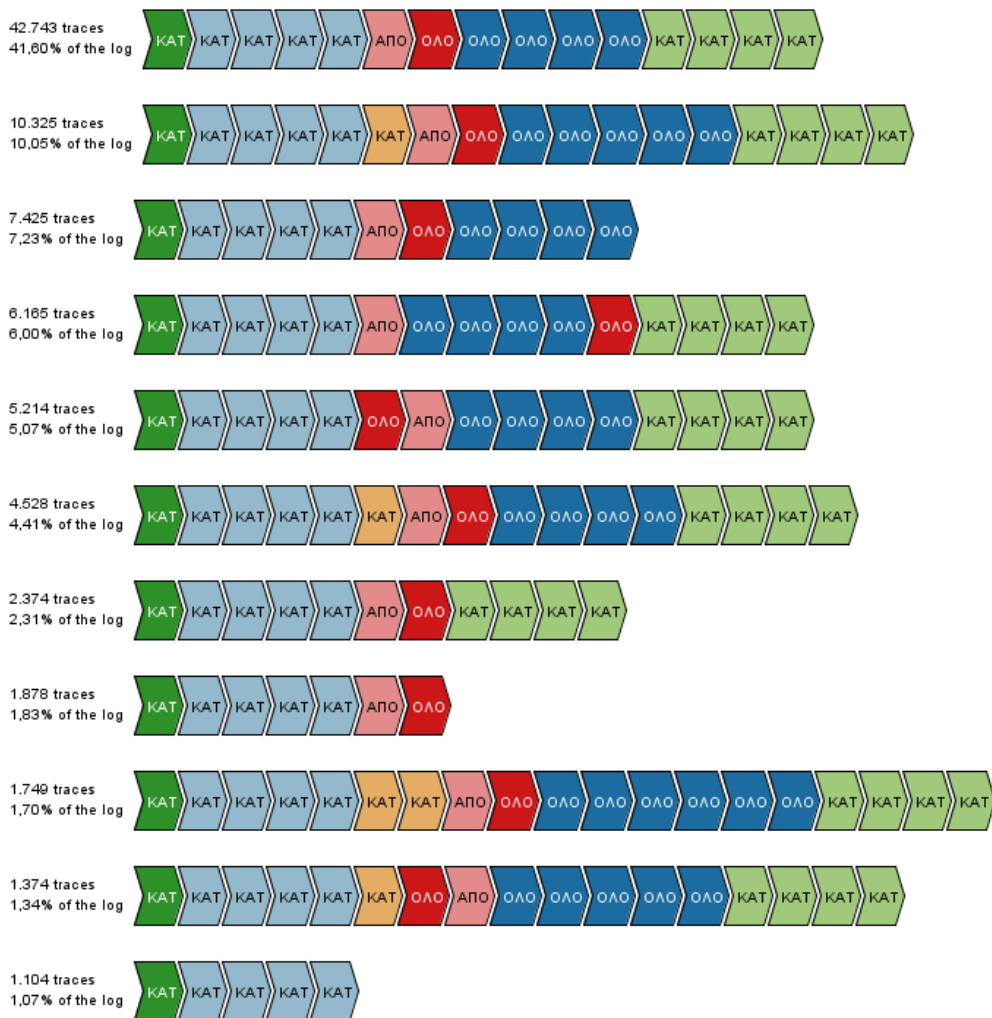


Start events		
Total number of classes: 2		
Class	Occurrences (absolute)	Occurrences (relative)
ΚΑΤΑΧΩΡΗΣΗ ΑΙΤΗΣΗΣ	101903	99,179%
ΟΛΟΚΛΗΡΩΣΗ ΒΑΘΜΟΛΟΓΙΑΣ ΜΑΘΗΜΑΤΟΣ ΣΤΟ ΒΚ	844	0,821%
End events		
Total number of classes: 8		
Class	Occurrences (absolute)	Occurrences (relative)
ΚΑΤΑΧΩΡΗΣΗ ΒΑΘΜΟΛΟΓΙΑΣ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ	85333	83,052%
ΟΛΟΚΛΗΡΩΣΗ ΒΑΘΜΟΛΟΓΙΑΣ ΜΑΘΗΜΑΤΟΣ ΣΤΟ ΒΚ	11930	11,611%
ΟΛΟΚΛΗΡΩΣΗ ΑΙΤΗΣΗΣ	3604	3,508%
ΚΑΤΑΧΩΡΗΣΗ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ	1255	1,221%
ΚΑΤΑΧΩΡΗΣΗ ΕΙΔΙΚΟΥ ΜΑΘΗΜΑΤΟΣ	336	0,327%
ΑΠΟΔΟΧΗ ΑΙΤΗΣΗΣ	163	0,159%
ΑΚΥΡΩΣΗ ΑΙΤΗΣΗΣ	97	0,094%
ΚΑΤΑΧΩΡΗΣΗ ΑΙΤΗΣΗΣ	29	0,028%

Εικόνα 21: Σύνοψη των γεγονότων έναρξης-λήξης στο δεύτερο αρχείο καταγραφής συμβάντων

- Ένα σημαντικό ποσοστό, το 11,611% των περιπτώσεων (11.930 περιπτώσεις) δείχνουν να ολοκληρώνονται με τη βαθμολόγηση στα ΒΚ, κάτι που είναι αδύνατο καθώς οι βαθμολογίες των ΒΚ μεταφέρονται στη συνέχεια στο πληροφοριακό σύστημα. Επίσης, σε 844 περιπτώσεις δείχνει να ξεκινάει η διαδικασία από την ΟΛΟΚΛΗΡΩΣΗ ΒΑΘΜΟΛΟΓΙΑΣ ΣΤΑ ΒΚ, κάτι που είναι επίσης αδύνατο. Επειδή η πληροφορία για τη βαθμολόγηση στα ΒΚ προέρχεται από άλλο αρχείο που εξήχθη από διαφορετικό πληροφοριακό σύστημα, αυτή η πιθανή ασυνέπεια των χρονικών σημείων των δύο πληροφοριακών συστημάτων θα ήταν καλό να διερευνηθεί, καθώς δημιουργεί λανθασμένη πληροφορία η οποία όμως λαμβάνεται υπόψη κατά την εξόρυξη της διαδικασίας και την ανάλυση.

Τέλος, εάν επιλέξουμε την απεικόνιση “Explore Event Log” και δούμε όλες τις πιθανές διαδρομές των δραστηριοτήτων της διαδικασίας με φθίνουσα συχνότητα εμφάνισης, παίρνουμε την παρακάτω εικόνα:



Εικόνα 22: Διαδρομές δραστηριοτήτων με φθίνουσα συχνότητα (2ο event log)

Στην παραπάνω εικόνα αυτό που χρήζει διερεύνησης είναι η παραπάνω παρατήρηση, όπου δηλαδή για 7.425 περιπτώσεις (7,23%) δείχνει να υπάρχει καταχώρηση βαθμολογίας στα ΒΚ αλλά όχι στο σύστημα και για 2.374 περιπτώσεις (2,31%) δείχνει να υπάρχει καταχώρηση βαθμολογίας στο σύστημα χωρίς να έχει προηγηθεί βαθμολόγηση στα ΒΚ.

## 5.2 Εξερεύνηση του πρώτου αρχείου καταγραφής συμβάντων με το εργαλείο γραφήματος Dotted Chart

Το Dotted Chart αποτελεί ένα ισχυρό οπτικό εργαλείο του ProM, στο οποίο απεικονίζονται σε μορφή γραφήματος τα γεγονότα συναρτήσει του χρόνου στον οποίο έλαβαν χώρα και βοηθά σε μια αρχική και γρήγορη επισκόπηση και κατανόηση του αρχείου καταγραφής συμβάντων ώστε κατά την αξιολόγηση των αποτελεσμάτων να λαμβάνουμε πληροφορίες σχετικά με τον τρόπο περαιτέρω επεξεργασίας των δεδομένων μας.

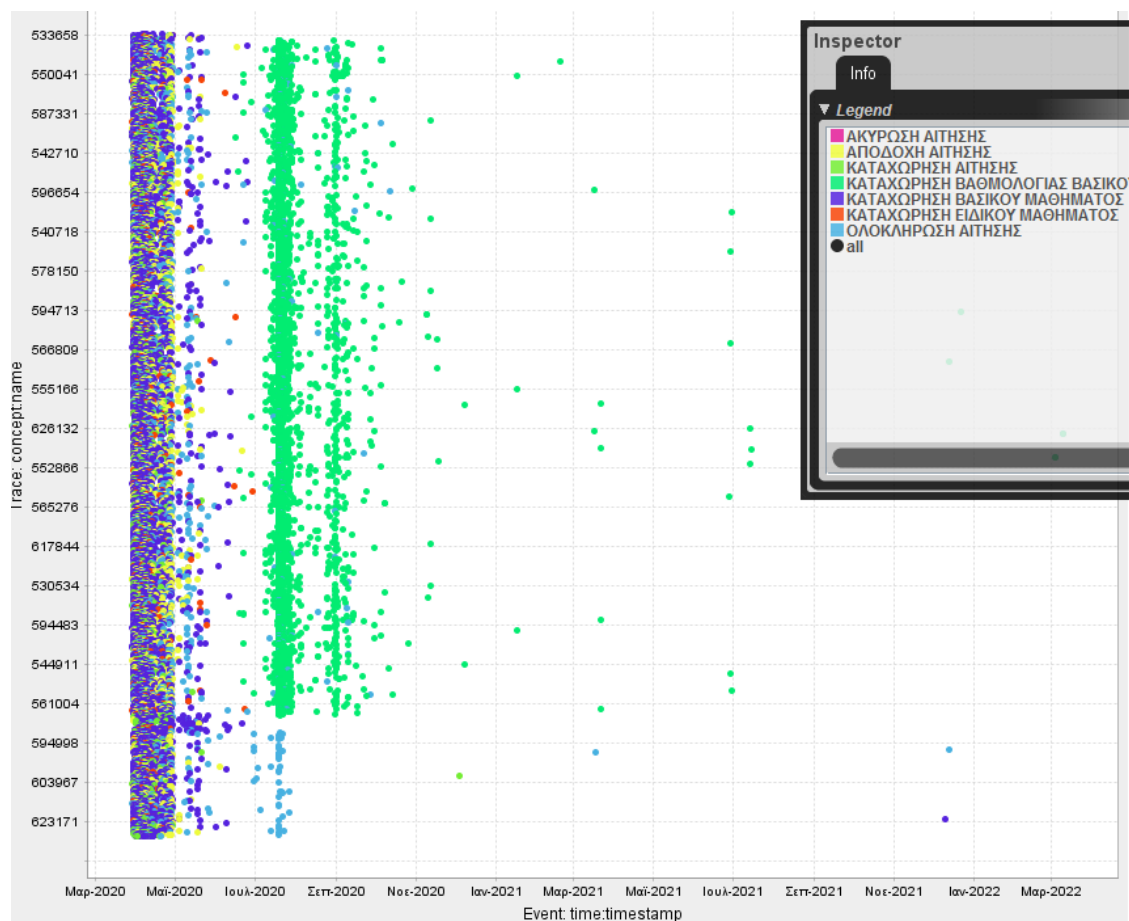
Κάνοντας χρήση του πρόσθετου “Project Log on DottedChart” βλέπουμε στην παρακάτω εικόνα την αρχική απεικόνιση του Dotted Chart.



Εικόνα 23: Αρχική απεικόνιση του αρχείου καταγραφής συμβάντων με τη χρήση του Dotted Chart

Στην αρχική αυτή απεικόνιση, παρατηρούμε δύο μεγάλες συμπυκνώσεις γεγονότων, μία την περίοδο Απριλίου-Μαΐου 2020 και μία την περίοδο Ιουλίου-Αυγούστου 2020, καθώς και μια μικρότερη συμπύκνωση την περίοδο του Σεπτεμβρίου 2020. Για το σχολικό έτος 2019-2020 για το οποίο εξετάζουμε τα δεδομένα, αυτές οι συμπυκνώσεις ανταποκρίνονται στις αναμενόμενες, καθώς για εκείνο το σχολικό έτος, λόγω της πανδημίας COVID-19, η καταχώρηση των αιτήσεων-δηλώσεων στο σύστημα MySchool πραγματοποιήθηκε την περίοδο Απριλίου-Μαΐου και η καταχώρηση των βαθμολογιών την περίοδο Ιουλίου-Αυγούστου. Η συγκέντρωση των γεγονότων την περίοδο του Σεπτεμβρίου αφορά στις καταχωρήσεις βαθμολογιών των υποψηφίων των Επαναληπτικών Πανελλαδικών Εξετάσεων.

Κατά τη μελέτη του Dotted Chart συστήνεται να θέτουμε την παράμετρο του χρώματος στο concept:name έτσι ώστε τα συμβάντα να έχουν διαφορετικό χρώμα για κάθε δραστηριότητα, όπως φαίνεται στην παρακάτω εικόνα:



Εικόνα 24: Απεικόνιση αρχείου καταγραφής συμβάντων με Dotted Chart, ανά CaseID (ApplicationID) με χρήση διαφορετικού χρώματος ανά συμβάν

Παρατηρώντας την παραπάνω εικόνα βλέπουμε ότι τα περισσότερα συμβάντα λαμβάνουν χώρα την περίοδο Απριλίου-Μαΐου (ΚΑΤΑΧΩΡΗΣΗ ΑΙΤΗΣΗΣ, ΚΑΤΑΧΩΡΗΣΗ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ, ΚΑΤΑΧΩΡΗΣΗ ΕΙΔΙΚΟΥ ΜΑΘΗΜΑΤΟΣ, ΑΚΥΡΩΣΗ ΑΙΤΗΣΗΣ, ΑΠΟΔΟΧΗ ΑΙΤΗΣΗΣ, ΟΛΟΚΛΗΡΩΣΗ ΑΙΤΗΣΗΣ) και μόνο η ΚΑΤΑΧΩΡΗΣΗ ΒΑΘΜΟΛΟΓΙΑΣ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ λαμβάνει χώρα τις περιόδους Ιουλίου-Αυγούστου και Σεπτεμβρίου.

Ωστόσο, υπάρχουν και συμβάντα που αφορούν τις πρώτες διαδικασίες που λαμβάνουν χώρα το Σεπτέμβριο, το οποίο σημαίνει ότι έχει πραγματοποιηθεί τροποποίηση της αίτησης-δήλωσης του υποψηφίου στο σύστημα Myschool. Επίσης, παρατηρούνται συμβάντα που αφορούν κυρίως στην καταχώρηση βαθμολογίας βασικού μαθήματος τα οποία εμφανίζονται στο επόμενο και ακόμη και στο μεθεπόμενο σχολικό έτος. Λαμβάνοντας υπόψη ότι δεν προβλέπεται από τη νομοθεσία η δυνατότητα τροποποίησης της βαθμολογίας ενός υποψηφίου, θα ήταν ενδιαφέρον οι περιπτώσεις αυτές να διερευνηθούν.

Σε μια δειγματοληπτική έρευνα που έγινε χρησιμοποιώντας το εργαλείο Instructor του ProM με το οποίο δίνεται η δυνατότητα να μελετάται αναλυτικά η κάθε περίπτωση ξεχωριστά, φαίνεται μία περίπτωση όπου η καταχώρηση μίας αίτησης-δήλωσης που αφορά στο σχολικό έτος 2019-2020, το οποίο και μελετάμε, έλαβε χώρα στις 13-12-

2021 (Εικόνα 18), καθώς και μία περίπτωση όπου η καταχώρηση των βαθμολογιών των βασικών μαθημάτων έλαβε χώρα στις 01-09-2020, ημερομηνία που δεν είχαν ακόμη πραγματοποιηθεί οι επαναληπτικές πανελλαδικές εξετάσεις του έτους 2019-2020 (Εικόνα 19). Η πρώτη προβληματική περίπτωση πιθανόν οφείλεται στην έλλειψη ελέγχου στο σύστημα MySchool, όπου επιτρέπεται να γίνει καταχώρηση αίτησης-δήλωσης σε προγενέστερο σχολικό έτος εάν αυτό για κάποιο λόγο έχει αλλάξει προσωρινά από το χειριστή του συστήματος για να δει π.χ. κάποια στοιχεία του προηγούμενου έτους και έχει ξεχαστεί να γίνει επαναφορά του στο παρόν σχολικό έτος.

831091 7 events
ΚΑΤΑΧΩΡΗΣΗ ΑΙΤΗΣΗΣ #1 complete 13.12.2021 20:19:12.530
ΚΑΤΑΧΩΡΗΣΗ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ #2 complete 13.12.2021 20:19:43.263
ΚΑΤΑΧΩΡΗΣΗ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ #3 complete 13.12.2021 20:19:43.280
ΚΑΤΑΧΩΡΗΣΗ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ #4 complete 13.12.2021 20:19:43.280
ΚΑΤΑΧΩΡΗΣΗ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ #5 complete 13.12.2021 20:19:43.280
ΑΠΟΔΟΧΗ ΑΙΤΗΣΗΣ #6 complete 13.12.2021 20:20:34.460
ΟΛΟΚΛΗΡΩΣΗ ΑΙΤΗΣΗΣ #7 complete 13.12.2021 20:20:34.797

Εικόνα 25: Μελέτη προβληματικής αίτησης-δήλωσης I

619057 11 events
ΚΑΤΑΧΩΡΗΣΗ ΑΙΤΗΣΗΣ #1 complete 15.04.2020 18:47:57.370
ΚΑΤΑΧΩΡΗΣΗ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ #2 complete 15.04.2020 18:48:31.730
ΚΑΤΑΧΩΡΗΣΗ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ #3 complete 15.04.2020 18:48:31.747
ΚΑΤΑΧΩΡΗΣΗ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ #4 complete 15.04.2020 18:48:31.747
ΚΑΤΑΧΩΡΗΣΗ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ #5 complete 15.04.2020 18:48:31.747
ΑΠΟΔΟΧΗ ΑΙΤΗΣΗΣ #6 complete 15.04.2020 18:49:43.667
ΟΛΟΚΛΗΡΩΣΗ ΑΙΤΗΣΗΣ #7 complete 17.04.2020 13:20:33.120
ΚΑΤΑΧΩΡΗΣΗ ΒΑΘΜΟΛΟΓΙΑΣ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ #8 complete 01.09.2020 10:15:26.463
ΚΑΤΑΧΩΡΗΣΗ ΒΑΘΜΟΛΟΓΙΑΣ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ #9 complete 01.09.2020 10:15:26.463
ΚΑΤΑΧΩΡΗΣΗ ΒΑΘΜΟΛΟΓΙΑΣ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ #10 complete 01.09.2020 10:15:26.463
ΚΑΤΑΧΩΡΗΣΗ ΒΑΘΜΟΛΟΓΙΑΣ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ #11 complete 01.09.2020 10:15:26.480

Εικόνα 26: Μελέτη προβληματικής αίτησης-δήλωσης II

Από την παραπάνω οπτική ανάλυση που παρέχουν τα εργαλεία του ProM, είναι σαφές πως αρκετές φορές υπάρχουν ορισμένες ασάφειες ή αμφισημίες στα δεδομένα, αρκετά συνηθισμένες όταν γίνεται ανάλυση των δεδομένων και είναι καλό αυτές να γνωστοποιούνται προκειμένου να επιλυθούν στο μέλλον κι έτσι τα μελλοντικά σύνολα δεδομένων να παρέχουν τη δυνατότητα καλύτερης και ορθότερης ανάλυσης.

### 5.3 Ανακάλυψη Διαδικασιών (Process Discovery) μέσω της εκτέλεσης αλγορίθμων εξόρυξης διαδικασιών

Προχωρώντας στην ανακάλυψη της διαδικασίας (process discovery), πραγματοποιείται αρχικά εκτέλεση διαφόρων αλγορίθμων που παρέχουν τα λογισμικά ProM και Disco με τα δεδομένα των αρχείων καταγραφής συμβάντων και στη συνέχεια γίνεται μια αποτίμηση των αποτελεσμάτων τους. Οι αλγόριθμοι αυτοί αφορούν στο πώς γίνεται η μετάβαση από τα δεδομένα του αρχείου καταγραφής συμβάντων (event data) σε ένα μοντέλο διεργασιών (process model) το οποίο περιγράφει και απεικονίζει τη συμπεριφορά που εμπεριέχεται στα δεδομένα αυτά.

Το μοντέλο διεργασιών που προκύπτει από τους αλγορίθμους του ProM συνήθως περιγράφεται χρησιμοποιώντας τη σημειολογία ενός Petri-Net. Στο Petri-Net, όλες οι δραστηριότητες ή οι ενέργειες που εμπεριέχονται σε μια διαδικασία μοντελοποιούνται ως τετράγωνα. Τα τετράγωνα αυτά συνδέονται μεταξύ τους με κύκλους μέσω βελών, οι οποίοι κύκλοι αναπαριστούν τις μεταβάσεις από τη μια δραστηριότητα στην άλλη. Τέλος, ένα σημαντικό στοιχείο ενός Petri-net είναι το token, το οποίο είναι στην ουσία ένα σύμβολο που τοποθετείται στην αρχή, ακολουθεί την πορεία του κάθε ίχνους δραστηριοτήτων της διεργασίας και εάν το μοντέλο είναι σωστό καταλήγει στο τέλος.

Για το process discovery στην παρούσα εργασία, θα εκτελεστούν οι αλγόριθμοι Alpha Miner, Heuristics Miner, Inductive Miner και Fuzzy Miner του ProM και στη συνέχεια θα χρησιμοποιηθεί και το εργαλείο Disco. Προκειμένου να αξιολογηθούν τα αποτελέσματα του κάθε αλγορίθμου θα πρέπει αρχικά να εξεταστεί κατά πόσο το μοντέλο διεργασιών που έχει παραχθεί είναι sound, χωρίς σφάλματα, δηλαδή το μοντέλο ολοκληρώνεται σωστά και το token σε κάθε περίπτωση ξεκινάει από την αρχή και καταλήγει στο τέλος. Στη συνέχεια, θα πρέπει να γίνει σύγκριση του μοντέλου που παράχθηκε με το αρχείο καταγραφής συμβάντων, εξετάζοντας τις τέσσερις (4) ποιοτικές παραμέτρους της καταλληλότητας, της απλότητας, της ακρίβειας και της γενίκευσης, οι οποίες όπως έχει αναφερθεί είναι αντικρουόμενες μεταξύ τους και συνήθως όταν προσπαθούμε να βελτιώσουμε κάποια από αυτές, μειώνεται η ποιότητα σε τουλάχιστον μία άλλη παράμετρο.

Η καταλληλότητα (fitness) είναι η κύρια παράμετρος, που εκτιμά κατά πόσο τα δεδομένα μπορούν να αναπαραχθούν κατά το μεγαλύτερο μέρος τους από το μοντέλο διεργασιών που έχει δημιουργηθεί. Η απλότητα (simplicity) αφορά στο βαθμό πολυπλοκότητας που θα έχει το μοντέλο και ο στόχος είναι να είναι εύκολα κατανοητό το μοντέλο που έχει παραχθεί για να μπορεί και να ερμηνευτεί. Η ακρίβεια (precision) είναι επίσης μία σημαντική παράμετρος, διότι δεν επιθυμούμε το μοντέλο διεργασιών να επιτρέπει κάθε συμπεριφορά, γιατί τότε αυτό θα σημαίνει ότι είναι πολύ χαλαρό (underfitting). Τέλος η παράμετρος της γενίκευσης (generalization) υποδηλώνει το κατά πόσο ανταποκρίνεται το παραγόμενο μοντέλο και σε τυχόν μελλοντικά δεδομένα κι όχι μόνο σε αυτά που περιέχει το συγκεκριμένο αρχείο καταγραφής συμβάντων (overfitting).

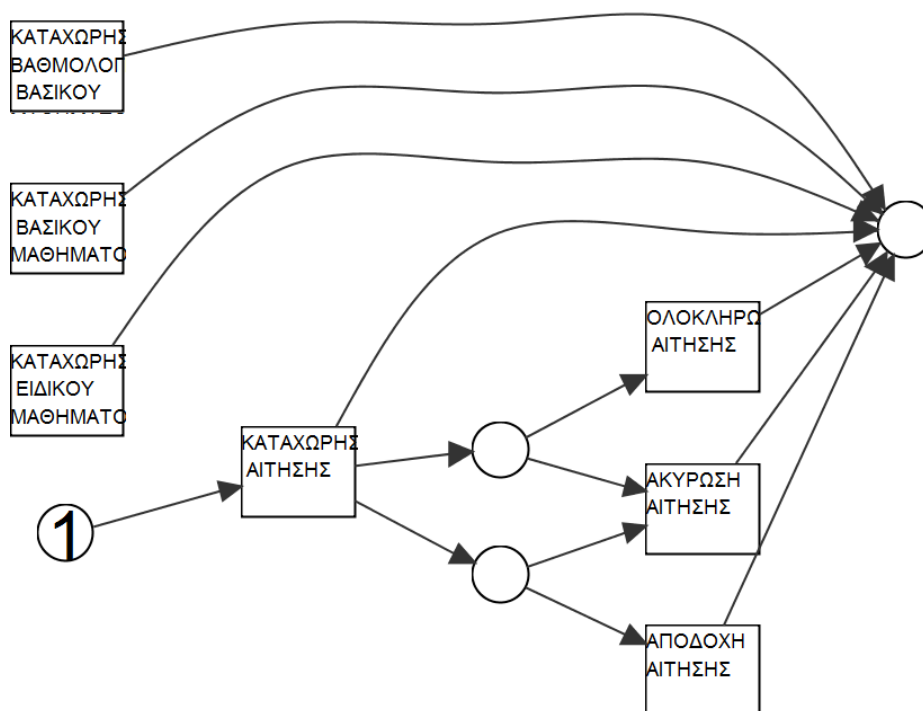
Ο στόχος είναι να κρατηθεί μια ισορροπία σ' αυτές τις αντικρουόμενες παραμέτρους. Συνήθως δεν μπορούμε να βελτιστοποιήσουμε και τις 4 ταυτόχρονα, επομένως ο καθένας πρέπει να θέσει τις προτεραιότητές του σχετικά με αυτό το θέμα.

### 5.3.1 Alpha Miner

Ο Alpha Miner είναι ο πρώτος αλγόριθμος ο οποίος, στο πλαίσιο της εξόρυξης διαδικασιών, γεφυρώνει το κενό μεταξύ των δεδομένων που βρίσκονται στα αρχεία καταγραφής συμβάντων και της ανακάλυψης ενός μοντέλου διεργασιών. Δεν εγγυάται την παραγωγή sound μοντέλων διεργασιών και δεν μπορεί να διαχειριστεί το θόρυβο στα δεδομένα.

Για την εφαρμογή του Alpha Miner πάνω στο αρχείο καταγραφής συμβάντων, επιλέγεται το πρόσθετο "Alpha Miner".

Το Petri-Net που προκύπτει από την αρχική έκδοση του Alpha είναι το παρακάτω:

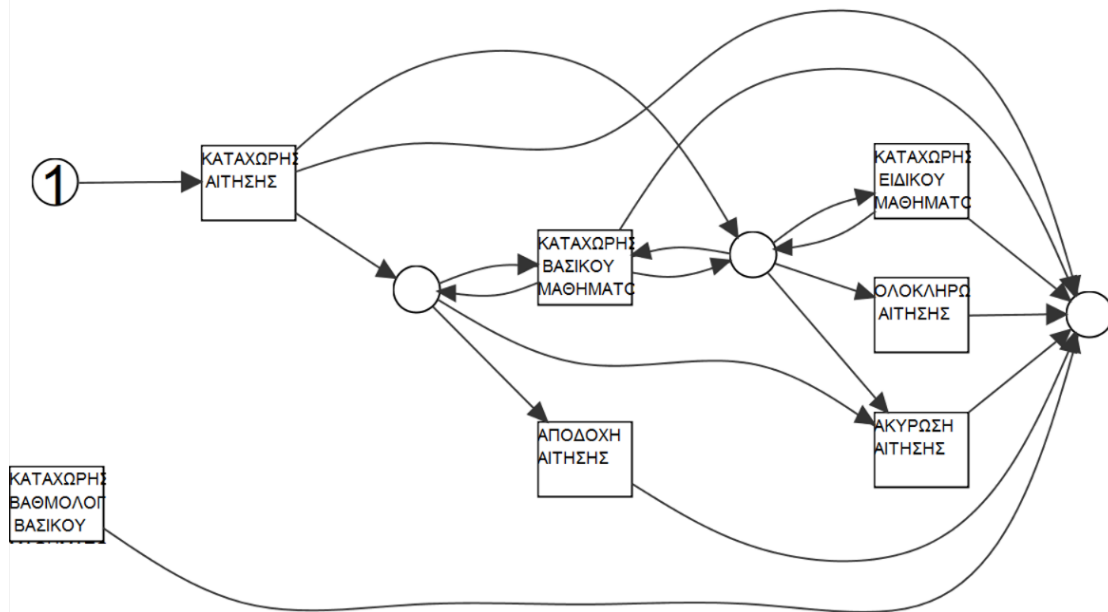


Εικόνα 27: Κλασική έκδοση Alpha Miner

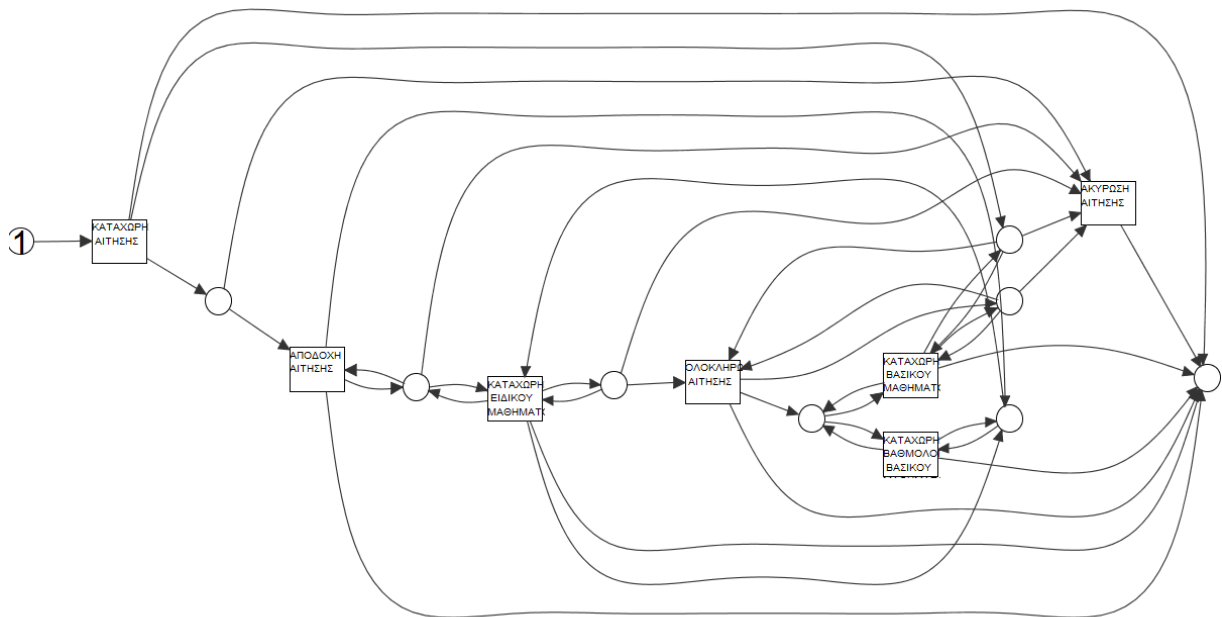
Παρατηρούμε ότι το μοντέλο ικανοποιεί το κριτήριο της απλότητας αλλά δεν απεικονίζονται σωστά οι συνδέσεις μεταξύ των δραστηριοτήτων, επομένως το μοντέλο σίγουρα δεν ικανοποιεί το κριτήριο της καταλληλότητας.

Εάν εφαρμόσουμε και τις άλλες εκδόσεις που παρέχει ο αλγόριθμος Alpha Miner, το μοντέλο που προκύπτει είναι ελάχιστα πιο αντιπροσωπευτικό της διαδικασίας, αλλά

και πάλι δεν ικανοποιούνται τα κριτήρια βασικών παραμέτρων. Ενδεικτικά παρουσιάζονται τα Petri-Net που προκύπτουν από την εφαρμογή των εκδόσεων Alpha+ και Alpha++.



Εικόνα 28: Έκδοση Alpha+ του Alpha Miner

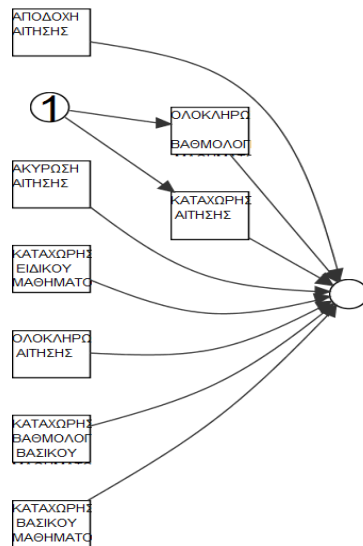


Εικόνα 29: Έκδοση Alpha++ του Alpha Miner

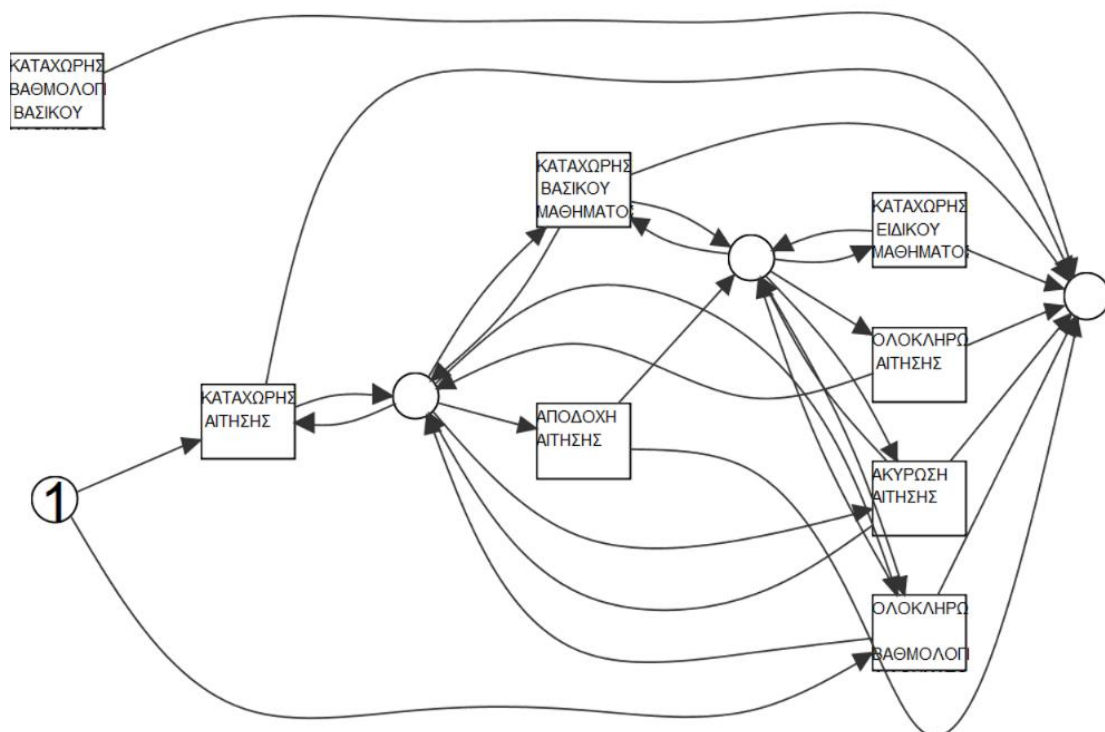
Ακόμη πιο προβληματικά είναι τα αποτελέσματα της εφαρμογής του Alpha Miner στο αρχείο καταγραφής συμβάντων που περιέχει και την πληροφορία της Ολοκλήρωσης Βαθμολογίας Βασικού Μαθήματος στα ΒΚ.



Παρουσιάζονται παρακάτω τα αποτελέσματα των εκδόσεων Alpha και Alpha+, όπου η πρώτη δεν καταφέρνει να βρει τις σωστές συνδέσεις μεταξύ των δραστηριοτήτων και η δεύτερη απεικονίζει ένα μοντέλο με πολλές συνδέσεις που δεν είναι ακριβές και είναι αρκετά γενικό.



Εικόνα 30: Κλασική έκδοση Alpha Miner στο event log με στοιχεία από τα ΒΚ

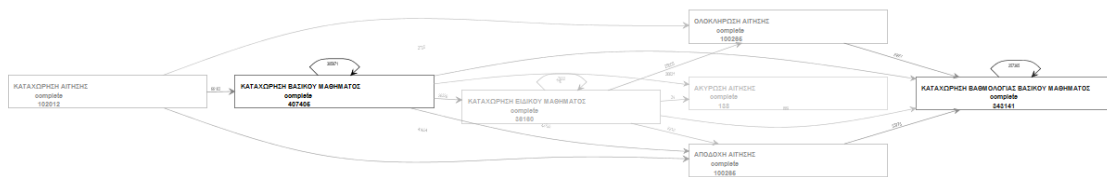


Εικόνα 31: Έκδοση Alpha+ του Alpha Miner στο event log με στοιχεία από τα ΒΚ

### 5.3.2 Heuristics Miner

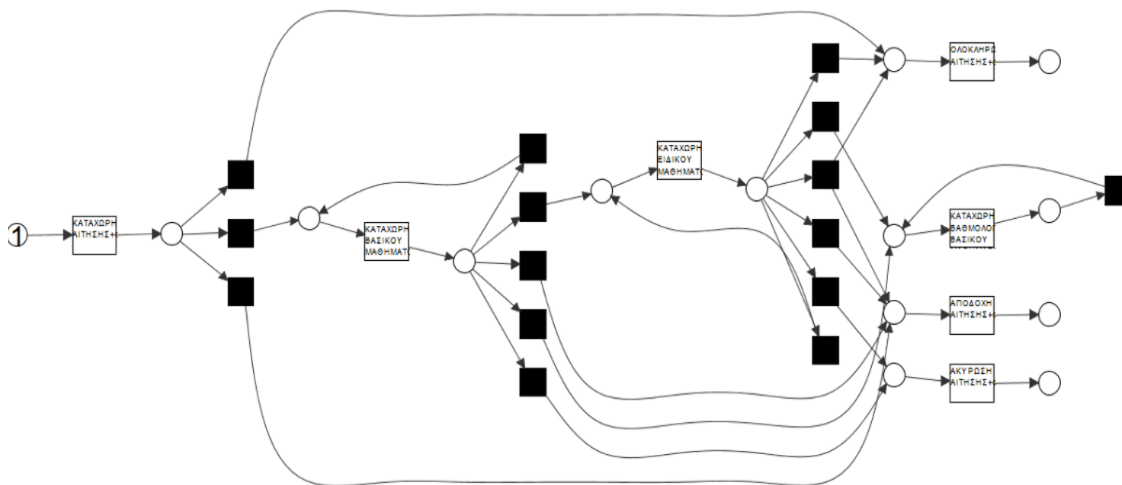
Ο αλγόριθμος Heuristics Miner είναι στην πραγματικότητα μια βελτιωμένη έκδοση του Alpha Miner. Η βασική του βελτίωση είναι ότι λαμβάνει υπόψη τις συχνότητες με τις οποίες εμφανίζονται τα ίχνη των δραστηριοτήτων στο αρχείο καταγραφής συμβάντων κι επομένως έχει τη δυνατότητα να φιλτράρει μια μη-συνήθη συμπεριφορά και το θόρυβο.

Για την εφαρμογή του Heuristics Miner πάνω στα δύο αρχεία καταγραφής συμβάντων, επιλέγεται το πρόσθετο “Mine for a Heuristics Net using Heuristics Miner”, με την εφαρμογή του οποίου δεν προκύπτει petri-net, αλλά το παρακάτω μοντέλο (με fitness 0,6466):



Εικόνα 32: Αρχικό μοντέλο Heuristics Miner

Στο μοντέλο αυτό εφαρμόζεται το πρόσθετο “Convert Heuristics Net into Petri net” και προκύπτει το παρακάτω petri-net, το οποίο μοντελοποιεί το πρώτο αρχείο καταγραφής συμβάντων:

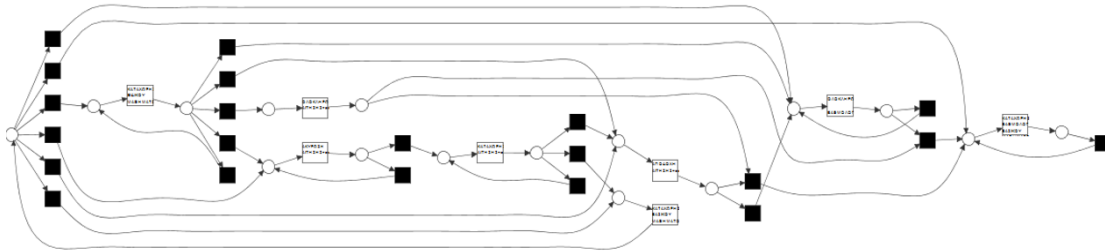


Εικόνα 33: Heuristics Miner (petri-net)

Το μοντέλο αυτό δείχνει πιο λογικό και αντιπροσωπευτικό από αυτά που προέκυπταν από την εφαρμογή του Alpha Miner όσον αφορά στη σειρά και στις συνδέσεις μεταξύ των δραστηριοτήτων, οπότε έχει αυξηθεί ο βαθμός της καταλληλότητας. Ωστόσο το μοντέλο δεν ολοκληρώνεται σωστά όπως φαίνεται στους τελευταίους κόμβους, μετά από την εκτέλεση των οποίων δεν προχωράει σε κάποια άλλη δραστηριότητα.

Επίσης, ο αλγόριθμος αυτός δημιουργεί τους λεγόμενους silent nodes, δηλαδή κόμβους που απεικονίζονται ως μαύρα τετράγωνα και δεν περιέχουν κάποια δραστηριότητα που εκτελείται αλλά χρησιμοποιούνται απλά για τη δρομολόγηση. Η ύπαρξη των κόμβων

αυτών αυξάνει την πολυπλοκότητα του μοντέλου και μιλάμε επομένως για μείωση της ποιοτικής παραμέτρου της απλότητας εφόσον δεν φαίνεται ξεκάθαρα ποια είναι η πορεία του token και γενικώς δεν είναι εμφανής η όλη διαδικασία. Τέλος, εφόσον έχει προηγηθεί κάποιο φιλτράρισμα του θορύβου στα δεδομένα, παρουσιάζεται αυξημένη η παράμετρος της γενίκευσης.



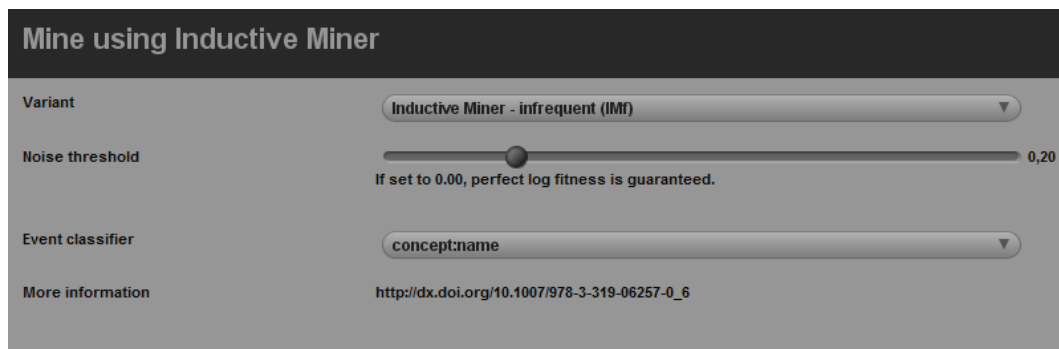
Εικόνα 34: Heuristics Miner στο event log με στοιχεία από τα ΒΚ

Το μοντέλο που προκύπτει από την εφαρμογή του Heuristics Miner στο δεύτερο αρχείο καταγραφής συμβάντων, το οποίο περιέχει επιπλέον την πληροφορία των ΒΚ, φαίνεται με ακόμη περισσότερο μειωμένη την παράμετρο της απλότητας, καθώς υπάρχουν ακόμη περισσότερα μαύρα τετράγωνα και περισσότερες συνδέσεις μεταξύ των δραστηριοτήτων. Γενικώς, η ύπαρξη αρκετών συνδέσεων μειώνει επιπλέον και την παράμετρο της ακρίβειας, καθώς το μοντέλο επιτρέπει περισσότερες συμπεριφορές από αυτές που θα έπρεπε.

### 5.3.3 Inductive Miner

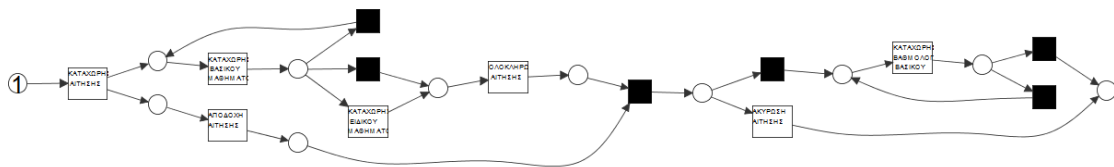
Ο Inductive Miner είναι βελτίωση των προηγούμενων δύο αλγορίθμων, αποτελεί την παρούσα στιγμή την κορυφαία λύση στην ανακάλυψη διαδικασιών και εγγυάται τη δημιουργία ενός sound μοντέλου διεργασιών, η σειρά δηλαδή των δραστηριοτήτων και οι συνδέσεις μεταξύ τους δεν οδηγούν τη διαδικασία σε τέλμα και το token ξεκινάει από την αρχή και καταλήγει πάντα στο τέλος, δεν μένει ποτέ κάπου στη μέση της διαδικασίας.

Για την εφαρμογή του Inductive Miner πάνω στα δύο αρχεία καταγραφής συμβάντων, επιλέγεται το πρόσθετο “Mine Petri Net using Inductive Miner”. Ο Inductive Miner δίνει την επιλογή να οριστεί κάποιο κατώφλι θορύβου, όπως φαίνεται και στην παρακάτω εικόνα, το οποίο υποδεικνύει πόση συμπεριφορά μπορεί να αγνοηθεί από τον αλγόριθμο και να μη ληφθεί υπόψη στην κατασκευή του μοντέλου, προκειμένου να πάρουμε ένα απλούστερο μοντέλο.

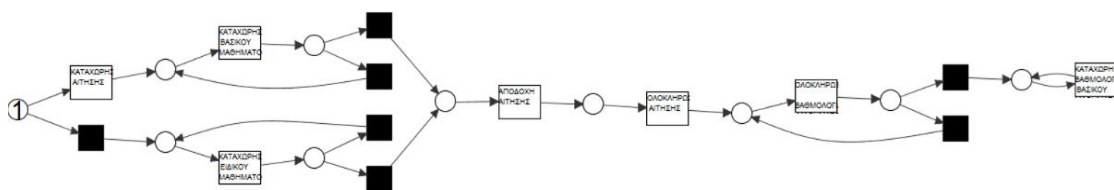


Εικόνα 35: Κατώφλι θορύβου στον Inductive Miner

Τα μοντέλα που προκύπτουν από τις διαφορετικές εκτελέσεις του Inductive Miner παρουσιάζονται παρακάτω:



Εικόνα 36: Inductive Miner με φίλτρο θορύβου 20%



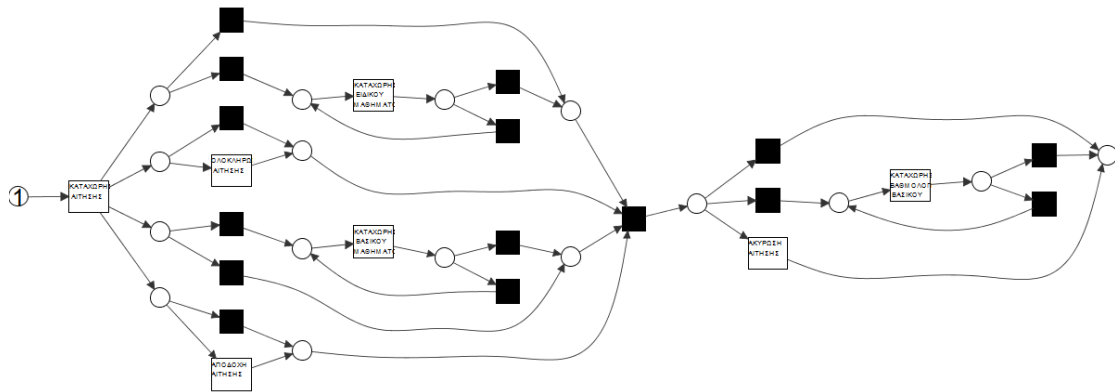
Εικόνα 37: Inductive Miner με φίλτρο θορύβου 20% στο event log με στοιχεία από τα ΒΚ

Και στα δύο αρχεία καταγραφής συμβάντων, τα μοντέλα που προκύπτουν με την προκαθορισμένη τιμή στην παράμετρο φιλτραρίσματος θορύβου (20%), ικανοποιούν τα κριτήρια της απλότητας, της καταλληλότητας και της ακρίβειας σε μεγάλο βαθμό. Ωστόσο το μοντέλο που προκύπτει από το αρχείο καταγραφής συμβάντων το οποίο περιέχει και τα δεδομένα των ΒΚ, λόγω του φιλτραρίσματος δεν εμφανίζει τη δραστηριότητα ΑΚΥΡΩΣΗ ΑΙΤΗΣΗΣ.

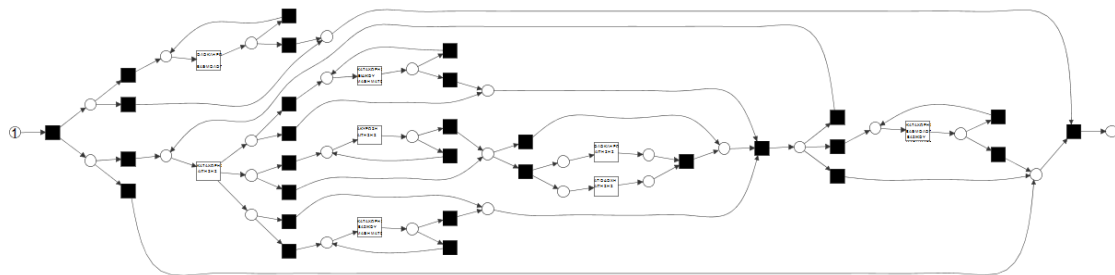
Η παράμετρος της γενίκευσης δεν δείχνει να έχει ιδιαίτερα αυξημένη τιμή, κάτι που είναι επίσης καλό στη συγκεκριμένη περίπτωση καθώς δεν θα έπρεπε το μοντέλο να ικανοποιεί συμπεριφορές που ξεφεύγουν από τη διαδικασία που προβλέπεται.

Στο σημείο αυτό, έχει ενδιαφέρον να δούμε πόσο σημαντικό ρόλο παίζει το φιλτράρισμα της μη συχνής συμπεριφοράς στα δεδομένα ενός αρχείου καταγραφής συμβάντων, καθώς, εάν ορίσουμε την παράμετρο για το κατώφλι θορύβου στο 0%, να

μην φιλτράρεται δηλ. καθόλου ο θόρυβος και οι μη συχνές συμπεριφορές, προκύπτουν μοντέλα που απεικονίζουν όλες τις δραστηριότητες και τα οποία ικανοποιούν όλες τις περιπτώσεις που υπάρχουν στα αρχεία καταγραφής συμβάντων. Ωστόσο τα μοντέλα αυτά είναι αρκετά πιο πολύπλοκα σε σχέση με τα προηγούμενα και εντέλει ο στόχος της ανακάλυψης διαδικασιών δεν είναι μόνο η υψηλή καταλληλότητα αλλά πρέπει να υπάρχει ισορροπία μεταξύ όλων των ποιοτικών παραμέτρων.



Εικόνα 38: Inductive Miner με φίλτρο θορύβου 0%



Εικόνα 39: Inductive Miner με φίλτρο θορύβου 0% στο event log με στοιχεία από τα ΒΚ

### 5.3.4 Inductive Visual Miner

Παράλληλα με την εφαρμογή του Inductive Miner, μπορούμε να εκτελέσουμε και τον Inductive Visual Miner, έναν αλγόριθμο ο οποίος επιτρέπει να εξετάσουμε και να διερευνήσουμε αρκετά γρήγορα τα πραγματικά δεδομένα. Επιτρέπει να δούμε αν υπάρχει κάποια «κρυμμένη» διαδρομή στα δεδομένα μας και επιπλέον δίνει τη δυνατότητα να γίνει επανάληψη της αναπαραγωγής των δεδομένων του αρχείου καταγραφής συμβάντων και να οπτικοποιηθεί με χρήση animation η ροή των διαφόρων συμβάντων στο μοντέλο διεργασίας που έχει παραχθεί και να έχουμε έτσι μια αρχική αίσθηση του πού ενδεχομένως δαπανάται αρκετός χρόνος στην όλη διεργασία. Ο Inductive Visual Miner έχει τη δυνατότητα να φιλτράρει συμπεριφορές που δεν αναπαράγονται συχνά.

Εκτελείται επιλέγοντας το πρόσθετο “Mine with Inductive Visual Miner” και αυτό που κάνει στη συνέχεια είναι ευθυγράμμιση του μοντέλου που ανακαλύπτεται με τα

δεδομένα που περιέχονται στο αρχείο καταγραφής συμβάντων. Υπάρχει δυνατότητα επιλογής των δραστηριοτήτων που εμφανίζονται (προεπιλογή 100%) καθώς και των μονοπατιών που συνδέουν τις δραστηριότητες μεταξύ τους (προεπιλογή 80%). Επίσης υπάρχει η δυνατότητα επιλογής εμφάνισης «Μονοπατιών και Αποκλίσεων» καθώς και «Μονοπατιών και Χρόνων Αναμονής».

Με τις παραπάνω επιλογές, εκτελείται το πρόσθετο για το πρώτο αρχείο καταγραφής συμβάντων και παίρνουμε τα παρακάτω αποτελέσματα, καθένα από τα οποία χωρίζεται από κάτω σε δύο επιπλέον εικόνες για καλύτερη προβολή τους:

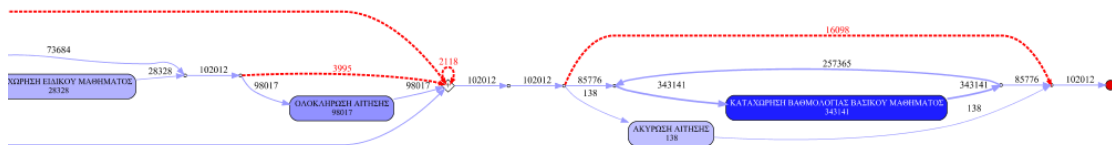
### 1) Inductive Visual Miner - Μονοπάτια και Αποκλίσεις



Εικόνα 40: Inductive Visual Miner - Μονοπάτια και Αποκλίσεις



Εικόνα 41: Inductive Visual Miner - Μονοπάτια και Αποκλίσεις (α' μέρος)



Εικόνα 42: Inductive Visual Miner - Μονοπάτια και Αποκλίσεις (β' μέρος)

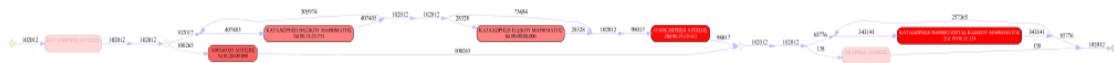
Κατά την επιλογή εμφάνισης «Μονοπάτια και Αποκλίσεις» παρατηρούνται με κόκκινες διακεκομμένες γραμμές τα σημεία στα οποία τα δεδομένα αποκλίνουν ως προς τη συμπεριφορά τους σε σχέση με ότι απεικονίζεται στο μοντέλο. Πάνω στις κόκκινες αυτές γραμμές εμφανίζεται επιπλέον και ο αριθμός των δεδομένων που ακολουθούν στο αποκλίνον μονοπάτι.

Οι αποκλίνουσες αυτές συμπεριφορές συνήθως δεν αποτελούν σφάλματα της συνολικής διαδικασίας αλλά πρόκειται για διαφορετικά επιτρεπόμενα σενάρια ακολουθίας των συμβάντων. Σε ορισμένες περιπτώσεις ωστόσο μπορεί να εντοπιστεί σε αυτή την απεικόνιση μια συμπεριφορά η οποία δεν θα έπρεπε να παρατηρείται.

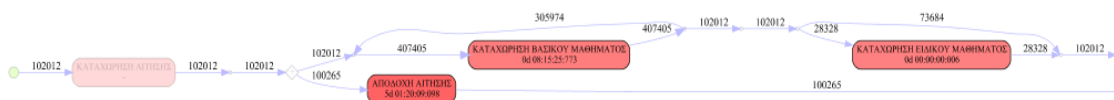
Στο συγκεκριμένο μοντέλο, παρατηρείται ότι για 16.098 περιπτώσεις παρακάμφθηκαν οι δραστηριότητες είτε της ΚΑΤΑΧΩΡΗΣΗΣ ΒΑΘΜΟΛΟΓΙΑΣ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ είτε της ΑΚΥΡΩΣΗΣ ΑΙΤΗΣΗΣ, σενάριο το οποίο είναι επιτρεπτό καθώς αφορά όλους τους

υποψηφίους οι οποίοι υπέβαλλαν αίτηση-δήλωση και στη συνέχεια δεν προσήλθαν να εξεταστούν σε όλα ή σε κάποια από τα μαθήματα. Από την άλλη, παρατηρείται ότι για 3.995 περιπτώσεις δεν πραγματοποιήθηκε ΟΛΟΚΛΗΡΩΣΗ ΑΙΤΗΣΗΣ, κάτι που όπως έχουμε επισημάνει και πιο πάνω δεν θα έπρεπε να επιτρέπεται στο πληροφοριακό σύστημα.

## 2) Inductive Visual Miner - Μονοπάτια και Χρόνοι Αναμονής



Εικόνα 43: Inductive Visual Miner - Μονοπάτια και Χρόνοι Αναμονής



Εικόνα 44: Inductive Visual Miner - Μονοπάτια και Χρόνοι Αναμονής (α' μέρος)



Εικόνα 45: Inductive Visual Miner - Μονοπάτια και Χρόνοι Αναμονής (β' μέρος)

Κατά την επιλογή εμφάνισης «Μονοπάτια και Χρόνοι Αναμονής» παρατηρούνται με εντονότερο χρώμα οι πιο χρονοβόρες δραστηριότητες της διαδικασίας και κάτω από την κάθε δραστηριότητα εμφανίζεται ο μέσος χρόνος ολοκλήρωσής της κι έτσι υπάρχει η δυνατότητα να παρατηρήσουμε εάν κάποια δραστηριότητα καθυστερεί αδικαιολόγητα προκειμένου να διερευνηθούν οι λόγοι καθυστέρησης.

Στο συγκεκριμένο μοντέλο οι δύο πιο χρονοβόρες δραστηριότητες είναι η ΟΛΟΚΛΗΡΩΣΗ ΑΙΤΗΣΗΣ και η ΚΑΤΑΧΩΡΗΣΗ ΒΑΘΜΟΛΟΓΙΑΣ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ με 20 και 21 μέρες αντίστοιχα, διαστήματα τα οποία είναι αναμενόμενα καθώς είναι τα διαστήματα που ορίζονται και προβλέπονται από τη νομοθεσία και τις σχετικές αποφάσεις που εκδίδονται.

Εάν περιμένουμε λίγη ώρα να εμφανιστεί και η ροή των συμβάντων στο μοντέλο, παρατηρούμε τη συγκέντρωση των συμβάντων στην τελευταία δραστηριότητα, καθώς πρέπει να έχει ολοκληρωθεί η προηγούμενη της προκειμένου αυτή να ξεκινήσει, κι έτσι όλα τα συμβάντα συσσωρεύονται πλέον σε μία δραστηριότητα (Εικόνα 46).

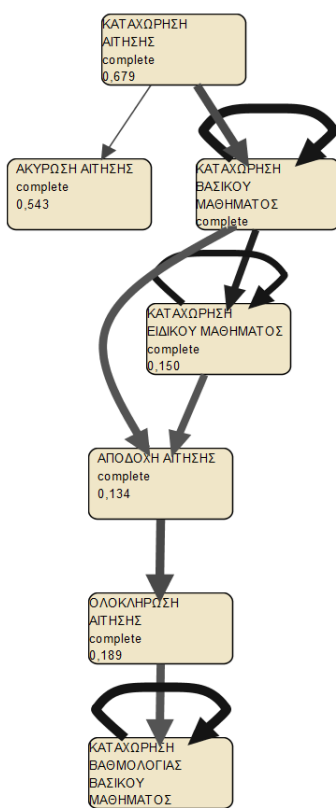


Εικόνα 46: Inductive Visual Miner - Ποή Συμβάντων

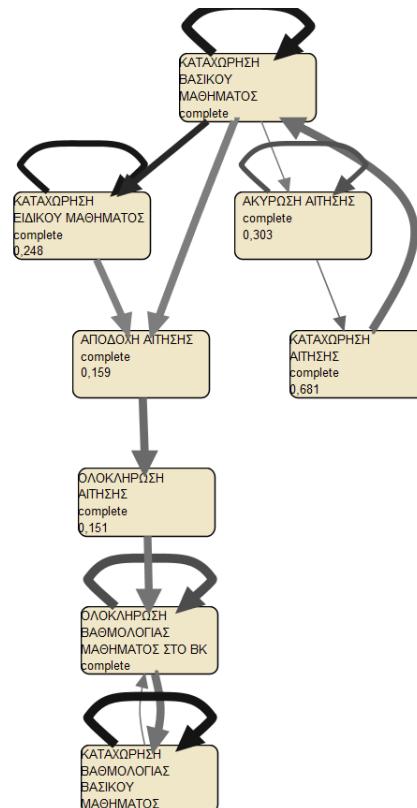
### 5.3.5 Fuzzy Miner

Ο Fuzzy Miner είναι ένας αλγόριθμος λίγο διαφορετικός από τους παραπάνω, μια και δεν ανακαλύπτει Petri-Nets αλλά γράφους διεργασιών (process graphs). Η ιδιαιτερότητα με τους γράφους διεργασιών είναι πως δεν μπορεί να γίνει διάκριση μεταξύ παραλληλίας και επιλογής, εάν δηλαδή δύο δραστηριότητες πραγματοποιούνται παράλληλα με οποιαδήποτε σειρά ή αν πραγματοποιείται η μία από τις δύο. Παρόλα αυτά, παραμένει ένα πολύ χρήσιμο εργαλείο για μια γρήγορη εξερεύνηση και γνώση των δεδομένων μας και για να έχουμε μια οπτικοποίηση της διαδικασίας, αλλά η δομή του γράφου παραμένει δύσκολο να ερμηνευτεί κι επομένως ο Fuzzy Miner δεν είναι εργαλείο για να εξεταστεί η ικανοποίηση των τεσσάρων βασικών παραμέτρων.

Ο Fuzzy Miner εκτελείται επιλέγοντας το πρόσθετο “Mine for a Fuzzy Model” και για τα δύο αρχεία καταγραφής συμβάντων που εξετάζουμε παίρνουμε τα παρακάτω αποτελέσματα:



Εικόνα 47: Fuzzy Model του πρώτου event log (Significance cutoff 0%)



Εικόνα 48: Fuzzy Model του event log με τα δεδομένα των ΒΚ (Significance cutoff 0%)



Παρόλο που δεν υπάρχει η δυνατότητα για περαιτέρω ανάλυση, ο Fuzzy Miner όπως φαίνεται ανακαλύπτει τον βασικό κορμό της διαδικασίας και στις δύο περιπτώσεις και σε κάθε περίπτωση αποτελεί έναν εύχρηστο τρόπο να έχει κανείς μια διαγραμματική αναπαράσταση της διαδικασίας.

### 5.3.6 Disco

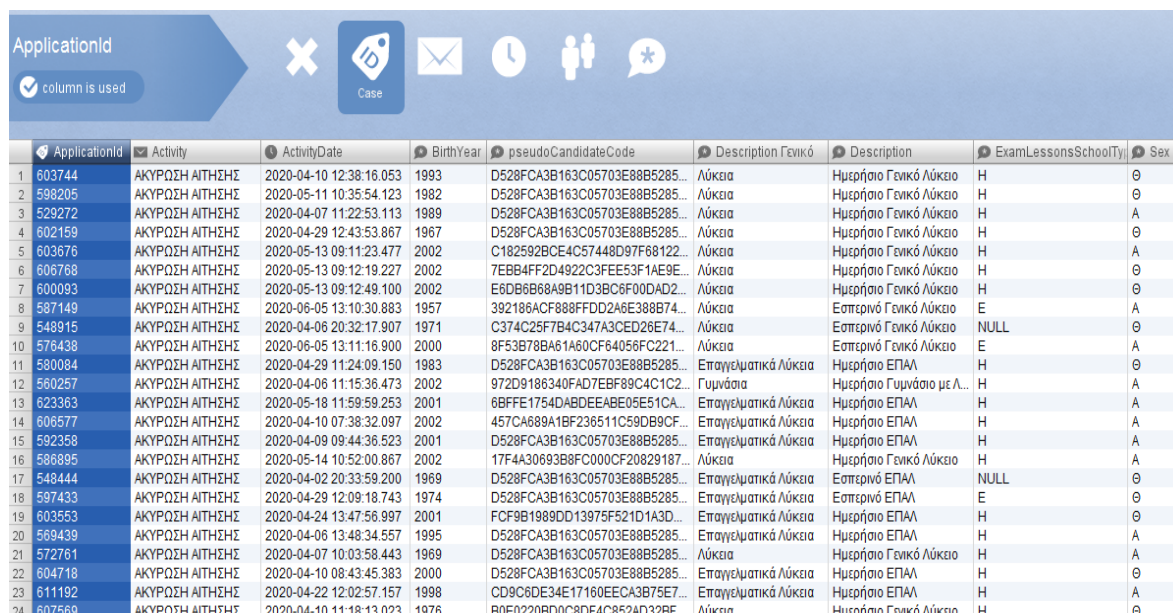
Από τη στιγμή που έγινε επεξεργασία και ενοποίηση των δεδομένων που είχαν σταλεί από τις αρμόδιες Διευθύνσεις του ΥΠΑΙΘ σε δύο τελικά αρχεία που αφορούν τη διαδικασία αιτήσεων-δηλώσεων για το σχολικό έτος 2019-2020, ένα που σταματά στην καταχώρηση βαθμολογίας βασικού μαθήματος στο πληροφοριακό σύστημα και περιέχει και ορισμένες επιπλέον πληροφορίες (είδος σχολείου κτλ) και ένα που περιέχει επιπλέον και την πληροφορία της ολοκλήρωσης βαθμολογίας των βασικών μαθημάτων στα Βαθμολογικά Κέντρα (ΒΚ), μπορούμε να εισάγουμε τα αρχεία αυτά στο εργαλείο Disco, το οποίο όπως έχει ήδη αναφερθεί είναι ένα εργαλείο εξόρυξης διαδικασιών που έχει μεγάλη χρηστικότητα, πιστότητα και αποδοτικότητα και καθιστά την εξόρυξη διαδικασιών εύκολη και γρήγορη.

Αρχικά πραγματοποιήθηκε η εισαγωγή του πρώτου αρχείου καταγραφής συμβάντων και στη συνέχεια ζητήθηκε η επιλογή των τριών βασικών παραμέτρων που ορίζονται κατά την εξόρυξη διαδικασιών. Όπως και στο ProM, ορίστηκαν οι εξής παράμετροι:

**CaseID:** ApplicationID

**Activity:** Activity

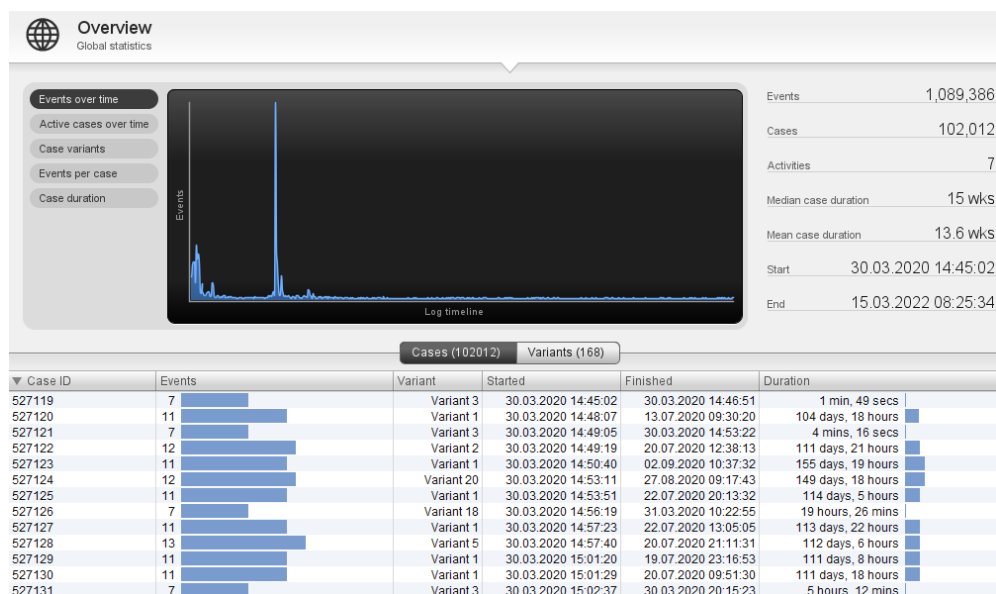
**Timestamp:** ActivityDate



ApplicationId	Activity	ActivityDate	BirthYear	pseudoCandidateCode	Description Γενικό	Description	ExamLessonsSchoolTy	Sex
1	603744	ΑΚΥΡΩΣΗ ΑΙΤΗΣΗΣ	2020-04-10 12:38:16.053	1993	D528FCA3B163C05703E88B5285...	Λύκεια	Ημερήσιο Γενικό Λύκειο	H
2	598205	ΑΚΥΡΩΣΗ ΑΙΤΗΣΗΣ	2020-05-11 10:35:54.123	1982	D528FCA3B163C05703E88B5285...	Λύκεια	Ημερήσιο Γενικό Λύκειο	H
3	529272	ΑΚΥΡΩΣΗ ΑΙΤΗΣΗΣ	2020-04-07 11:22:53.113	1989	D528FCA3B163C05703E88B5285...	Λύκεια	Ημερήσιο Γενικό Λύκειο	H
4	602159	ΑΚΥΡΩΣΗ ΑΙΤΗΣΗΣ	2020-04-29 12:43:53.867	1967	D528FCA3B163C05703E88B5285...	Λύκεια	Ημερήσιο Γενικό Λύκειο	H
5	603676	ΑΚΥΡΩΣΗ ΑΙΤΗΣΗΣ	2020-05-13 09:11:23.477	2002	C182592BCE4C57448D97F68122...	Λύκεια	Ημερήσιο Γενικό Λύκειο	H
6	606768	ΑΚΥΡΩΣΗ ΑΙΤΗΣΗΣ	2020-05-13 09:12:19.227	2002	7EBB4FF2D4922C3FEE53F1AE9E...	Λύκεια	Ημερήσιο Γενικό Λύκειο	H
7	600093	ΑΚΥΡΩΣΗ ΑΙΤΗΣΗΣ	2020-05-13 09:12:49.100	2002	E6DB6B68A9B11D3BC6F00AD2...	Λύκεια	Ημερήσιο Γενικό Λύκειο	H
8	587149	ΑΚΥΡΩΣΗ ΑΙΤΗΣΗΣ	2020-06-05 13:10:30.883	1957	392186ACF888FFDD2A6E388B74...	Λύκεια	Εσπερινό Γενικό Λύκειο	E
9	548915	ΑΚΥΡΩΣΗ ΑΙΤΗΣΗΣ	2020-04-06 20:32:17.907	1971	C374C25F7B4C347A3CED26E74...	Λύκεια	Εσπερινό Γενικό Λύκειο	NULL
10	576438	ΑΚΥΡΩΣΗ ΑΙΤΗΣΗΣ	2020-06-05 13:11:16.900	2000	8F53B78BA61A60CF64056FC221...	Λύκεια	Εσπερινό Γενικό Λύκειο	E
11	580084	ΑΚΥΡΩΣΗ ΑΙΤΗΣΗΣ	2020-04-29 11:24:09.150	1983	D528FCA3B163C05703E88B5285...	Επαγγελματικά Λύκεια	Ημερήσιο ΕΠΑΛ	H
12	560257	ΑΚΥΡΩΣΗ ΑΙΤΗΣΗΣ	2020-04-06 11:15:36.473	2002	972D9186340FAD7EBF89C4C1C2...	Γυμνάσιο	Ημερήσιο Γυμνάσιο με Λ...	H
13	623363	ΑΚΥΡΩΣΗ ΑΙΤΗΣΗΣ	2020-05-18 11:59:59.253	2001	6BF6E1754DABDEEABE05E51CA...	Επαγγελματικά Λύκεια	Ημερήσιο ΕΠΑΛ	H
14	606577	ΑΚΥΡΩΣΗ ΑΙΤΗΣΗΣ	2020-04-10 07:38:32.097	2002	457CA689A1BF236511C59DB9CF...	Επαγγελματικά Λύκεια	Ημερήσιο ΕΠΑΛ	H
15	592358	ΑΚΥΡΩΣΗ ΑΙΤΗΣΗΣ	2020-04-09 09:44:36.523	2001	D528FCA3B163C05703E88B5285...	Επαγγελματικά Λύκεια	Ημερήσιο ΕΠΑΛ	H
16	586895	ΑΚΥΡΩΣΗ ΑΙΤΗΣΗΣ	2020-05-14 10:52:00.867	2002	17F4A30693B8FC000C2F20829187...	Λύκεια	Ημερήσιο Γενικό Λύκειο	H
17	548444	ΑΚΥΡΩΣΗ ΑΙΤΗΣΗΣ	2020-04-02 20:33:59.200	1969	D528FCA3B163C05703E88B5285...	Επαγγελματικά Λύκεια	Εσπερινό ΕΠΑΛ	NULL
18	597433	ΑΚΥΡΩΣΗ ΑΙΤΗΣΗΣ	2020-04-29 12:09:18.743	1974	D528FCA3B163C05703E88B5285...	Επαγγελματικά Λύκεια	Εσπερινό ΕΠΑΛ	E
19	603553	ΑΚΥΡΩΣΗ ΑΙΤΗΣΗΣ	2020-04-24 13:47:56.997	2001	FCF9B1989DD13975F521D1A3D...	Επαγγελματικά Λύκεια	Ημερήσιο ΕΠΑΛ	H
20	569439	ΑΚΥΡΩΣΗ ΑΙΤΗΣΗΣ	2020-04-06 13:48:34.557	1995	D528FCA3B163C05703E88B5285...	Επαγγελματικά Λύκεια	Ημερήσιο ΕΠΑΛ	H
21	572761	ΑΚΥΡΩΣΗ ΑΙΤΗΣΗΣ	2020-04-07 10:03:58.443	1969	D528FCA3B163C05703E88B5285...	Λύκεια	Ημερήσιο Γενικό Λύκειο	H
22	604718	ΑΚΥΡΩΣΗ ΑΙΤΗΣΗΣ	2020-04-10 08:43:45.383	2000	D528FCA3B163C05703E88B5285...	Επαγγελματικά Λύκεια	Ημερήσιο ΕΠΑΛ	H
23	611192	ΑΚΥΡΩΣΗ ΑΙΤΗΣΗΣ	2020-04-22 12:02:57.157	1998	CD9C6DE34E17160EECA3B75E7...	Επαγγελματικά Λύκεια	Ημερήσιο ΕΠΑΛ	H
24	607569	ΑΚΥΡΩΣΗ ΑΙΤΗΣΗΣ	2020-04-10 11:18:13.023	1976	B0E0220BD0C8DF4C852AD32BF...	Λύκεια	Ημερήσιο Γενικό Λύκειο	H

Εικόνα 49: Επιλογή παραμέτρων στο Disco

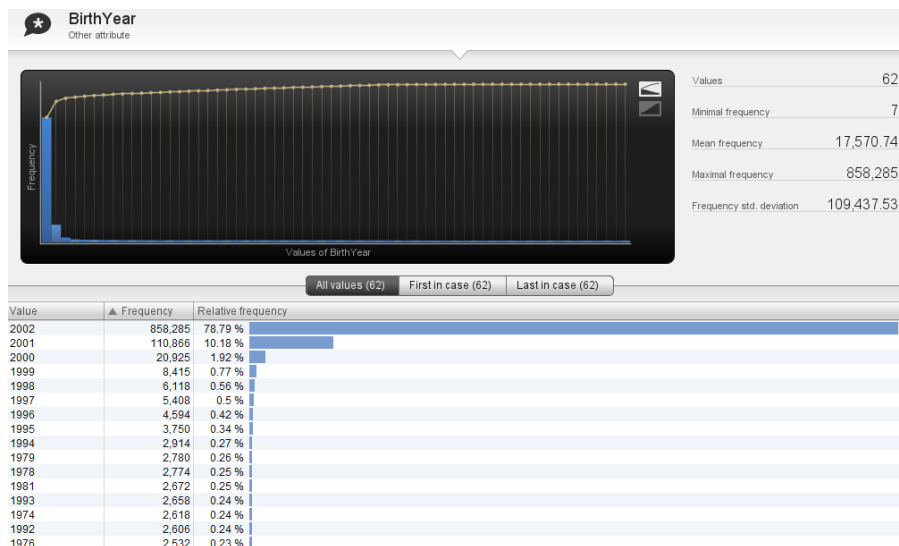
Το Disco παρέχει αρκετά εργαλεία επισκόπησης της διαδικασίας για την οποία γίνεται ανάλυση:



Εικόνα 50: Γενική επισκόπηση στο Disco

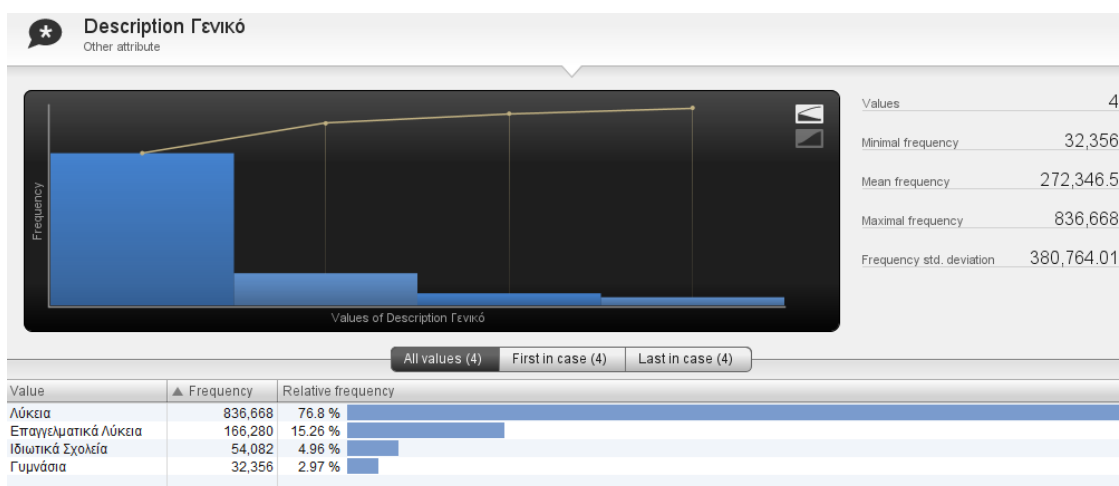
Με μια πρώτη γενική επισκόπηση, παρατηρούμε ότι υπάρχουν 1.089.386 συμβάντα για 102.012 περιπτώσεις της διαδικασίας αιτήσεων-δηλώσεων και 7 διαφορετικές δραστηριότητες. Η μέση διάρκεια των περιπτώσεων της διαδικασίας είναι 13,6 εβδομάδες και παρατηρούμε ότι το πρώτο συμβάν λαμβάνει χώρα στις 30-3-2020 και το τελευταίο στις 15-3-2022, κάτι που είχαμε παρατηρήσει και κατά την επισκόπηση στο ProM και αποτελεί παράδοξο που χρήζει διερεύνησης. Στο γράφημα φαίνεται ότι υπάρχει αρχικά συγκεντρωμένη δραστηριότητα, η οποία αφορά στη διαδικασία καταχώρησης των επιλογών της αίτησης-δήλωσης και αργότερα μία δραστηριότητα που αφορά στην καταχώρηση βαθμολογίας των βασικών μαθημάτων, κυρίως τον Ιούλιο του 2020 και λιγότερο το Σεπτέμβριο του 2020 (στις επαναληπτικές εξετάσεις). Από εκεί και πέρα, υπάρχει πολλή χαμηλή κίνηση συμβάντων τα οποία ωστόσο συνεχίζονται για αρκετό διάστημα (Μάρτιος 2022).

Το Disco παρέχει μια άμεση εικόνα κατανομής συχνοτήτων και για τις υπόλοιπες παραμέτρους που περιέχονται στο αρχείο καταγραφής συμβάντων. Ενδεικτικά παρουσιάζονται οι παρακάτω κατανομές των συμβάντων ανά έτος γέννησης του υποψηφίου και ανά είδος σχολείου:



Εικόνα 51: Κατανομή των συμβάντων ανά έτος γέννησης του υποψηφίου

Όσον αφορά στο έτος γέννησης, περίπου το 90% των περιπτώσεων αναφέρονται σε υποψήφιους που τελείωσαν το Λύκειο το έτος που εξετάζεται ή το προηγούμενο. Ωστόσο υπάρχουν 60 ακόμη τιμές στα έτη γέννησης που καταλαμβάνουν αρκετά μικρότερα ποσοστά.

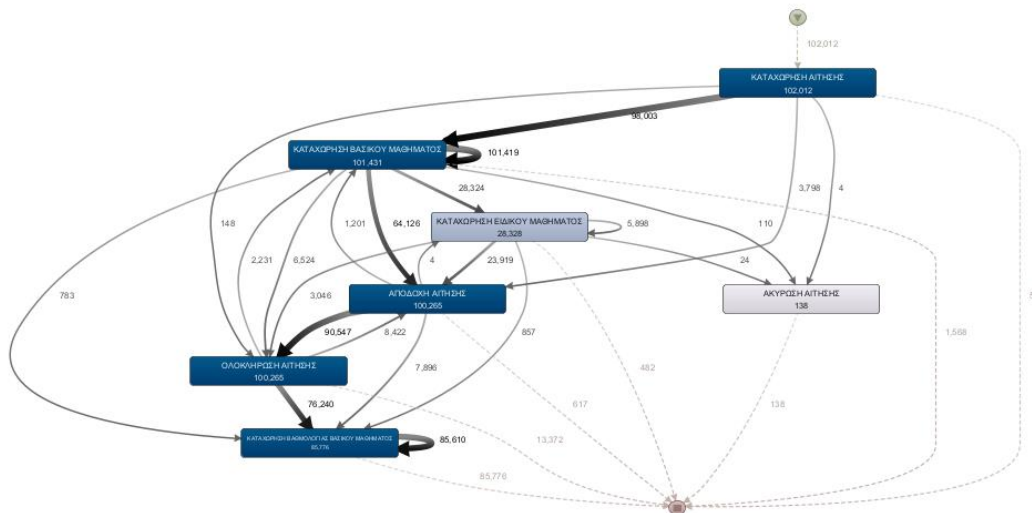


Εικόνα 52: Κατανομή των συμβάντων ανά είδος Λυκείου

Όσον αφορά στο είδος του σχολείου, παρατηρείται ότι το 76,8% αναφέρεται σε υποψήφιους Γενικών Λυκείων και 15,26% σε υποψήφιους Επαγγελματικών Λυκείων, ενώ μικρότερα ποσοστά αναφέρονται σε υποψηφίους ιδιωτικών σχολείων ή γυμνασίων με λυκειακές τάξεις.

Μετά από την εισαγωγή του αρχείου δεδομένων στο Disco, δημιουργείται άμεσα ένας χάρτης που αποτελεί τη γραφική αναπαράσταση της προς ανάλυση διαδικασίας, παρόμοια με αυτή που προκύπτει από την εκτέλεση του Fuzzy Miner στο ProM, και είναι ένας εύκολος και άμεσος τρόπος να ερμηνευτεί η διαδικασία βασιζόμενοι σε πραγματικά γεγονότα. Στην παρακάτω εικόνα εμφανίζεται ο χάρτης της διαδικασίας για

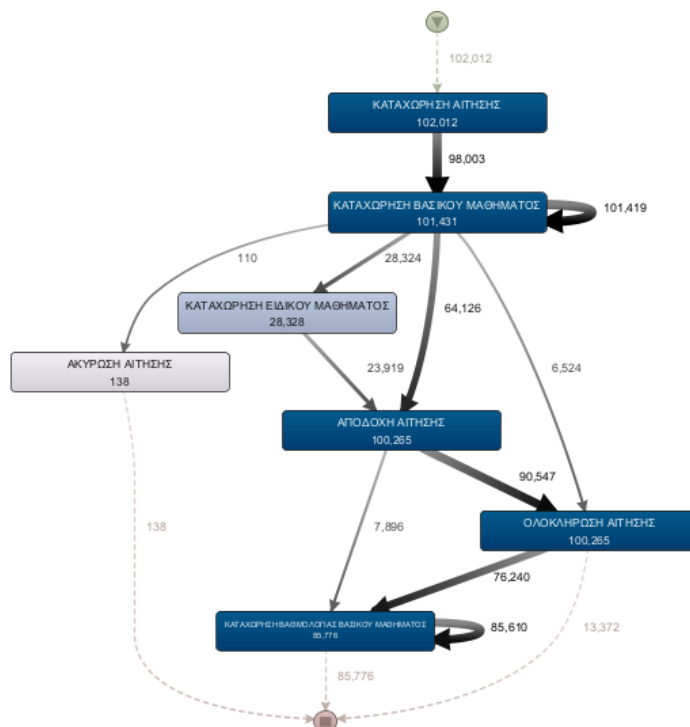
το 100% των δραστηριοτήτων και για το 100% των διαδρομών που καταγράφονται στο event log:



Εικόνα 53: Χάρτης διαδικασίας (100% Activities - 100% Paths)

Παρόλο που πρόκειται για μια σχετικά απλή και δομημένη διαδικασία, επειδή εμφανίζονται όλες οι διαδρομές που έχουν καταγραφεί ο χάρτης που προκύπτει δεν μπορεί να ερμηνευτεί εύκολα και να βγει ασφαλές συμπέρασμα.

Μειώνοντας την απεικόνιση των μονοπατιών στο 20%, φιλτράρονται οι λιγότερο συχνές διαδρομές κι έτσι εμφανίζεται ένας πιο απλοποιημένος χάρτης:



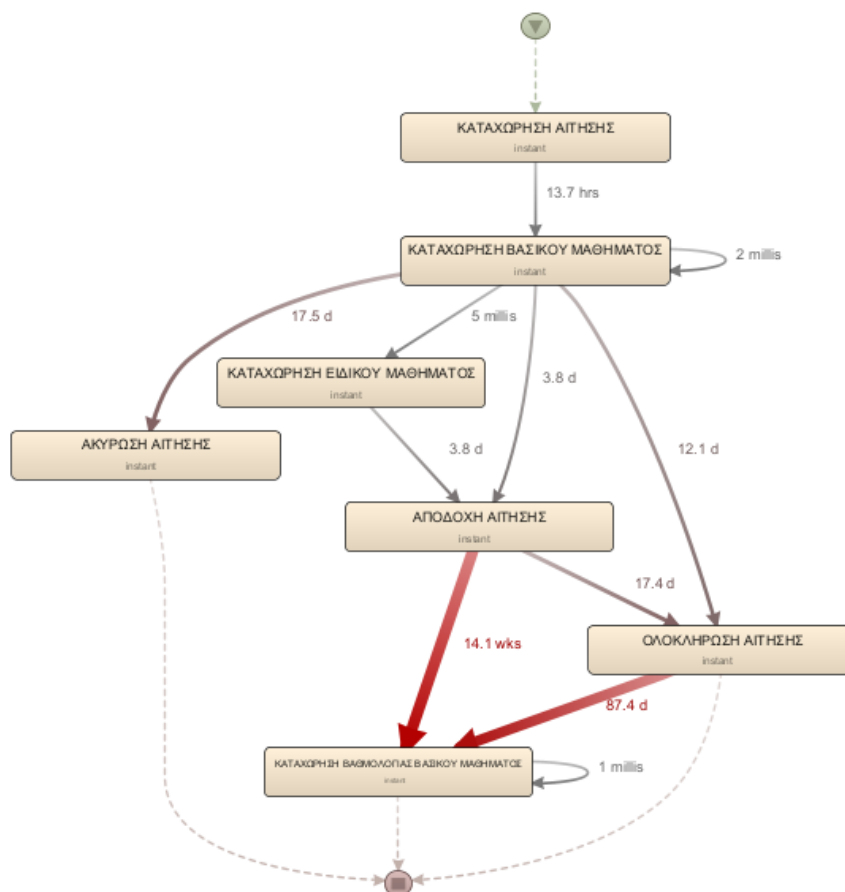
Εικόνα 54: Χάρτης διαδικασίας (100% Activities - 20% Paths)

Τώρα το γράφημα που προκύπτει απεικονίζει άριστα τη διαδικασία όπως αυτή προβλέπεται, πράγμα που σημαίνει ότι αυτή τηρείται στην μεγάλη πλειοψηφία των περιπτώσεων.

Η διαβάθμιση του μπλε χρώματος υποδεικνύει τη συχνότητα εμφάνισης της συγκεκριμένης δραστηριότητας. Στη συγκεκριμένη περίπτωση με έντονο μπλε χρώμα εμφανίζονται οι δραστηριότητες του βασικού κορμού της διαδικασίας μας (ΚΑΤΑΧΩΡΗΣΗ ΑΙΤΗΣΗΣ – ΚΑΤΑΧΩΡΗΣΗ ΒΑΣΙΚΩΝ ΜΑΘΗΜΑΤΩΝ – ΑΠΟΔΟΧΗ ΑΙΤΗΣΗΣ – ΟΛΟΚΛΗΡΩΣΗ ΑΙΤΗΣΗΣ – ΟΛΟΚΛΗΡΩΣΗ ΒΑΘΜΟΛΟΓΙΑΣ ΒΑΣΙΚΩΝ ΜΑΘΗΜΑΤΩΝ), με λιγότερο έντονο χρώμα εμφανίζεται η ΚΑΤΑΧΩΡΗΣΗ ΕΙΔΙΚΟΥ ΜΑΘΗΜΑΤΟΣ που είναι προαιρετική και με ακόμη λιγότερο έντονο χρώμα η ΑΚΥΡΩΣΗ ΑΙΤΗΣΗΣ που λαμβάνει χώρα σπάνια.

Τα βέλη που δείχνουν τις μεταβάσεις από τη μια δραστηριότητα στην άλλη έχουν δίπλα τους τον αριθμό των μεταβάσεων αυτών και το πάχος τους διαμορφώνεται ανάλογα με τον αριθμό αυτό. Στο σημείο αυτό φαίνεται και ότι η ΚΑΤΑΧΩΡΗΣΗ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ και η ΚΑΤΑΧΩΡΗΣΗ ΒΑΘΜΟΛΟΓΙΑΣ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ αποτελούν δραστηριότητες που επαναλαμβάνονται μέσα στην ίδια διαδικασία, καθώς πρόκειται για τέσσερα μαθήματα ανά υποψήφιο.

Τέλος, επιλέγοντας την απεικόνιση Performance στο Disco, εμφανίζεται ο ίδιος χάρτης στον οποίο αναγράφεται η διάρκεια της κάθε δραστηριότητας:

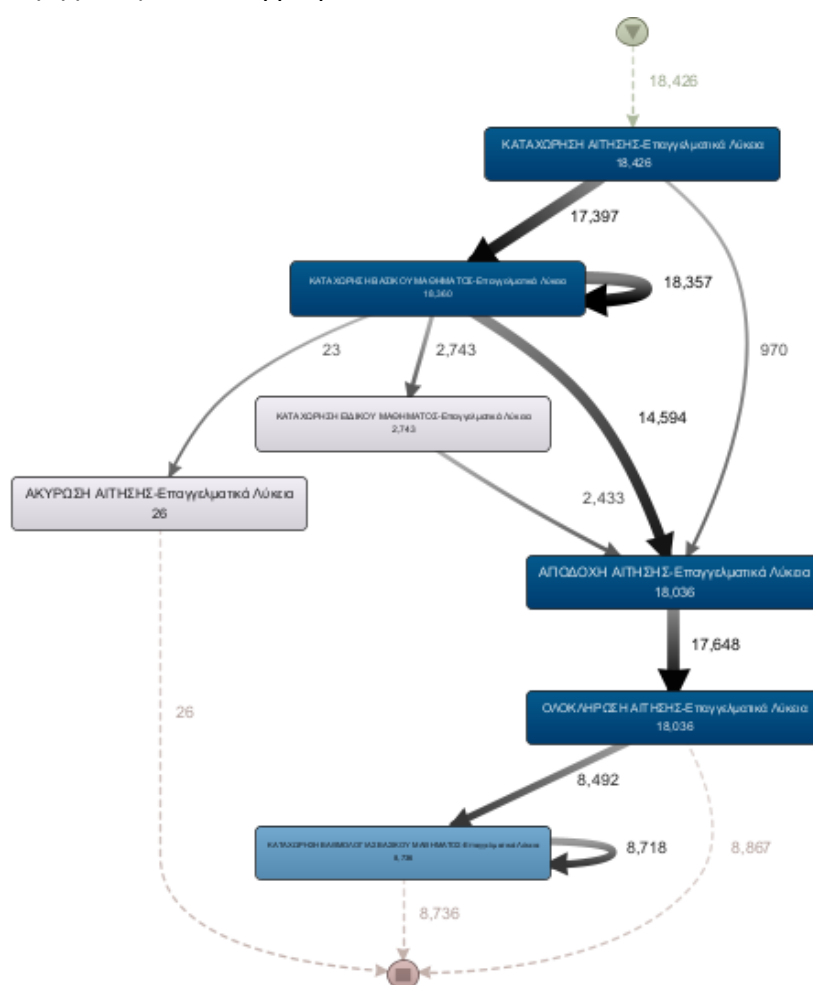


Εικόνα 55: Χάρτης μέσης διάρκειας δραστηριοτήτων

Στην περίπτωση αυτή, δίπλα στα βέλη των μεταβάσεων από τη μία δραστηριότητα στην άλλη φαίνονται οι μέσοι χρόνοι της κάθε μετάβασης και το πάχος του κάθε βέλους είναι ανάλογο της διάρκειας της μετάβασης. Στην παρούσα ανάλυση, η περισσότερη

αναμονή εμφανίζεται πριν από την ΚΑΤΑΧΩΡΗΣΗ ΒΑΘΜΟΛΟΓΙΑΣ ΒΑΣΙΚΩΝ ΜΑΘΗΜΑΤΩΝ, κάτι που είναι λογικό όμως εφόσον μεσολαβεί η εξέταση των υποψηφίων στα μαθήματα και η βαθμολόγησή τους στα Βαθμολογικά Κέντρα. Επίσης, αξιοσημείωτοι είναι οι μέσοι χρόνοι ΑΚΥΡΩΣΗΣ ΑΙΤΗΣΗΣ και ΟΛΟΚΛΗΡΩΣΗΣ ΑΙΤΗΣΗΣ οι οποίοι είναι 17,5 και 17,4 μέρες αντίστοιχα. Δεδομένου ότι το διάστημα που ορίζεται από την εγκύκλιο για τη διαδικασία των αιτήσεων-δηλώσεων συνήθως δεν ξεπερνά τις 2 εβδομάδες, είναι εμφανές ότι ο μέσος χρόνος αυξάνεται τελικά από τις περιπτώσεις που πραγματοποιούνται εκπρόθεσμες ενέργειες στο αρχείο των αιτήσεων-δηλώσεων.

Ένα πολύ εύχρηστο εργαλείο του Disco είναι η δυνατότητα γρήγορου φιλτραρίσματος των διαφόρων χαρακτηριστικών του αρχείου και η άμεση απεικόνιση του φιλτραρισμένου αρχείου. Ενδιαφέρον παρουσιάζει η παρακάτω απεικόνιση του χάρτη για τους υποψηφίους των Επαγγελματικών Λυκείων:

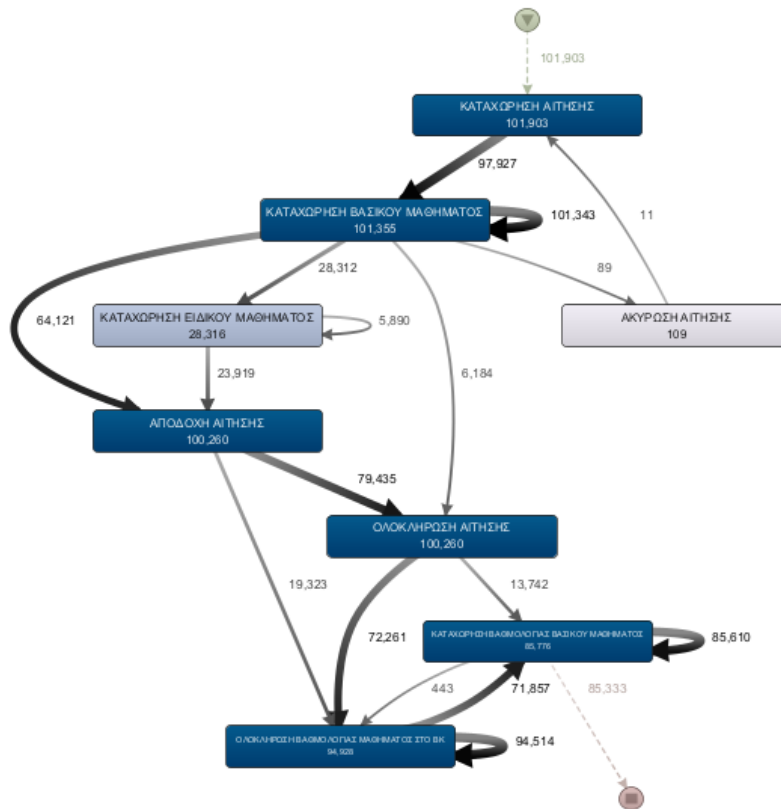


Εικόνα 56: Χάρτης διαδικασίας για τα Επαγγελματικά Λύκεια (100% Activities - 20% Paths)

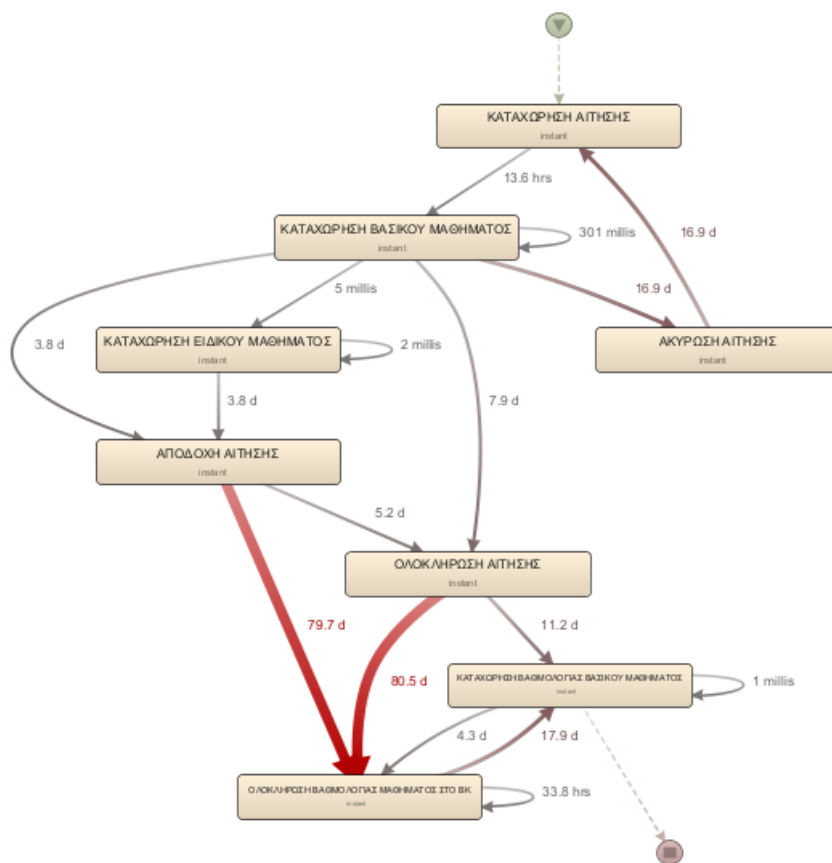
Παρατηρούμε εδώ ότι, βάσει των συχνοτήτων εμφάνισης των δραστηριοτήτων, ο βασικός κορμός της διαδικασίας διαφοροποιείται και δείχνει να σταματάει στην ΟΛΟΚΛΗΡΩΣΗ ΑΙΤΗΣΗΣ, ενώ η ΚΑΤΑΧΩΡΗΣΗ ΒΑΘΜΟΛΟΓΙΑΣ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ εμφανίζεται με μικρότερη συχνότητα, πράγμα που υποδεικνύει ότι μεγάλο ποσοστό των υποψηφίων των Επαγγελματικών Λυκείων (περίπου 50%) δεν προσέρχεται στην

εξέταση των μαθημάτων. Η σειρά ΑΠΟΔΟΧΗ-ΟΛΟΚΛΗΡΩΣΗ-ΒΑΘΜΟΛΟΓΗΣΗ είναι και η σωστή σειρά με την οποία θα πρέπει να εκτελούνται οι ενέργειες.

Τέλος, εάν εισάγουμε στο Disco το δεύτερο ενοποιημένο αρχείο, που περιέχει και τα δεδομένα από τα ΒΚ, προκύπτουν τα παρακάτω γραφήματα:



Εικόνα 57: Χάρτης διαδικασίας για το αρχείο με τα δεδομένα των ΒΚ (100% Activities - 20% Paths)



Εικόνα 58: Χάρτης μέσης διάρκειας δραστηριοτήτων για το αρχείο με τα δεδομένα των ΒΚ

Όπως υποδεικνύει και στους παραπάνω χάρτες η πορεία και το πάχος των βελών, ο βασικός κορμός των δραστηριοτήτων δείχνει κι εδώ να είναι ο αναμενόμενος, δηλ. ΚΑΤΑΧΩΡΗΣΗ ΑΙΤΗΣΗΣ – ΚΑΤΑΧΩΡΗΣΗ ΒΑΣΙΚΩΝ ΜΑΘΗΜΑΤΩΝ – ΑΠΟΔΟΧΗ ΑΙΤΗΣΗΣ – ΟΛΟΚΛΗΡΩΣΗ ΑΙΤΗΣΗΣ – ΟΛΟΚΛΗΡΩΣΗ ΒΑΘΜΟΛΟΓΙΑΣ ΒΑΣΙΚΩΝ ΜΑΘΗΜΑΤΩΝ – ΚΑΤΑΧΩΡΗΣΗ ΒΑΘΜΟΛΟΓΙΑΣ ΒΑΣΙΚΩΝ ΜΑΘΗΜΑΤΩΝ. Επίσης, όσον αφορά στη διάρκεια και στις καθυστερήσεις μετάβασης, δεν φαίνεται να υπάρχει κάποια καθυστέρηση στη διαδικασία ούτε τώρα που εμπλέκονται και τα Βαθμολογικά Κέντρα, καθώς η μόνη καθυστέρηση είναι δικαιολογημένη και οφείλεται στο χρόνο που χρειάζεται για να προετοιμαστεί και να πραγματοποιηθεί η διαδικασία εξέτασης, βάσει των στοιχείων που αντλούνται από τις αιτήσεις-δηλώσεις.

### 5.3.7 Παρατηρήσεις από τη διαδικασία του Process Discovery

Παρόλο που η διαδικασία των αιτήσεων-δηλώσεων των Πανελλαδικών εξετάσεων είναι μια διαδικασία αρκετά δομημένη και σαφώς ορισμένη, επειδή υπεισέρχεται σε μεγάλο βαθμό μέσα της το ανθρώπινο στοιχείο (υποψήφιοι που εξετάζονται, χειριστές του πληροφοριακού συστήματος) παρουσιάζει τελικά κατά την ανακάλυψή της την δυνατότητα αρκετών εναλλακτικών διαδρομών.

Έχοντας εφαρμόσει τους αλγόριθμους και τα εργαλεία ανακάλυψης διεργασιών που είχαμε στη διάθεσή μας παρατηρούμε πως οι αλγόριθμοι Alpha και Heuristics Miner δεν κατάφεραν να ανακαλύψουν και να μοντελοποιήσουν σωστά τη διαδικασία.



Συγκεκριμένα ο Heuristics έδειξε να κάνει μια αρκετά καλή προσέγγιση της διαδικασίας, αλλά δεν κατάφερε να οδηγήσει σε ένα sound μοντέλο που να ολοκληρώνεται σωστά σε κάθε περίπτωση. Αντίθετα, ο Inductive Miner και τα δύο εργαλεία παραγωγής γράφων, Fuzzy Miner και Disco, πραγματοποίησαν σωστή ανακάλυψη της διαδικασίας και έδωσαν τη δυνατότητα μιας ορθής απεικόνισης αυτής. Ειδικά ο Inductive Miner σε συνδυασμό με τον Inductive Visual Miner, έδωσε χρήσιμες πληροφορίες για τις αποκλίσεις και τους χρόνους καθυστέρησης στις διάφορες δραστηριότητες της διαδικασίας.

#### **5.4 Έλεγχος Συμμόρφωσης (Conformance Checking)**

Μετά από την εκτέλεση του βήματος όπου από τα δεδομένα που ήταν αποθηκευμένα στα αρχεία καταγραφής συμβάντων δημιουργήθηκαν τα μοντέλα διεργασιών μέσω των διαφόρων αλγορίθμων εξόρυξης διεργασιών, συνεχίζουμε με τον έλεγχο συμμόρφωσης (conformance checking) όπου τα δεδομένα συνδυάζονται με τα παραγόμενα μοντέλα με κύριο στόχο την ανίχνευση του κατά πόσο υπάρχει συμφωνία μεταξύ τους ή όχι και την κατανόηση της παρουσίας πιθανών αποκλίσεων καθώς και της φύσης αυτών των αποκλίσεων. Ο έλεγχος συμμόρφωσης πραγματοποιείται προκειμένου είτε να εντοπιστούν προβλήματα στην εκτέλεση της υπάρχουσας διαδικασίας και να διερευνηθούν για να διορθωθούν στο μέλλον είτε για να βελτιωθεί το υπάρχον μοντέλο διεργασίας έτσι ώστε να αποτυπώνει ορθότερα τη διαδικασία την οποία απεικονίζει.

Με τις τεχνικές του ελέγχου συμμόρφωσης στην ουσία εξετάζεται η παράμετρος της καταλληλότητας (fitness) σύμφωνα με την οποία το μοντέλο συνάδει σε μεγάλο βαθμό με το αρχείο καταγραφής συμβάντων, αναπαράγει δηλαδή το μεγαλύτερο μέρος της συμπεριφοράς που εμφανίζεται σε αυτό χωρίς να υπάρχουν αποκλίσεις. Ο έλεγχος αυτός πραγματοποιείται κάνοντας ευθυγράμμιση (alignment) του παραγόμενου μοντέλου με τα δεδομένα, ελέγχοντας μία προς μία τις καταγεγραμμένες διαδικασίες με τις διαδικασίες που απεικονίζει το μοντέλο, και παρατηρώντας έτσι κατά πόσο το μοντέλο περιγράφει τις αλληλουχίες γεγονότων που περιέχονται στο αρχείο καταγραφής συμβάντων ή κατά πόσο τα δεδομένα αποκλίνουν από τη μοντελοποίηση. Κατά την ευθυγράμμιση αυτή ελέγχεται αν μία σύνδεση δραστηριοτήτων υπάρχει στο event log και όχι στο μοντέλο καθώς και το αντίθετο και κάθε ασυμβατότητα ή απόκλιση υπολογίζεται και στη συνέχεια αθροίζεται προκειμένου να προκύψει ένα τελικό νούμερο που ορίζει την καταλληλότητα (fitness) του παραγόμενου μοντέλου.

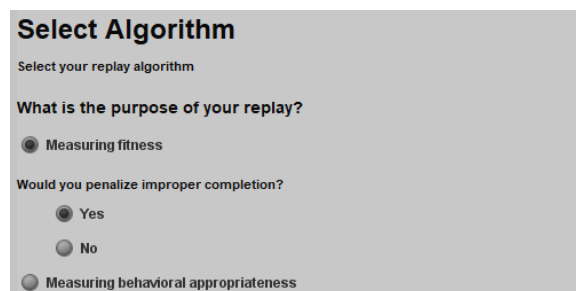
Στο ProM, ο έλεγχος συμμόρφωσης πραγματοποιείται λαμβάνοντας ως είσοδο ένα αρχείο καταγραφής συμβάντων και ένα μοντέλο διεργασιών και εκτελώντας στη συνέχεια το πρόσθετο “Replay a Log on Petri Net for Conformance Analysis”, και επιστρέφεται αρχείο με ενδείξεις σχετικά με την αντιστοιχία των συμπεριφορών που περιέχονται στο αρχείο καταγραφής σε σχέση με αυτές τις οποίες αναπαριστά το

μοντέλο. Σημειώνεται ότι με αυτό το πρόσθετο παρέχεται η εγγύηση ότι θα έχουμε τη βέλτιστη ευθυγράμμιση.

Ο βασικός στόχος του ελέγχου συμμόρφωσης στην παρούσα εργασία είναι να εξασφαλιστεί αρχικά ότι η διαδικασία των αιτήσεων-δηλώσεων των Πανελλαδικών εξετάσεων όπως πραγματικά συμβαίνει και καταγράφεται στο πληροφοριακό σύστημα ακολουθεί και ευθυγραμμίζεται με τους κανόνες που την διέπουν, καθώς πρόκειται για μια ιδιαίτερα ευαίσθητη και νευραλγική διαδικασία που πρέπει να πραγματοποιείται κάτω από αρχές διαφάνειας και συνέπειας. Ένας ακόμη στόχος είναι να εντοπιστούν τυχόν αδυναμίες του πληροφοριακού συστήματος ή/και τυχόν αστοχίες στη χρήση του, κάτι που θα μπορούσε να διορθωθεί μελλοντικά ώστε το σύστημα να καλύπτει πληρέστερα τις ανάγκες της διαδικασίας και να απαιτείται όσο το δυνατόν μικρότερη παρέμβαση από τους χειριστές. Τέλος, παρόλο που πρόκειται για μια διαδικασία με αρκετά ξεκάθαρους κανόνες, τυχόν παρατηρήσεις που προκύπτουν από τον έλεγχο συμμόρφωσης θα μπορούσαν ενδεχομένως να αποτελέσουν πρόταση για επαναξιολόγηση και αναθεώρηση μέρους της διαδικασίας από τους αρμόδιους.

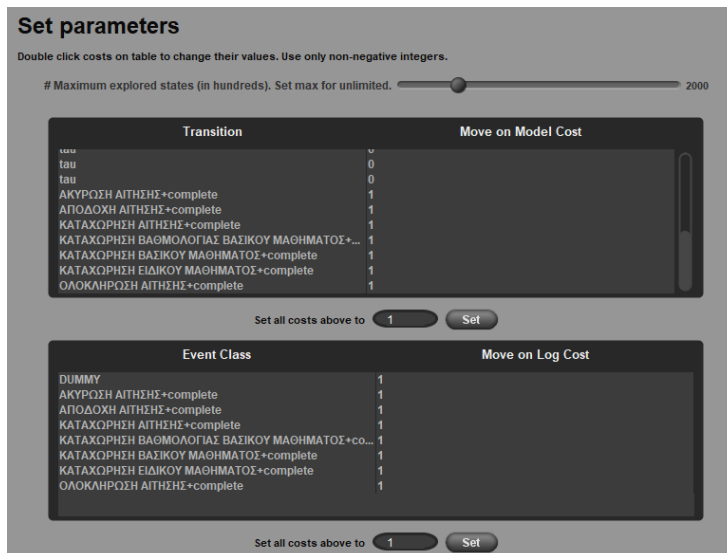
#### 5.4.1 Έλεγχος συμμόρφωσης του μοντέλου που προέκυψε από τον Inductive Miner με φίλτρο θορύβου 20%

Κατά την εφαρμογή του πρόσθετου “Replay a Log on Petri Net for Conformance Analysis” πάνω στο πρώτο αρχείο καταγραφής συμβάντων και στο μοντέλο που προέκυψε από την εκτέλεση του αλγορίθμου Inductive Miner με φίλτρο θορύβου 20%, επελέγη ως στόχος η μέτρηση της καταλληλότητας (measuring fitness) και η τιμωρία μη ολοκλήρωσης των διαδρομών (penalize improper completion), όπως φαίνεται στην παρακάτω εικόνα:



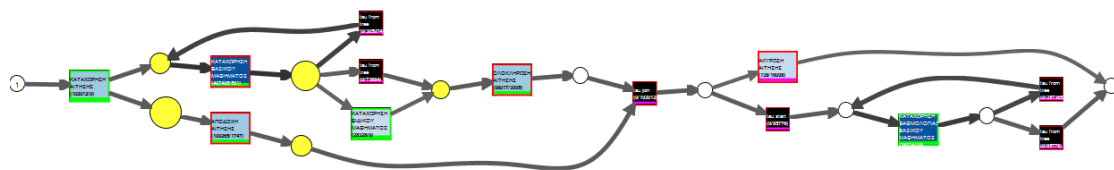
Εικόνα 59: Επιλογή αλγορίθμου ελέγχου συμμόρφωσης

Τα κόστη των αποκλίσεων μεταξύ αρχείου καταγραφής συμβάντων και μοντέλου παρέμειναν ισότιμα, στις προεπιλεγμένες τιμές που εμφανίζονται στην παρακάτω εικόνα, καθώς ο βασικός στόχος είναι η επίτευξη υψηλής καταλληλότητας κι όχι το να δοθεί κάποια βαρύτητα στις λανθάνουσες διαδρομές ή στις παρακάμψεις είτε του μοντέλου είτε του event log.

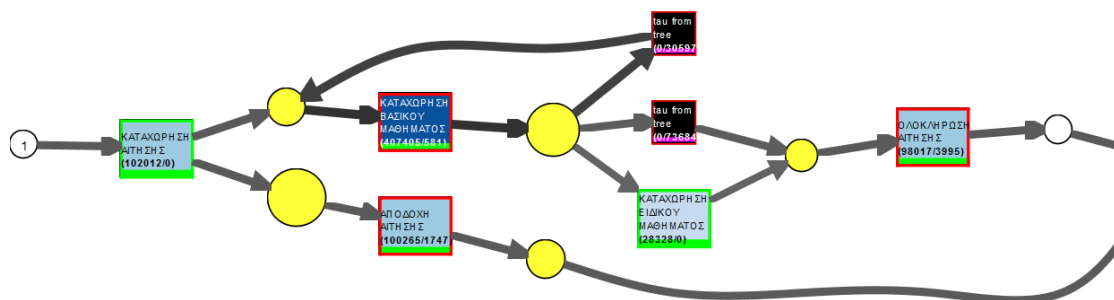


Εικόνα 60: Παραμετροποίηση κοστών ελέγχου συμμόρφωσης

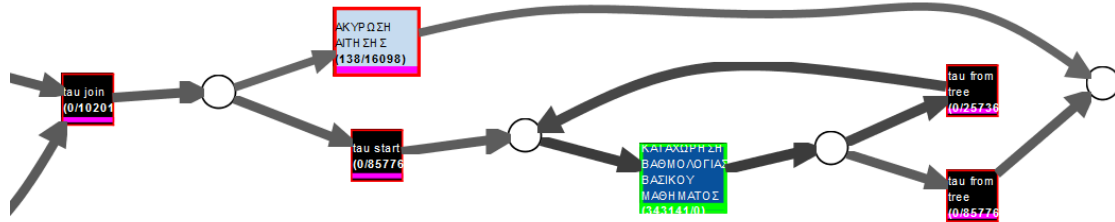
Με τις επιλογές αυτές, προκύπτει το παρακάτω Petri Net, το οποίο χωρίζεται από κάτω σε δύο επιπλέον εικόνες για καλύτερη προβολή του:



Εικόνα 61: Έλεγχος συμμόρφωσης πρώτου αρχείου καταγραφής συμβάντων με μοντέλο Inductive Miner και φίλτρο θορύβου 20%



Εικόνα 62: Έλεγχος συμμόρφωσης πρώτου αρχείου καταγραφής συμβάντων με μοντέλο Inductive Miner και φίλτρο θορύβου 20% (α' μέρος)



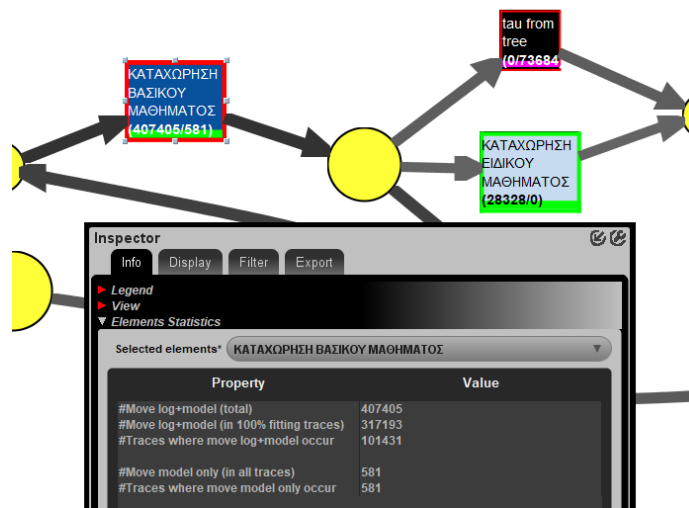
Εικόνα 63: Έλεγχος συμμόρφωσης πρώτου αρχείου καταγραφής συμβάντων με μοντέλο Inductive Miner και φίλτρο θορύβου 20% (β' μέρος)

Η διαβάθμιση στο μπλε χρώμα της κάθε δραστηριότητας υποδεικνύει τη συχνότητα με την οποία εκτελείται, όπου το πιο έντονο μπλε υποδηλώνει μεγαλύτερη συχνότητα. Στην παρούσα διαδικασία, με πιο σκούρο χρώμα εμφανίζονται οι δραστηριότητες που αφορούν στην ΚΑΤΑΧΩΡΗΣΗ ή στη ΒΑΘΜΟΛΟΓΗΣΗ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ, κάτι αναμενόμενο από τη στιγμή που για κάθε αίτηση-δήλωση υπάρχουν τέσσερα βασικά μαθήματα κι επομένως μιλάμε για δραστηριότητες που εκτελούνται με τετραπλάσια συχνότητα. Από τις υπόλοιπες δραστηριότητες, με παρόμοια συχνότητα εμφανίζονται οι δραστηριότητες ΚΑΤΑΧΩΡΗΣΗΣ, ΑΠΟΔΟΧΗΣ και ΟΛΟΚΛΗΡΩΣΗΣ ΑΙΤΗΣΗΣ, που ορίζουν τη βασική σειρά της διαδικασίας. Με μικρότερη συχνότητα εμφανίζεται η ΚΑΤΑΧΩΡΗΣΗ ΕΙΔΙΚΟΥ ΜΑΘΗΜΑΤΟΣ, καθώς αυτή είναι προαιρετική και με ακόμη μικρότερη η ΑΚΥΡΩΣΗ ΑΙΤΗΣΗΣ, κάτι όχι τόσο συνηθισμένο στη διαδικασία των αιτήσεων-δηλώσεων.

Γύρω από ορισμένες δραστηριότητες εμφανίζεται ένα κόκκινο περίγραμμα, που υποδηλώνει ότι για τη δραστηριότητα αυτή δεν υπάρχει πλήρης συμμόρφωση μεταξύ του μοντέλου και του event log. Οι δραστηριότητες για τις οποίες το μοντέλο και το event log συμμορφώνονται απόλυτα είναι η ΚΑΤΑΧΩΡΗΣΗ ΑΙΤΗΣΗΣ, η ΚΑΤΑΧΩΡΗΣΗ ΕΙΔΙΚΟΥ ΜΑΘΗΜΑΤΟΣ και η ΚΑΤΑΧΩΡΗΣΗ ΒΑΘΜΟΛΟΓΙΑΣ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ. Ειδικά για την τελευταία δραστηριότητα, αυτό είναι αρκετά σημαντικό, καθώς δεν φαίνεται να υπάρχει βαθμολόγηση πριν ολοκληρωθεί οριστικά η διαδικασία καταχώρησης των αιτήσεων-δηλώσεων στο πληροφοριακό σύστημα.

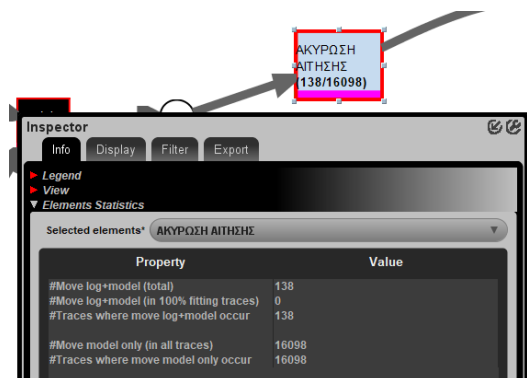
Για τις υπόλοιπες δραστηριότητες, στις οποίες εμφανίζεται το κόκκινο περίγραμμα, φαίνεται ότι τα συμβάντα του event log δεν εκτελέστηκαν σωστά, δηλαδή δεν υπήρχε απόλυτος συγχρονισμός με την απεικόνιση στο μοντέλο. Κάτω από κάθε δραστηριότητα εμφανίζεται μία πράσινη/φούξια μπάρα, που υποδεικνύει την αναλογία σωστών/λανθασμένων εκτελέσεων, ώστε να φαίνεται ποια είναι τελικά η συχνότητα των αποκλίσεων. Οι μπάρες αυτές δείχνουν η μεν πράσινη τη συχνότητα με την οποία η δραστηριότητα εκτελείται σωστά, δηλαδή υπάρχει συγχρονισμός μεταξύ μοντέλου και event log, και η δε φούξια τη συχνότητα με την οποία υπάρχει κίνηση μόνο στο μοντέλο κι όχι στο event log, επομένως πρόκειται για αποκλίνουσα συμπεριφορά.

Οι περισσότερες δραστηριότητες της διαδικασίας έχουν πράσινη μπάρα, που σημαίνει ότι είναι κατά πολύ περισσότερες οι σύγχρονες κινήσεις κι επομένως μιλάμε για πολύ μικρό ποσοστό αποκλίσεων από το μοντέλο. Αξιοσημείωτη είναι η περίπτωση των αποκλίσεων στην ΚΑΤΑΧΩΡΗΣΗ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ, για την οποία, όπως φαίνεται στην παρακάτω εικόνα, 407.405 φορές εκτελέστηκε σωστά αλλά 581 φορές εκτελέστηκε στο μοντέλο αλλά όχι και στα δεδομένα, μια απόκλιση που δε θα έπρεπε να υπάρχει, ούτε σε τόσο μικρό ποσοστό, καθώς η κάθε αίτηση του υποψηφίου κανονικά συνοδεύεται αυτομάτως από τη δήλωση ομάδας προσανατολισμού και άρα και βασικών μαθημάτων.



Εικόνα 64: Εμφάνιση αποκλίσεων στην ΚΑΤΑΧΩΡΗΣΗ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ

Φούξια μπάρα, δηλαδή μεγάλο ποσοστό αποκλίσεων, εμφανίζεται στη δραστηριότητα της ΑΚΥΡΩΣΗΣ ΑΙΤΗΣΗΣ, όπου, όπως φαίνεται στην παρακάτω εικόνα, 138 φορές η δραστηριότητα εκτελέστηκε συγχρονισμένα και σε 16.098 περιπτώσεις η δραστηριότητα αυτή δεν εμφανίστηκε στα ίχνη του event log.



Εικόνα 65: Εμφάνιση αποκλίσεων στην ΑΚΥΡΩΣΗ ΑΙΤΗΣΗΣ

Στον παρακάτω πίνακα παρουσιάζεται συνοπτικά για κάθε δραστηριότητα της διαδικασίας το πλήθος των συγχρονισμένων κινήσεων μεταξύ event log και μοντέλου και το πλήθος των κινήσεων που εμφανίστηκαν μόνο στο μοντέλο κι όχι στα δεδομένα:

<b>ΔΡΑΣΤΗΡΙΟΤΗΤΑ</b>	<b>ΠΛΗΘΟΣ ΣΥΓΧΡΟΝΙΣΜΕΝΩΝ ΚΙΝΗΣΕΩΝ</b>	<b>ΠΛΗΘΟΣ ΚΙΝΗΣΕΩΝ ΜΟΝΟ ΣΤΟ ΜΟΝΤΕΛΟ</b>
ΚΑΤΑΧΩΡΗΣΗ ΑΙΤΗΣΗΣ	102.012	0
ΚΑΤΑΧΩΡΗΣΗ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ	407.405	581
ΚΑΤΑΧΩΡΗΣΗ ΕΙΔΙΚΟΥ ΜΑΘΗΜΑΤΟΣ	28.328	0
ΑΠΟΔΟΧΗ ΑΙΤΗΣΗΣ	100.265	1.747
ΟΛΟΚΛΗΡΩΣΗ ΑΙΤΗΣΗΣ	98.017	3.995
ΑΚΥΡΩΣΗ ΑΙΤΗΣΗΣ	138	16.098
ΚΑΤΑΧΩΡΗΣΗ ΒΑΘΜΟΛΟΓΙΑΣ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ	343.141	0

Πίνακας 1: Πλήθος συγχρονισμένων κινήσεων μεταξύ μοντέλου Inductive Miner 20% και Event Log

Σε γενικές γραμμές τα νούμερα ανταποκρίνονται στη διαδικασία, καθώς η καταχώρηση βασικών μαθημάτων και η βαθμολόγηση αυτών είναι περίπου τετραπλάσια της καταχώρησης των αιτήσεων, με την βαθμολόγηση να υπολείπεται λίγο, υποδηλώνοντας τις περιπτώσεις αυτές των υποψηφίων όπου υπέβαλλαν αίτηση-δήλωση αλλά δεν εξετάστηκαν κι ως εκ τούτου δεν βαθμολογήθηκαν. Όπως είδαμε παραπάνω κατά τη διερεύνηση των επιλογών οπτικοποίησης, έτσι κι εδώ, δείχνει να υπάρχει μια σύγχυση με τις διαδικασίες ΑΠΟΔΟΧΗΣ και ΟΛΟΚΛΗΡΩΣΗΣ ΑΙΤΗΣΗΣ, οι οποίες εμφανίζουν ορισμένες αποκλίσεις που μάλλον προέρχονται από το γεγονός ότι στο event log που προέκυψε από το πληροφοριακό σύστημα δεν φαίνεται πάντα να τηρείται η σωστή σειρά στις διαδικασίες αυτές.

Οι κύκλοι που υποδηλώνουν τις μεταβάσεις μεταξύ των δραστηριοτήτων έχουν κίτρινο ή λευκό χρώμα. Οι κίτρινοι κύκλοι δείχνουν ότι κατά τη μετάβαση αυτή υπάρχουν διαδρομές που αποκλίνουν, υπάρχουν δηλαδή συμβάντα στο event log για τα οποία όταν βρισκόταν το token σε εκείνη τη μετάβαση ακολούθησε μια μη αναμενόμενη διαδρομή.

Επιλέγοντας την εμφάνιση κάποιων συνολικών στατιστικών, παίρνουμε το παρακάτω αποτέλεσμα:



ΔΡΑΣΤΗΡΙΟΤΗΤΑ	ΠΛΗΘΟΣ ΣΥΓΧΡΟΝΙΣΜΕΝΩΝ ΚΙΝΗΣΕΩΝ	ΠΛΗΘΟΣ ΚΙΝΗΣΕΩΝ ΜΟΝΟ ΣΤΟ ΜΟΝΤΕΛΟ
ΚΑΤΑΧΩΡΗΣΗ ΑΙΤΗΣΗΣ	102.012	0
ΚΑΤΑΧΩΡΗΣΗ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ	407.405	581
ΚΑΤΑΧΩΡΗΣΗ ΕΙΔΙΚΟΥ ΜΑΘΗΜΑΤΟΣ	28.328	0
ΑΠΟΔΟΧΗ ΑΙΤΗΣΗΣ	100.265	1.747
ΟΛΟΚΛΗΡΩΣΗ ΑΙΤΗΣΗΣ	98.017	3.995
ΑΚΥΡΩΣΗ ΑΙΤΗΣΗΣ	138	0
ΚΑΤΑΧΩΡΗΣΗ ΒΑΘΜΟΛΟΓΙΑΣ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ	343.141	0

Πίνακας 2: Πλήθος συγχρονισμένων κινήσεων μεταξύ μοντέλου Inductive Miner 10% και Event Log

Όσον αφορά τα γενικά στατιστικά, έχουμε μεγαλύτερη τιμή στην καταλληλότητα σε σχέση με το προηγούμενο μοντέλο, και συγκεκριμένα 0,98786.

Property	Value
Calculation Time (ms)	4.083999715719719
Num. States	93.67638121005372
<b>Trace Fitness</b>	<b>0.9878601886704421</b>
Title of Visualization	Alignments of XES Event Log...
Exit code of alignment for tra...	1.0
Model move cost empty trace	4.0
Number of LPs solved	1.8898953064345372
Queued States	115.41353958357841
Raw Fitness Cost	0.16079480845390737
Raw Model Fitness	0.004500407203039

Εικόνα 68: Συνολικά στατιστικά ελέγχου συμμόρφωσης (Inductive Miner – 10%)

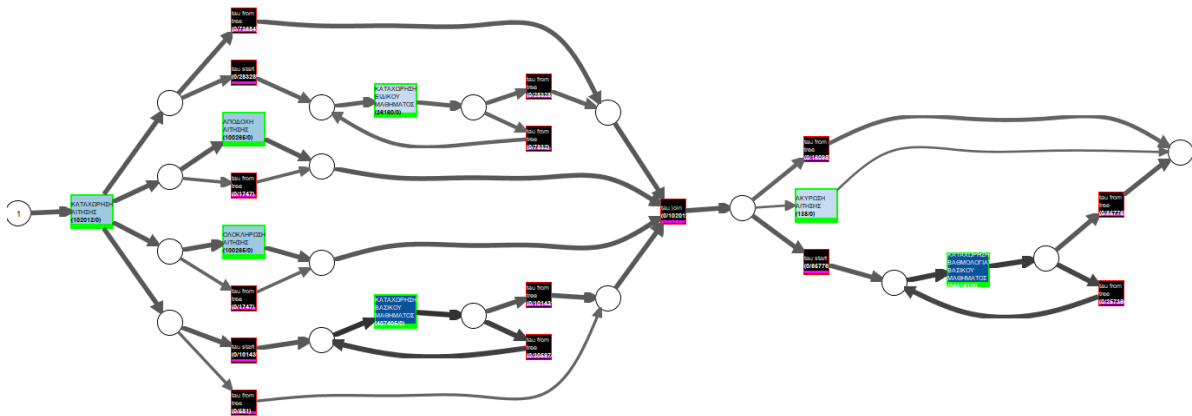
Τέλος εφαρμόζοντας το πρόσθετο “Measure Precision/Generalization” πάνω στο event log, στο μοντέλο και στο petrinet του ελέγχου συμμόρφωσης, έχουμε και τις τιμές των ποιοτικών παραμέτρων της ακρίβειας και της γενίκευσης, οι οποίες είναι 0,62164 και 0,99983 αντίστοιχα.

### 5.4.3 Έλεγχος συμμόρφωσης του μοντέλου που προέκυψε από τον Inductive Miner με φίλτρο θορύβου 0%

Κατά την εφαρμογή του πρόσθετου “Replay a Log on Petri Net for Conformance Analysis” πάνω στο πρώτο αρχείο καταγραφής συμβάντων και στο μοντέλο που προέκυψε από την εκτέλεση του αλγορίθμου Inductive Miner με φίλτρο θορύβου 0%,



χωρίς δηλαδή να φιλτράρονται οι μη συχνές κινήσεις, προκύπτει το παρακάτω Petri Net:



Εικόνα 69: Έλεγχος συμμόρφωσης πρώτου αρχείου καταγραφής συμβάντων με μοντέλο Inductive Miner και φίλτρο θορύβου 0%

Στην περίπτωση αυτή παρατηρούμε ότι δεν υπάρχει σε καμία δραστηριότητα κόκκινο πλαίσιο που σημαίνει ότι υπάρχει πλήρης συμμόρφωση μεταξύ του μοντέλου και του event log, δηλ. όλες οι κινήσεις γίνονται συγχρονισμένα. Αυτό συμβαίνει γιατί πλέον δεν φιλτράρεται καμία σπάνια κίνηση κι έτσι δημιουργείται ένα μοντέλο που έχει βασιστεί σε όλη την πληροφορία του event log. Όπως βλέπουμε πιο κάτω, η καταλληλότητα βρίσκεται στο 100% ωστόσο το μοντέλο που προέκυψε στερείται σε μεγάλο βαθμό απλότητας σε σχέση με τα προηγούμενα, καθώς περιλαμβάνει αρκετούς silent nodes οι οποίοι δημιουργούν πολύπλοκες δρομολογήσεις.

Global Statistics (non-filtered traces)	
Property	Value
Calculation Time (ms)	5.989371446496491
Num. States	322.2830353291769
<b>Trace Fitness</b>	<b>1.0</b>
Title of Visualization	Alignments of XES Event Log...
Exit code of alignment for tra...	1.0
Model move cost empty trace	1.0
Number of LPs solved	3.123730541504921
Queued States	384.3591734305768
Raw Fitness Cost	0.0
Move Model Fitness	1.0

Εικόνα 70: Συνολικά στατιστικά ελέγχου συμμόρφωσης (Inductive Miner – 0%)

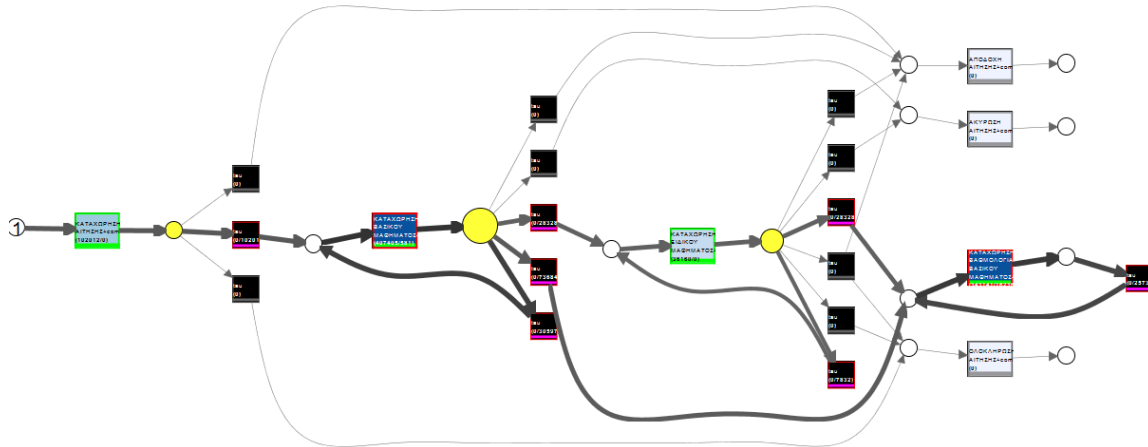
Οι τιμές των ποιοτικών παραμέτρων της ακρίβειας και της γενίκευσης είναι 0,38738 και 0,99941 αντίστοιχα.

#### 5.4.4 Έλεγχος συμμόρφωσης του μοντέλου που προέκυψε από τον Heuristics Miner

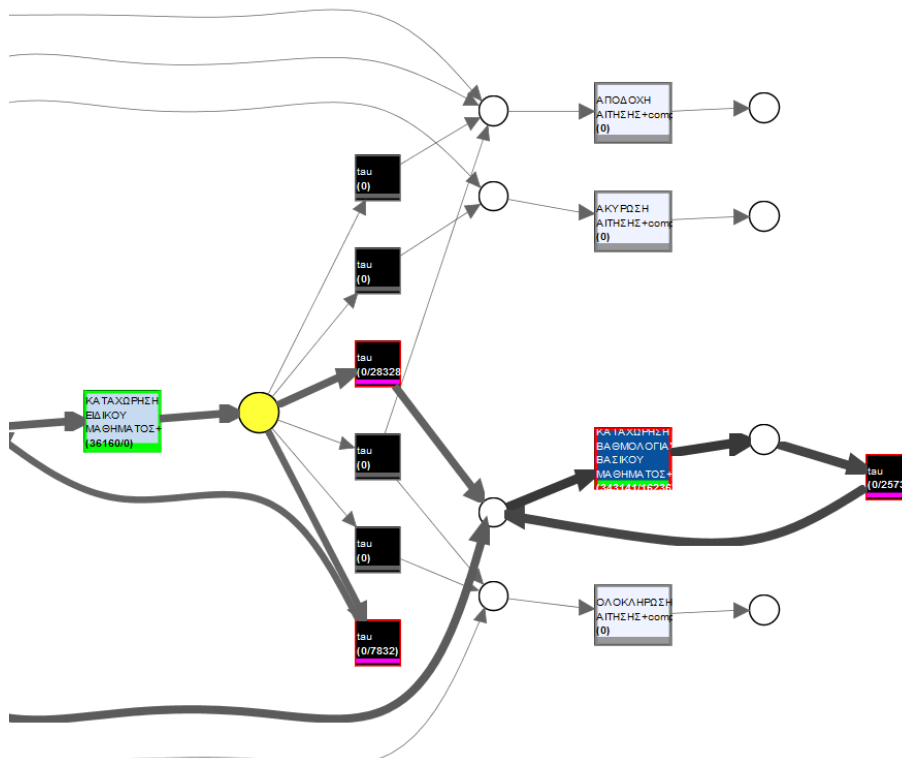
Πραγματοποιήθηκε τέλος ενδεικτικός έλεγχος συμμόρφωσης στο μοντέλο που προέκυψε από τον Heuristics Miner, το οποίο όπως είχε επισημανθεί και παραπάνω δεν ήταν sound μοντέλο, δηλ. δεν ολοκληρωνόταν σωστά. Κατά την εκτέλεση του

πρόσθετου “Replay a Log on Petri Net for Conformance Analysis” ζητήθηκε να οριστεί τερματικός κόμβος και ορίστηκε ο κόμβος που αφορά στην ΚΑΤΑΧΩΡΗΣΗ ΒΑΘΜΟΛΟΓΙΑΣ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ, που είναι συνήθως και η τελευταία δραστηριότητα της διαδικασίας μας.

Το Petri Net που δημιουργήθηκε απεικονίζεται παρακάτω, και πιο κάτω απεικονίζεται το τελευταίο του μέρος:



Εικόνα 71: Έλεγχος συμμόρφωσης πρώτου αρχείου καταγραφής συμβάντων με μοντέλο Heuristics Miner (α')

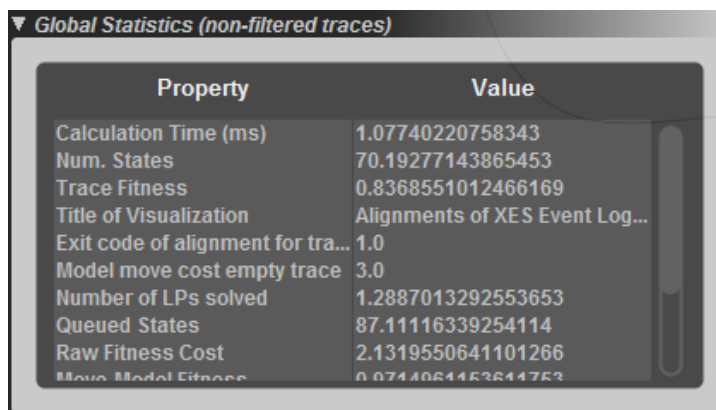


Εικόνα 72: Έλεγχος συμμόρφωσης πρώτου αρχείου καταγραφής συμβάντων με μοντέλο Heuristics Miner (β')

Όπως φαίνεται καλύτερα στην παραπάνω εικόνα, με το να οριστεί τερματικός κόμβος στο PetriNet έχουν μείνει εκτός διαδικασίας οι δραστηριότητες ΑΠΟΔΟΧΗΣ, ΑΚΥΡΩΣΗΣ

και ΟΛΟΚΛΗΡΩΣΗΣ ΑΙΤΗΣΗΣ, για τις οποίες φαίνεται να μην γίνονται καθόλου κινήσεις στο μοντέλο.

Στα γενικά στατιστικά, έχουμε αρκετά χαμηλότερη τιμή στην καταλληλότητα, και συγκεκριμένα 0,8368, ενώ εφαρμόζοντας το πρόσθετο “Measure Precision/Generalization” οι τιμές των ποιοτικών παραμέτρων της ακρίβειας και της γενίκευσης, οι οποίες είναι 0,37655 και 0,99998 αντίστοιχα.



Property	Value
Calculation Time (ms)	1.07740220758343
Num. States	70.19277143865453
Trace Fitness	0.8368551012466169
Title of Visualization	Alignments of XES Event Log...
Exit code of alignment for tra...	1.0
Model move cost empty trace	3.0
Number of LPs solved	1.2887013292553653
Queued States	87.11116339254114
Raw Fitness Cost	2.1319550641101266
Raw Model Fitness	0.9744064453844753

Εικόνα 73: Συνολικά στατιστικά ελέγχου συμμόρφωσης (Heuristics Miner)

#### 5.4.5 Παρατηρήσεις από τη διαδικασία του Conformance Checking

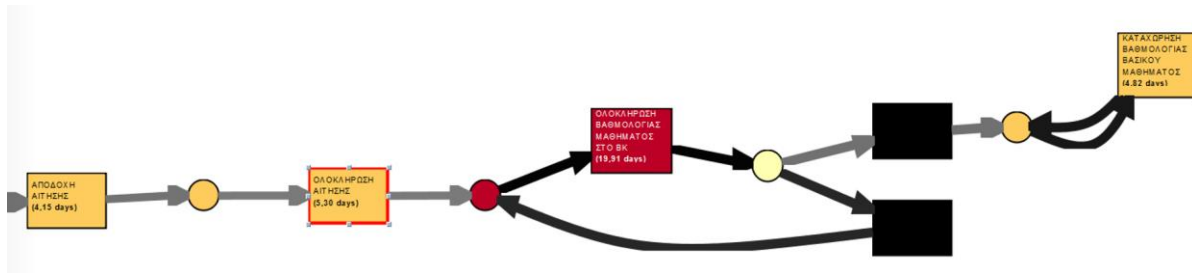
Από τον έλεγχο συμμόρφωσης είναι εμφανές ότι τα μοντέλα που δημιουργήθηκαν από την εφαρμογή του Inductive Miner αναπαράγουν σε μεγάλο βαθμό τα δεδομένα του αρχείου καταγραφής συμβάντων και έχουν μεγάλο βαθμό καταλληλότητας. Αντιθέτως, το μοντέλο που δημιουργήθηκε από τον Heuristics Miner, παρόλο που έδειχνε αρκετά αντιπροσωπευτικό, δεν ολοκληρώνεται σωστά και επιπλέον δεν υπάρχει η δυνατότητα να κάνουμε σωστό έλεγχο συμμόρφωσης από τη στιγμή που δεν υπάρχει σε αυτό τερματικός κόμβος.

Από τα τρία μοντέλα του Inductive Miner για τα οποία πραγματοποιήθηκε έλεγχος συμμόρφωσης, καταλήγουμε σε αυτό στο οποίο έχει επιλεγεί φίλτρο θορύβου 10%, επειδή συνδυάζει την απλότητα με μεγάλο βαθμό καταλληλότητας. Οι μη συγχρονισμένες κινήσεις που παρατηρήθηκαν αφορούν πολύ μικρό ποσοστό συμβάντων και αυτό που θα πρέπει να διερευνηθεί κυρίως είναι η ορθότητα της αλληλουχίας των δραστηριοτήτων ΑΠΟΔΟΧΗ και ΟΛΟΚΛΗΡΩΣΗ ΑΙΤΗΣΗΣ.

Η ανακάλυψη διαδικασιών επομένως ήταν επιτυχής και ανακαλύφθηκαν μοντέλα με πολύ καλές τιμές στις τέσσερις ποιοτικές παραμέτρους, η ανάλυση των οποίων δείχνει ότι κατά το μεγαλύτερο ποσοστό η διαδικασία των αιτήσεων-δηλώσεων εκτελείται και καταγράφεται σωστά στο πληροφοριακό σύστημα και δεδομένου του μεγάλου όγκου των δεδομένων υπάρχουν πολύ λίγες περιπτώσεις με αποκλίνουσες συμπεριφορές.

Αξίζει τέλος να σημειωθεί ότι έγινε προσπάθεια να γίνει έλεγχος συμμόρφωσης και στα μοντέλα που προέκυψαν από την εφαρμογή των αλγορίθμων ανακάλυψης στο δεύτερο αρχείο καταγραφής συμβάντων με την πληροφορία από τα ΒΚ, και αυτός δεν ήταν





Εικόνα 75: Ανάλυση απόδοσης δεύτερου αρχείου καταγραφής συμβάντων με την πληροφορία των ΒΚ, με μοντέλο Inductive Miner και φίλτρο θορύβου 20% ( $\beta'$ )

Οι δραστηριότητες ΑΠΟΔΟΧΗΣ και ΟΛΟΚΛΗΡΩΣΗΣ ΑΙΤΗΣΗΣ σηματοδοτούν το τέλος της διαδικασίας των αιτήσεων-δηλώσεων στο πληροφοριακό σύστημα, με μέσους χρόνους 4,15 και 5,30 μέρες αντίστοιχα. Οι 10 μέρες που προκύπτουν αθροιστικά είναι περίπου και το διάστημα που ορίστηκε από την υπουργική απόφαση ως διάστημα υποβολής αιτήσεων-δηλώσεων, επομένως η διαδικασία έχει εκτελεστεί χωρίς καθυστέρηση.

Η ΟΛΟΚΛΗΡΩΣΗ ΒΑΘΜΟΛΟΓΗΣΗΣ ΣΤΑ ΒΚ έχει μεγαλύτερο χρόνο αναμονής, ο οποίος επίσης δικαιολογείται βάσει της οργανωτικής διαδικασίας που περιγράφεται παραπάνω.

Αυτό που θα μπορούσε να διορθωθεί, είναι ο χρόνος που απαιτείται προκειμένου η βαθμολογία των μαθημάτων από τα Βαθμολογικά Κέντρα να καταχωρηθεί στο πληροφοριακό σύστημα, ο οποίος είναι περίπου 5 ημέρες όπως φαίνεται στη δραστηριότητα ΚΑΤΑΧΩΡΗΣΗΣ ΒΑΘΜΟΛΟΓΙΑΣ ΒΑΣΙΚΟΥ ΜΑΘΗΜΑΤΟΣ. Αυτός ο χρόνος θα μπορούσε να ελαχιστοποιηθεί εάν υπήρχε ένα ενιαίο πληροφοριακό σύστημα στο οποίο θα έχουν πρόσβαση και τα βαθμολογικά κέντρα και οι σχολικές μονάδες ταυτόχρονα.

## 5.6 Σύνθεση συμπερασμάτων

Από την έλλειψη διαθέσιμων δεδομένων για τη συνολική διαδικασία των Πανελλαδικών Εξετάσεων και από την παραπάνω διαδικασία της εξόρυξης στα δεδομένα που τελικά δόθηκαν, είναι φανερό ότι παρατηρείται μια κατάτμηση των δεδομένων που αφορούν στη διαδικασία αυτή και το κάθε Τμήμα του Υπουργείου Παιδείας και Θρησκευμάτων διαθέτει τα δικά του στοιχεία, ανάλογα με τις αρμοδιότητές του, τα οποία και διαθέτει στα υπόλοιπα Τμήματα προκειμένου να υπάρξει η απαραίτητη σύνδεση που θα υποστηρίξει ηλεκτρονικά τη συνολική διαδικασία. Αυτό δυσχεραίνει την συγκέντρωση και την ενοποίησή τους, καθώς απαιτούνται επιπλέον εργασίες της εξαγωγής δεδομένων από κάθε Τμήμα, κάτι που δεν είναι πάντα εφικτό.

Αξίζει να σημειωθεί ότι την παρούσα χρονική στιγμή η καταγραφή και ηλεκτρονική καταχώρηση των δεδομένων των διαδικασιών των Πανελλαδικών Εξετάσεων πραγματοποιείται σε τέσσερα διαφορετικά πληροφοριακά συστήματα: την εφαρμογή

ΠΣ-MySchool για τα στοιχεία των υποψηφίων και των αιτήσεων-δηλώσεων, την εφαρμογή των βαθμολογικών κέντρων για την καταχώρηση των βαθμολογιών των μαθημάτων για τον κάθε υποψήφιο, την εφαρμογή «Σύστημα Ηλεκτρονικού Μηχανογραφικού» για την καταχώριση του μηχανογραφικού δελτίου του κάθε υποψηφίου με την καταγραφή των επιλογών του και τέλος την εφαρμογή «Ηλεκτρονικές Εγγραφές Επιτυχόντων» για την ηλεκτρονική εγγραφή του κάθε υποψηφίου στη σχολή επιτυχίας του στην Τριτοβάθμια Εκπαίδευση.

Τα παραπάνω πληροφοριακά συστήματα διαθέτουν εξαρτήσεις μεταξύ τους, προκειμένου να μεταφέρονται δεδομένα από το ένα στο άλλο, πράγμα που ωστόσο καθιστά τη συνολική διαδικασία πιο πολύπλοκη και χρονοβόρα από άποψης υποστήριξης και διαχείρισης. Επιπροσθέτως, αυτός ο διασκορπισμός των δεδομένων σε βάσεις από διαφορετικά πληροφοριακά συστήματα σε συνδυασμό με το μεγάλο όγκο τους λόγω του πολύ μεγάλου πλήθους των υποψηφίων, καθιστά τη διαδικασία εξόρυξης αρκετά δύσκολη όπως αναφέρθηκε και παραπάνω καθώς, ακόμη και αν παρέχονται όλα τα δεδομένα από όλες τις πλατφόρμες, απαιτείται χρονοβόρα προεπεξεργασία τους καταρχάς για να έρθουν στην κατάλληλη μορφή αλλά κυρίως για να δημιουργηθούν σε μορφή αρχείου καταγραφής συμβάντων οι εξαρτήσεις που θα κάνουν το σύνολο των δεδομένων αξιοποιήσιμο από μια εφαρμογή εξόρυξης διαδικασιών.

Για όλους τους παραπάνω λόγους, θα ήταν καλό να εξεταστεί μελλοντικά είτε η δημιουργία ενός νέου συστήματος που θα γεφυρώσει την επικοινωνία των υπαρχόντων με αυτοματοποιημένο τρόπο είτε η ενοποίηση των προαναφερθέντων πληροφοριακών συστημάτων σε ένα καινούριο, ενισχύοντας έτσι τις αρχές της διαλειτουργικότητας και δημιουργώντας εντέλει ένα σύστημα πληρέστερο όσον αφορά στην πληροφορία που θα καταχωρείται, θα επεξεργάζεται και θα παρέχεται εάν ζητηθεί με μορφή πιο ολοκληρωμένων συνόλων εκπαιδευτικών δεδομένων (datasets), προκειμένου να είναι και σε θέση να εξορυχτούν καλύτερα με διάφορες μεθόδους και εργαλεία. Υπάρχει ανάγκη λοιπόν για την ύπαρξη ενός συστήματος υποστήριξης της διαδικασίας των Πανελλαδικών Εξετάσεων προσανατολισμένου στην εξόρυξη διαδικασιών ώστε να μπορεί να γίνεται έλεγχος της σωστής λειτουργίας όλων των διεργασιών που λαμβάνουν χώρα σε όλη τη διάρκεια των εξετάσεων, ακριβώς επειδή αυτές είναι πολλές και έχουν συγκεκριμένο πλαίσιο. Με τον τρόπο αυτό θα είναι πιο εύκολο να πραγματοποιείται συχνός έλεγχος του κατά πόσο το πλαίσιο αυτό τηρείται και του τι μπορεί να τροποποιηθεί και να βελτιωθεί προκειμένου να βελτιστοποιηθεί η λειτουργία του συνόλου των διαδικασιών.

Παρακάτω απεικονίζεται μία απλή και αφαιρετική διαγραμματική αναπαράσταση του πώς θα μπορούσαν να συνδέονται οι πίνακες με τα δεδομένα των διαδικασιών των Πανελλαδικών Εξετάσεων, όπως έχει δημιουργηθεί σε σχεσιακό περιβάλλον MS-Access:







## 6. ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΜΕΛΛΟΝΤΙΚΗ ΕΡΓΑΣΙΑ

Η Εξόρυξη Διαδικασιών (Process Mining) είναι μια σχετικά νέα τεχνική η οποία αρχικά δημιουργήθηκε προκειμένου να γίνεται διάγνωση των πραγματικών διαδικασιών που λαμβάνουν χώρα σε μια εταιρεία ή έναν οργανισμό, των λόγων για τους οποίους συμβαίνουν αυτές, των πιθανών αποκλίσεων και συμφορήσεων κατά την εκτέλεσή τους και τέλος να δημιουργείται ένα κατάλληλο μοντέλο που θα περιγράφει όλα τα παραπάνω. Στην παρούσα εργασία εφαρμόστηκαν τεχνικές εξόρυξης διαδικασιών στη διαδικασία υποβολής αιτήσεων-δηλώσεων και καταχώρησης βαθμολογιών των υποψηφίων των Πανελλαδικών Εξετάσεων.

Η διαδικασία των αιτήσεων-δηλώσεων των Πανελλαδικών Εξετάσεων δεν είναι μια περίπλοκη διαδικασία, αλλά είναι μια διαδικασία που υπάγεται σε συγκεκριμένο πλαίσιο. Σε όρους εξόρυξης διαδικασιών ανήκει στις επονομαζόμενες *Isagna processes*, όπου με μικρή προσπάθεια μπορεί να παραχθεί ένα μοντέλο διεργασιών που να συμφωνεί και να έχει μεγάλο συντελεστή καταλληλότητας (*fitness*), χωρίς να χρειαστεί κάποιο φιλτράρισμα στα δεδομένα. Η εκτέλεση της διεργασίας είναι ήδη δομημένη και αυτό διαφαίνεται και στα δεδομένα που υπάρχουν στο αρχείο καταγραφής συμβάντων. Από την άλλη, λόγω του γεγονότος ότι πρόκειται για μια δομημένη διεργασία, κάθε περαιτέρω γνώση και παρατήρηση από την ανάλυση των δεδομένων χρησιμοποιώντας τεχνικές εξόρυξης διαδικασιών ή άλλες τεχνικές θα μπορεί να έχει σχετικά μικρή επίδραση σε μελλοντικές διεργασίες. Αποτελεί μια διεργασία με ξεκάθαρους κανόνες και ροή από τη μία δραστηριότητα στην άλλη κι αυτό που μπορούμε κυρίως να εξετάσουμε είναι αν ο πολύ μεγάλος όγκος των δεδομένων συμφωνεί με αυτή τη ροή και αν υπάρχουν σοβαρές αποκλίσεις και καθυστερήσεις κατά την εκτέλεση της διαδικασίας.

Μετά από την εφαρμογή των τεχνικών εξόρυξης πραγματοποιήθηκε επιτυχής ανακάλυψη αντιπροσωπευτικών μοντέλων διεργασιών και παράλληλα χρησιμοποιήθηκαν διάφορα απεικονιστικά εργαλεία που οδήγησαν στη δημιουργία μιας κατανοητής μορφής της διαδικασίας. Ο αλγόριθμος *Inductive Miner* του εργαλείου *ProM* κατάφερε να εξάγει ένα ορθό μοντέλο, το οποίο στη συνέχεια επαληθεύτηκε με τον έλεγχο συμμόρφωσης, κάτι που δεν κατάφεραν οι αλγόριθμοι *Alpha* και *Heuristics Miner*. Επίσης, ο αλγόριθμος *Fuzzy Miner* και το εργαλείο *Disco* πραγματοποίησαν μια ορθή απεικόνιση της διαδικασίας με τη μορφή χάρτη όπου επιλέγοντας παραμέτρους υπάρχει η δυνατότητα άμεσης εποπτείας είτε πιο λεπτομερούς ή πιο απλοποιημένης μορφής της.

Η έλλειψη των δεδομένων που τελικά παραδόθηκαν και η επεξεργασία αυτών κατέδειξε την αναγκαιότητα ύπαρξης ενός ενιαίου πληροφοριακού συστήματος, προσανατολισμένου στο *process mining*, το οποίο θα υποστηρίζει τη συνολική διαδικασία οργάνωσης των Πανελλαδικών Εξετάσεων κι έτσι θα υπάρχει δυνατότητα εξαγωγής όλων των απαραίτητων δεδομένων σε συγκεντρωτική μορφή. Ήταν επίσης εμφανές ότι είναι απαραίτητη η δυνατότητα πρόσβασης και άντλησης καθώς και η εντύπωση στα δεδομένα αυτά, καθώς δεν ήταν εύκολη η επικοινωνία των

διαφορετικών Τμημάτων του ΥΠΑΙΘ προκειμένου να εξάγουν και να αντιστοιχήσουν τα δεδομένα από τα διαφορετικά συστήματα. Επιπροσθέτως, παρατηρήθηκε στα δεδομένα απουσία χρονικών σημείων που να σηματοδοτούν την αρχή και το τέλος των δραστηριοτήτων, κάτι που θα έδινε τη δυνατότητα προσδιορισμού της διάρκειάς τους και μιας πιο σωστής ανάλυσης απόδοσης που θα εντόπιζε πιο εύκολα τις πιθανές συμφορήσεις. Τέλος, η ανάλυση των αποτελεσμάτων που προέκυψαν ανέδειξε την έλλειψη ορισμένων ελέγχων στο πληροφοριακό σύστημα οι οποίοι δεν θα πρέπει να επιτρέπουν στους χρήστες να παραβιάζουν τη σειρά αυστηρά καθορισμένων δραστηριοτήτων.

Η επιλογή της διαδικασίας των Πανελλαδικών Εξετάσεων η οποία παρουσιάζει χαμηλό βαθμό πολυπλοκότητας, ήταν ένα έναυσμα για την ενασχόληση με την εξόρυξη διαδικασιών και την απόκτηση κάποιας εμπειρίας η οποία στο μέλλον θα μπορούσε να χρησιμοποιηθεί προκειμένου να πραγματοποιηθεί εκτενέστερη ανάλυση, με περισσότερα δεδομένα και για περισσότερα σχολικά έτη για να μπορεί έτσι να γίνει και σύγκριση μεταξύ των ετών. Το μεγαλύτερο ενδιαφέρον που παρουσιάζει αυτή η διαδικασία άλλωστε είναι ο τεράστιος όγκος των δεδομένων, ο οποίος δεν αφήνει περιθώρια για ελλιπή ανάλυση. Αυτή είναι και η μελλοντική πρόκληση επομένως, η ύπαρξη ενός ενιαίου συστήματος που θα επιτρέψει την εξαγωγή δεδομένων που θα αφορούν ολόκληρη τη διαδικασία των Πανελλαδικών Εξετάσεων κι όχι μόνο αυτή της αίτησης-δήλωσης και κατ' επέκταση η εφαρμογή των τεχνικών εξόρυξης στα δεδομένα αυτά, κάτι που θα μας δώσει τη δυνατότητα εξαγωγής συμπερασμάτων για το σύνολο της διαδικασίας κι όχι μόνο για ένα μέρος της. Αξίζει ωστόσο να σημειωθεί εδώ ότι η χρήση του εργαλείου ProM απαιτεί μεγάλη υπολογιστική ισχύ, καθώς όταν έγινε προσπάθεια να εισαχθούν δεδομένα αιτήσεων-δηλώσεων έξι (6) ετών δεν ήταν δυνατό το ProM να προχωρήσει στη δημιουργία αρχείου καταγραφής συμβάντων. Το Disco αντίθετα είναι αρκετά πιο ελαφρύ πρόγραμμα και η εμπορική του έκδοση έχει τη δυνατότητα να διαχειριστεί πολύ μεγάλα αρχεία δεδομένων, ενώ η ακαδημαϊκή φίλτραρε κάποιο ποσοστό αυτών.

Συμπεραίνουμε εντέλει ότι η εξόρυξη διαδικασιών είναι μια ισχυρή τεχνική όπου με την ορθή χρήση της μπορεί να εφαρμοστεί σε πολλαπλούς τομείς και να μοντελοποιήσει διάφορες διαδικασίες ώστε να γίνει πιο κατανοητή η συμπεριφορά τους. Αυτό με τη σειρά του θα οδηγήσει στη βελτιστοποίηση των διαδικασιών αυτών ή των πληροφοριακών συστημάτων που τις υποστηρίζουν με στόχο την ευθυγράμμιση του τι πραγματικά συμβαίνει με αυτό που ορίζεται από το νομοθετικό ή το κανονιστικό πλαίσιο.

Τέλος, δεδομένου ότι η διαδικασία των Πανελλαδικών Εξετάσεων είναι μια νευραλγική διαδικασία που αφορά μεγάλο όγκο ανθρώπινου δυναμικού, θα ήταν καλό να συσταθεί από το Υπουργείο Παιδείας και Θρησκευμάτων μια ομάδα εργασίας με γνώσεις σχεδιασμού και ανάλυσης διαδικασιών αλλά και διαχείρισης πληροφοριακών συστημάτων και δεδομένων, η οποία θα εφαρμόσει τεχνικές και εργαλεία εξόρυξης διαδικασιών για να καταθέσει στη συνέχεια βελτιωτικές προτάσεις στο σύνολο της διαδικασίας, προς όφελος όλων των εμπλεκόμενων σε αυτή.

## 7. ΒΙΒΛΙΟΓΡΑΦΙΑ - ΑΝΑΦΟΡΕΣ

1. Aalst, W. Van Der 2011. Process Mining: Discovery, Conformance and Enhancement of Business Processes. Springer-Verlag, Berlin
2. van der Aalst W. (2016) Data Science in Action. In: Process Mining. Springer, Berlin, Heidelberg
3. W.M.P. van der Aalst, A.J.M.M. Weijters, Process mining: a research agenda, Computers in Industry, Volume 53, Issue 3, 2004, Pages 231-244, ISSN 0166-3615, <https://doi.org/10.1016/j.compind.2003.10.001>.
4. W. M. P. van der Aalst and C. W. Gunther, "Finding Structure in Unstructured Processes: The Case for Process Mining," *Seventh International Conference on Application of Concurrency to System Design (ACSD 2007)*, Bratislava, 2007, pp. 3-12, doi: 10.1109/ACSD.2007.50.
5. Reimann, Peter & Frerejean, Jimmy & Thompson, Kate. (2009). Using process mining to identify models of group decision making in chat data. Computer Supported Collaborative Learning Practices, CSCL 2009 Conference Proceedings - 9<sup>th</sup> International Conference. 98-107. 10.3115/1600053.1600068
6. Trcka, Nikola & Pechenizkiy, Mykola. (2009). From Local Patterns to Global Models: Towards Domain Driven Educational Process Mining. ISDA 2009 - 9th International Conference on Intelligent Systems Design and Applications. 1114-1119. 10.1109/ISDA.2009.159.
7. Southavilay, Vilaythong & Yacef, Kalina & Calvo, Rafael. (2010). Process Mining to Support Students' Collaborative Writing. 257-266.
8. Baradwaj, Brijesh & Pal, Saurabh. (2011). Mining Educational Data to Analyze Students' Performance. International Journal of Advanced Computer Science and Applications. 2. 63-69. 10.14569/IJACSA.2011.020609.
9. van der Aalst, W., et al.: Process mining manifesto. In: Proceedings of BPM International Workshops, vol. 99, pp. 169–194 (2012)
10. Wil van der Aalst. 2012. Process Mining: Overview and Opportunities. ACM Trans. Manage. Inf. Syst. 3, 2, Article 7 (July 2012), 17 pages. DOI:<https://doi.org/10.1145/2229156.2229157>
11. Janes, Andrea & Maggi, Fabrizio & Marrella, Andrea & Montali, Marco. (2017). From Zero to Hero: A Process Mining Tutorial. 625-629. 10.1007/978-3-319-69926-4\_55.
12. Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. & Byers, A. H. (2011). Big Data: The Next Frontier for Innovation, Competition, and Productivity (). McKinsey Global Institute .
13. Trcka, N., Pechenizkiy, M., & Aalst, van der, W. M. P. (2011). Process mining from educational data. In C. Romero, S. Ventura, M. Pechenizkiy, & R. Baker (Eds.), *Handbook of Educational Data Mining* (pp. 123-142). (Chapman & Hall/CRC Data Mining and Knowledge Discovery Series). CRC Press.

14. Anuwatvisit S, Tungkasthan A, Premchaiswadi W. Bottleneck mining and Petri net simulation in education situations. In: ICT and Knowledge Engineering (ICT & Knowledge Engineering), 10th International Conference. IEEE; Bangkok, Thailand; 2012, 244–251.
15. Fernández-Gallego B, Lama M, Vidal JC, Mucientes M. Learning analytics framework for educational virtual worlds. *Proc Comput Sci* 2013, 25:443–447.
16. Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 3.
17. Cairns AH, Gueni B, Fhima M, Cairns A, David S, Khelifa N. Towards custom-designed professional training contents and curriculums through educational process mining. In: The Fourth International Conference on Advances in Information Mining and Management. Paris, France; 2014, 53–58.
18. Günther, C. W., & Rozinat, A. (2012). Disco: discover your processes. In N. Lohmann, & S. Moser (Eds.), *Proceedings of the Demonstration Track of the 10th International Conference on Business Process Management (BPM 2012)* (pp. 40-44). (CEUR Workshop Proceedings; Vol. 940). CEUR-WS.org. <http://ceur-ws.org/Vol-940/>
19. Gunther, C. W., & Verbeek, H. M. W. (2014). *XES - standard definition*. (BPM reports; Vol. 1409). BPMcenter. org.
20. Bogarín, Alejandro & Romero, Cristóbal & Cerezo, Rebeca & Sánchez-Santillán, Miguel. (2014). Clustering for improving Educational process mining. *ACM International Conference Proceeding Series*. 11-15. 10.1145/2567574.2567604.
21. Barreiros BV, Lama M, Mucientes M, Vidal JC. Softlearn: a process mining platform for the discovery of learning paths. In: 14th International Conference on Advanced Learning Technologies. IEEE, Athens, Greece; 2014, 373–375.
22. van Eck M.L., Lu X., Leemans S.J.J., van der Aalst W.M.P. (2015) PM2: A Process Mining Project Methodology. In: Zdravkovic J., Kirikova M., Johannesson P. (eds) *Advanced Information Systems Engineering. CAiSE 2015. Lecture Notes in Computer Science*, vol 9097. Springer, Cham. [https://doi.org/10.1007/978-3-319-19069-3\\_19](https://doi.org/10.1007/978-3-319-19069-3_19)
23. Mukala, P., Buijs, J. C. A. M., & Aalst, van der, W. M. P. (2015). Uncovering learning patterns in a MOOC through conformance alignments. (BPM reports; Vol. 1509). BPMcenter. org.
24. Cairns, A.H., Gueni, B., Fhima, M., Cairns, A., David, S., Khelifa, N., & Dautier, P. (2015). Process Mining in the Education Domain. *International journal on advances in intelligent systems*, 8, 219-232.
25. Wang R, Zaïane OR. Discovering process in curriculum data to provide recommendation. In: *Proceedings of the 5th International Conference on Educational Data Mining*, Madrid, Spain, 2015, 580–581.
26. "IEEE Standard for eXtensible Event Stream (XES) for Achieving Interoperability in Event Logs and Event Streams," in *IEEE Std 1849-2016*, vol., no., pp.1-50, 11 Nov. 2016, doi: 10.1109/IEEESTD.2016.7740858.

27. Romero, C., Cerezo, R., Bogarín, A. and Sánchez-Santillán, M. (2016). EDUCATIONAL PROCESS MINING. In Data Mining and Learning Analytics (eds S. ElAtia, D. Ipperciel and O.R. Zaïane).<https://doi.org/10.1002/9781118998205.ch1>
28. Bogarín, Alejandro & Cerezo, Rebeca & Romero, Cristóbal. (2017). A survey on educational process mining. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 8. 10.1002/widm.1230.
29. Tóth K, Rölke H, Goldhammer F, Barkow I. Educational process mining: new possibilities for understanding students' problem-solving skills. In: Educational Research and Innovation. Paris, France: OECD Publishing; 2017, 193–209.
30. Kashyap, Diksha. "Educational administration: Meaning, nature and other details." (2020).
31. 'Introduction into Process Mining with ProM', online course, Eindhoven University of Technology