



ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

ΑΝΤΙΣΤΟΙΧΙΣΗ ΜΟΝΤΕΛΩΝ ΕΠΙΧΕΙΡΗΣΙΑΚΩΝ ΔΙΑΔΙΚΑΣΙΩΝ

ΣΗΜΕΙΩΣΕΙΣ ΓΙΑ ΤΟ ΜΑΘΗΜΑ

ΤΕΧΝΟΛΟΓΙΑ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΙΑΚΩΝ ΔΙΑΔΙΚΑΣΙΩΝ

Α. ΤΣΑΛΓΑΤΙΔΟΥ, Χ. ΤΣΑΓΚΑΝΗ

2020

Πρόλογος

Το παρόν σύγγραμμα παρέχει συνοδευτικό υλικό για τη διευκόλυνση της κατανόησης θεμάτων που παρουσιάστηκαν σε μεταπτυχιακούς φοιτητές, στη διάλεξη που αφορούσε την Αντιστοίχιση Επιχειρησιακών Διαδικασιών.

Με το παρόν επιχειρείται μία συνοπτική, αλλά συγχρόνως συμπαγής παρουσίαση όλων των εννοιών που θα διευκολύνουν τον αναγνώστη-φοιτητή να κατανοήσει τη σημασία της Αντιστοίχισης Επιχειρησιακών Διαδικασιών και τις τεχνικές που χρησιμοποιούνται. Επίσης ενδεικτικά παρουσιάζονται κάποιες υλοποιήσεις που προσφέρουν την αντιστοίχιση επιχειρησιακών διαδικασιών χρησιμοποιώντας διαφορετικές τεχνικές.

Καθώς η εξοικείωση με τις έννοιες της Αντιστοίχισης Επιχειρησιακών Διαδικασιών δεν είναι προαπαιτούμενη, αυτές εισάγονται, όπου αυτό απαιτείται και στο ανάλογο επίπεδο ανάλυσης με απώτερο στόχο τη διευκόλυνση του αναγνώστη-φοιτητή κατά την κατανόηση των εννοιών-τεχνικών της συγκεκριμένης θεματικής ενότητας.

Περιεχόμενα

1.	Εισαγωγή	3
2.	Σημασία Αντιστοίχισης Επιχειρησιακών Διαδικασιών.....	3
2.1	Σημερινή Κατάσταση.....	3
2.2	Τομείς εφαρμογής-Αναμενόμενα Οφέλη	4
3.	Διαφορές σε Επιχειρησιακές Διαδικασίες	5
4.	Αντιστοίχιση Επιχειρησιακών Διαδικασιών - Τεχνικές Αντιστοίχισης	6
4.1	Τεχνικές Αντιστοίχισης Επιχειρησιακών Διαδικασιών - Μετρικές Ομοιότητας.....	7
4.2	Ομοιότητα Ονομασιών.....	8
4.3	Δομική Ομοιότητα.....	11
4.4	Ομοιότητα Προφίλ Συμπεριφοράς (Behavioural Profile Similarity)	16
5.	Εργαλεία.....	20
5.1	Τεχνική «Triple-S»	20
5.2	Τεχνική «RefMod-Mine/NSCM».....	21
5.3	Ανάλυση των Εργαλείων και Ευρήματα.....	23

1. Εισαγωγή

Η αντιστοίχιση επιχειρησιακών διαδικασιών, έχει ως απώτερο στόχο να εκμεταλλευτεί τη γνώση που υπάρχει ήδη, τόσο σε διαφορετικές επιχειρησιακές διαδικασίες, όσο και σε παραλλαγές της ίδιας επιχειρησιακής διαδικασίας ως μια πηγή πληροφοριών και να χρησιμοποιηθεί κατά τη δημιουργία νέων διαδικασιών ή τη βελτιστοποίηση υφισταμένων επιχειρησιακών διαδικασιών.

2. Σημασία Αντιστοίχισης Επιχειρησιακών Διαδικασιών

2.1 Σημερινή Κατάσταση

Σήμερα οι επιχειρησιακές διαδικασίες υπάρχουν διάσπαρτες σε διάφορες μηχανογραφικές εφαρμογές μέσα σε έναν οργανισμό και αποτυπώνονται χρησιμοποιώντας διαφορετικές φορμαλιστικές μεθόδους (πχ. BPMN, BPEL κτλ.). Ο εξαιρετικά μεγάλος όγκος των επιχειρησιακών διαδικασιών μέσα σε έναν οργανισμό θέτει επιτακτική την ανάγκη της σωστής διαχείρισης τους.

Στο πλαίσιο της διαχείρισης των επιχειρησιακών διαδικασιών εντάσσεται και η ανεύρεση επιχειρησιακών διαδικασιών με απώτερο στόχο: την παροχή πολύτιμων πληροφοριών για τη βελτιστοποίηση των υπαρχουσών διαδικασιών, την επαναχρησιμοποίηση των διαδικασιών επιλύοντας λειτουργικά προβλήματα και την ανακάλυψη προτιμώμενων πρακτικών. Επισημαίνεται ότι η ανεύρεση μοντέλων διαδικασιών μέσα σε έναν οργανισμό και η προσαρμογή τους προκειμένου να καλύψουν τις τρέχουσες ανάγκες των επιχειρήσεων έχει αποδειχτεί ως πιο αποτελεσματική και λιγότερο επιρρεπής σε λάθη πρακτική, συγκρινόμενη με την ανάπτυξη μοντέλων διαδικασιών από το μηδέν.

Η διαδικασία της ανεύρεσης επιχειρησιακών διαδικασιών ή μέρος αυτών μπορεί να υποστηριχτεί μέσω της αντιστοίχισης μοντέλων επιχειρησιακών διαδικασιών. Σημειώνεται ότι αντιστοίχιση γραφικών παραστάσεων (πχ. μοντέλα επιχειρησιακών διαδικασιών) είναι η διαδικασία της ανακάλυψης συσχετίσεων ανάμεσα σε γραφικές παραστάσεις μέσω της εφαρμογής ενός αλγόριθμου αντιστοίχισης.

Αξίζει να σημειωθεί ότι σήμερα, υπάρχουν διαφορετικές μηχανές αναζήτησης που εφαρμόζουν ανάκτηση διαδικασιών από υπάρχοντα αποθετήρια και μπορούν να ταξινομηθούν χρησιμοποιώντας διαφορετικές οπτικές όπως, ο προσφερόμενος τύπος του ερωτήματος, ο τύπος περιεχομένου (π.χ. πηγαίος κώδικας / μοντέλα επιχειρησιακών διαδικασιών), ο τύπος ευρετηρίασης και οι αλγόριθμοι που υιοθετούνται για την

αντιστοίχιση και την κατάταξη των επιχειρησιακών διαδικασιών. Ωστόσο, το πρόβλημα με τις παραδοσιακές μηχανές αναζήτησης είναι ότι στις περισσότερες περιπτώσεις βασίζονται στην αναζήτηση με λέξεις-κλειδιά και με ομοιότητα κειμένου [4]. Παρόλο που οι παραδοσιακές μηχανές αναζήτησης είναι χρήσιμες σε περιπτώσεις όπου ένας χρήστης ψάχνει για ένα μοντέλο που περιέχει μια δραστηριότητα με μια συγκεκριμένη λέξη-κλειδί στην ονομασία της δεν είναι σαφές σε ποιο βαθμό αυτές οι μηχανές είναι κατάλληλες για ερωτήματα ομοιότητας μοντέλων διαδικασιών, καθώς δε λαμβάνουν υπόψη ούτε τη δομή ούτε τη συμπεριφορά των μοντέλων επιχειρησιακών διαδικασιών.

Επομένως κρίνεται απαραίτητη για τη βέλτιστη διαχείριση των μοντέλων η δυνατότητα ανεύρεσης μοντέλων διαδικασιών με παρόμοια δομή. Αυτό επιτυγχάνεται μέσω διαφορετικών μηχανισμών αντιστοίχισης όπως αυτοί αναλύονται στην επόμενη ενότητα 4.

2.2 Τομείς εφαρμογής-Αναμενόμενα Οφέλη

Ο αυξανόμενος προσανατολισμός στις διαδικασίες των σύγχρονων πληροφοριακών συστημάτων και οι αρχιτεκτονικές που προσανατολίζονται στις υπηρεσίες οδήγησε στην ύπαρξη των αποθετηρίων με εκατοντάδες ή ακόμα και χιλιάδες μοντέλων διαδικασιών που αποτελούν μια μεγάλη πηγή γνώσης για τους οργανισμούς. Η ανάκτηση τέτοιων μοντέλων διαδικασιών μέσω της αντιστοίχισης, σύμφωνα με τις ανάγκες των χρηστών από αποθετήρια επιχειρησιακών διαδικασιών έχει γίνει ένα απαραίτητο χαρακτηριστικό, δεδομένου ότι μπορεί να έχει διαφορετικές εφαρμογές όπως:

- Η επαναχρησιμοποίηση, δηλαδή η εύρεση σχετικών μοντέλων σε μια αποθήκη μοντέλων διαδικασιών για σκοπούς όπως, η πρόληψη επικαλύψεων μοντέλων καθώς και η βοήθεια κατά τη διάρκεια του σχεδιασμού της διαδικασίας.
- Προσδιορισμός κοινών μοντέλων διαδικασιών στο πλαίσιο των συγχωνεύσεων διαφορετικών οργανισμών
- Η μέτρηση της συμμόρφωσης μεταξύ μοντέλων που καθορίζουν τις επιχειρησιακές διαδικασίες και μοντέλων που προκύπτουν από την υλοποίηση των διαδικασιών

Ως παράδειγμα εφαρμογής αναφέρεται μια περίπτωση που αφορά στον Τραπεζικό χώρο και συγκεκριμένα ένα τμήμα που ελέγχει τους πιστωτικούς κινδύνους των πελατών το οποίο πρέπει να επεκτείνει τις διαδικασίες του με επιπλέον ελέγχους προκειμένου να ενισχύσει τη διαδικασία αξιολόγησης των πιστωτικών κινδύνων που συνδέονται με την έκδοση νέων πιστωτικών καρτών. Μέσω της αντιστοίχισης μπορεί να ενσωματώσει στη διαδικασία του

ένα κομμάτι διαδικασίας που σχετίζεται με επιπλέον ελέγχους που μπορεί να χρησιμοποιείται σε άλλες διαδικασίες, όπως αυτή της έγκρισης δανείου.

3. Διαφορές σε Επιχειρησιακές Διαδικασίες

Πριν την παρουσίαση των τεχνικών που χρησιμοποιούνται για την αντιστοίχιση των Επιχειρησιακών Διαδικασιών κρίνεται σκόπιμο να αναφερθούν ορισμένες από τις διαφορές που μπορούν να εμφανιστούν μεταξύ διαφορετικών μοντέλων επιχειρησιακών διαδικασιών [3]. Όσον αφορά στις δραστηριότητες που περιλαμβάνονται στις διαδικασίες εντοπίζονται οι παρακάτω διαφορές:

1. Δραστηριότητες να έχουν παραλειφθεί, δηλαδή μία δραστηριότητα που επιτελεί ένα συγκεκριμένο έργο στην μία διαδικασία, να μην υπάρχει στην άλλη διαδικασία και επίσης στην δεύτερη διαδικασία να μην υπάρχει και καμία δραστηριότητα που να επιτελεί παρόμοιο έργο.
2. Δραστηριότητες να επιτελούν το ίδιο έργο, αλλά να το επιτυγχάνουν με διαφορετικό τρόπο.
3. Δραστηριότητες σε διαφορετικό επίπεδο ανάλυσης, δηλαδή μία δραστηριότητα μιας διαδικασίας να αποτυπώνεται στην άλλη διαδικασία μέσω ενός συνόλου δραστηριοτήτων.
4. Ισοδύναμες Συλλογές Δραστηριοτήτων, δηλαδή και στις δύο διαδικασίες να υπάρχουν σύνολα από δραστηριότητες που επιτελούν το ίδιο έργο, παρότι όλες οι δραστηριότητες που αποτελούν το σύνολο στην μία διαδικασία να είναι ανόμοιες με όλες τις δραστηριότητες του συνόλου της άλλης διαδικασίας.
5. Εν μέρει ισοδύναμες Συλλογές Δραστηριοτήτων, δηλαδή στην μία διαδικασία να υπάρχει ένα σύνολο από δραστηριότητες που επιτελεί ένα έργο, το οποίο να επιτελείται μερικώς από κάποιο σύνολο δραστηριοτήτων της άλλης διαδικασίας.

Αντίστοιχα, ορισμένες από τις διαφορές που μπορούν να εμφανιστούν μεταξύ δύο μοντέλων επιχειρησιακών διαδικασιών όσον αφορά τους εξουσιοδοτημένους ρόλους που περιλαμβάνονται σε αυτά είναι:

6. Διαφορετικοί ρόλοι, δηλαδή μία κοινή δραστηριότητα και των δύο διαδικασιών να έχει ανατεθεί σε διαφορετικό επιχειρησιακό ρόλο στην μία διαδικασία σε σχέση με την άλλη διαδικασία.

7. Μεμονωμένοι ρόλοι αντί συλλογικών ρόλων, δηλαδή μία κοινή δραστηριότητα και των δύο διαδικασιών να έχει ανατεθεί στην μία διαδικασία σε συγκεκριμένο ρόλο, ενώ στην άλλη διαδικασία να έχει ανατεθεί σε ένα σύνολο από διαφορετικούς ρόλους.
8. Διαφορετικοί συλλογικοί ρόλοι, δηλαδή μία κοινή δραστηριότητα και των δύο διαδικασιών να έχει ανατεθεί σε δύο σύνολα με τελείως διαφορετικούς ρόλους.

Τέλος, ορισμένες από τις διαφορές που μπορούν να εμφανιστούν μεταξύ δύο μοντέλων επιχειρησιακών διαδικασιών όσον αφορά τις ροές των δραστηριοτήτων που περιλαμβάνονται σε αυτά είναι:

9. Διαφορετικές εξαρτήσεις, δηλαδή δύο ισοδύναμες δραστηριότητες των δύο διαδικασιών να επηρεάζονται ή να εξαρτώνται από διαφορετικές δραστηριότητες στην μία διαδικασία σε σχέση με την άλλη διαδικασία.
10. Επιπλέον εξαρτήσεις, δηλαδή η ίδια δραστηριότητα να περιέχει εξαρτήσεις στη μια διαδικασία που είναι υποσύνολο των εξαρτήσεων της άλλης διαδικασίας.
11. Δραστηριότητες που είναι ισοδύναμες να εμφανίζονται σε διαφορετικά σημεία των δύο διαδικασιών.
12. Δραστηριότητες να επαναλαμβάνονται, δηλαδή να υπάρχουν ισοδύναμες δραστηριότητες που στην μία διαδικασία να επαναλαμβάνονται, ενώ στην άλλη διαδικασία να επιτελούνται μία φορά.
13. Δραστηριότητες να έχουν διαφορετικές συνθήκες εκτέλεσης, δηλαδή να υπάρχουν ισοδύναμες δραστηριότητες που και στις δύο διαδικασίες να αλληλεπιδρούν ή να εξαρτώνται από παρόμοιες δραστηριότητες, αλλά να αλλάζουν οι συνθήκες με τις οποίες αποφασίζεται το αν θα εκτελεστούν.

4. Αντιστοίχιση Επιχειρησιακών Διαδικασιών - Τεχνικές Αντιστοίχισης

Στο χώρο της αντιστοίχισης των επιχειρησιακών διαδικασιών χρησιμοποιούνται τεχνικές αντιστοίχισης γράφων, οι οποίες βασίζουν την αντιστοίχιση στην επιθυμητή συμπεριφορά και περιεχόμενο της επιχειρησιακής διαδικασίας. Αυτό συμβαίνει διότι οι γράφοι είναι η πιο φυσική αναπαράσταση μιας επιχειρησιακής διαδικασίας. Μια επιχειρησιακή διαδικασία μπορεί να απεικονιστεί ως ένας κατευθυνόμενος γράφος, όπου κάθε κόμβος

αντιπροσωπεύει μια δραστηριότητα ή ένα συμβάν ή ένα σύνδεσμο και κάθε ακμή την κατεύθυνση του ελέγχου μεταξύ των κόμβων.

Επομένως η εύρεση ομοιοτήτων μεταξύ διαφορετικών μοντέλων επιχειρησιακών διαδικασιών μπορεί να πραγματοποιηθεί μέσω της χρήσης τεχνικών που χρησιμοποιούνται για την αντιστοίχιση γράφων.

Η Αντιστοίχιση Γράφων είναι μια ευρύτερη ερευνητική περιοχή και έχει κρατήσει την προσοχή των ερευνητών για αρκετές δεκαετίες, αν λάβουμε υπόψη ότι οι πρώτες δημοσιεύσεις σχετικές με αλγορίθμους για την αντιστοίχιση γράφων εκδόθηκαν στην δεκαετία του '70.

Σημειώνεται ότι στόχος του συγκεκριμένου μαθήματος δεν είναι να εμβαθύνει στην περιοχή της αντιστοίχισης γράφων αλλά να παρουσιάσει κάποιες έννοιες οι οποίες χρησιμοποιούνται από διαδεδομένες τεχνικές που εφαρμόζονται για την αντιστοίχιση επιχειρησιακών διαδικασιών.

4.1 Τεχνικές Αντιστοίχισης Επιχειρησιακών Διαδικασιών - Μετρικές Ομοιότητας

Για το προσδιορισμό της ομοιότητας διαφορετικών μοντέλων επιχειρησιακών διαδικασιών χρησιμοποιούνται διαφορετικές μετρικές οι οποίες αποσκοπούν στη βελτίωση της ποιότητας του αποτελέσματος και επικεντρώνουν σε διαφορετικά στοιχεία της διαδικασίας (π.χ. γεγονότα, δεδομένα εισόδου/εξόδου, ρόλους, κ.λπ.).

Οι βασικές μετρικές που χρησιμοποιούνται από την επιστημονική κοινότητα [1] για την μέτρηση της ομοιότητας των στοιχείων από τα οποία αποτελούνται τα μοντέλα που αναπαριστούν επιχειρησιακές διαδικασίες είναι οι εξής:

- Συντακτική Ομοιότητα (Syntactic similarity), η οποία βασίζεται στη σύγκριση των ονομασιών των στοιχείων που εμφανίζονται στα μοντέλα των διαδικασιών (ετικέτες δραστηριοτήτων, ετικέτες γεγονότων, κλπ.) κάνοντας συντακτική ανάλυση.
- Σημασιολογική Ομοιότητα (Semantic similarity), στην οποία λαμβάνεται υπόψη η σημασία των λέξεων που περιέχονται στις ονομασίες που έχουν τα στοιχεία των μοντέλων.
- Δομική Ομοιότητα (Structural similarity), στην οποία λαμβάνεται υπόψη η θέση των στοιχείων στα μοντέλα.

- Ομοιότητα Συμπεριφοράς (Behavioral similarity), στην οποία λαμβάνεται υπόψη η ο τρόπος εκτέλεσης των μοντέλων των διαδικασιών. Η Ομοιότητα Συμπεριφοράς προϋποθέτει παρόμοια εκτέλεση των διαφορετικών μοντέλων διαδικασιών.

Σε μαθηματικό επίπεδο, όλες αυτές οι μετρικές παίρνουν τιμές μεταξύ μηδέν (0) και ένα (1) και συμβολίζουν το βαθμό ομοιότητας των στοιχείων δύο μοντέλων, όπου η τιμή μηδέν (0) δείχνει ότι δεν υπάρχει καμία ομοιότητα μεταξύ των στοιχείων και της δομής τους, ενώ η τιμή ένα (1) δείχνει ότι τα στοιχεία και η δομή τους είναι όμοια. Οι περισσότερες λύσεις σύγκρισης γραφημάτων που αναπαριστούν επιχειρησιακές διαδικασίες συνδυάζουν περισσότερες από μία μετρικές για να παράγουν ένα σταθμισμένο βαθμό ομοιότητας μεταξύ των στοιχείων των γραφημάτων, ώστε να καταλήξουν σε όσο το δυνατόν ακριβέστερα αποτελέσματα.

Επομένως, στις περιπτώσεις όπου επιχειρήσεις έχουν δημιουργήσει αποθετήρια που περιέχουν μοντέλα των επιχειρηματικών διαδικασιών τους σε μορφή γραφημάτων κάποιας μορφής αναπαράστασης (Το πρότυπο BPMN χρησιμοποιείται συνήθως σε τέτοιες περιπτώσεις), η εύρεση του μοντέλου που έχει τις μεγαλύτερες ομοιότητες με κάποια συγκεκριμένη επιχειρηματική διαδικασία πραγματοποιείται μέσω της αντιστοίχισης του μοντέλου του ερωτήματος με τα μοντέλα του αποθετηρίου. Προς το σκοπό αυτό χρησιμοποιούνται τεχνικές σύγκρισης γραφημάτων με χαρακτηριστικά όπως αυτά που αναφέρονται παραπάνω.

4.2 Ομοιότητα Ονομασιών

4.2.1 Συντακτική Ομοιότητα

Προκειμένου να προσδιοριστεί η συντακτική ομοιότητα λαμβάνεται υπόψη η σύνταξη των ονομασιών των στοιχείων των μοντέλων και οι τεχνικές βρίσκουν τη βέλτιστη λύση, συγκρίνοντας αυτές τις ονομασίες ενώ λαμβάνουν επίσης υπόψη το συνολικό μέγεθος των μοντέλων. Οι λύσεις σε αυτήν την κατηγορία βασίζονται σε τεχνικές όπως οι: string-edit distance, n-gram, morphological analysis (stemming) και stop-word elimination [5].

Ενδεικτικά περιγράφεται η επικρατέστερη τεχνική “Levenshtein distance” η οποία υπολογίζει τον μικρότερο αριθμό των ενεργειών επεξεργασίας που είναι απαραίτητες για να τροποποιήσει μία συμβολοσειρά προκειμένου να παραχθεί μια άλλη συμβολοσειρά. Ο πιο συνηθισμένος τρόπος υπολογισμού είναι μέσω της δυναμικής προσέγγισης στο προγραμματισμό. Αναλυτικότερα, ένας πίνακας δημιουργείται μετρώντας στο (m,n)-κελί την απόσταση Levenshtein μεταξύ του m χαρακτήρα της μίας λέξης με το n χαρακτήρα της άλλης

λέξης. Ο πίνακας μπορεί να συμπληρωθεί από την πάνω και αριστερή γωνία προς τα κάτω και δεξιά γωνία. Κάθε μετάβαση οριζόντια ή κάθετα στην πρώτη γραμμή ή στήλη, αντιστοιχεί σε μία εισαγωγή ή διαγραφή, αντίστοιχα. Το κόστος ορίζεται συνήθως ως 1 για κάθε μία από τις ενέργειες (εισαγωγή ή διαγραφή).

Η τιμή για τα υπόλοιπα κελιά υπολογίζεται βάσει του ελάχιστου πάντα κόστους σε τοπικό επίπεδο. Εάν οι χαρακτήρες δεν ταιριάζουν τότε το επιπλέον κόστος είναι 1 ενώ εάν οι χαρακτήρες ταιριάζουν δεν υπάρχει επιπλέον κόστος. Με αυτό τον τρόπο ο αριθμός στην κάτω δεξιά γωνία είναι η απόσταση μεταξύ των δύο Levenshtein λέξεων. Στην Εικόνα 1 παρουσιάζεται ο πίνακας που υπολογίζει την απόσταση των ονομασιών «Meilenstein" και "Levenshtein".

		m	e	i	l	e	n	s	t	e	i	n
l	0	1	2	3	4	5	6	7	8	9	10	11
e	1	1	2	3	3	4	5	6	7	8	9	10
v	2	2	1	2	3	3	4	5	6	7	8	9
e	3	3	2	2	3	4	4	5	6	7	8	9
n	4	4	3	3	3	3	4	5	6	6	7	8
s	5	5	4	4	4	4	3	4	5	6	7	7
h	6	6	5	5	5	5	4	3	4	5	6	7
t	7	7	6	6	6	6	5	4	4	5	6	7
e	8	8	7	7	7	7	6	5	4	5	6	7
i	9	9	8	8	8	7	7	6	5	4	5	6
n	10	10	9	8	9	8	8	7	6	5	4	5
	11	11	10	9	9	9	8	8	7	6	5	4

Εικόνα 1: Παράδειγμα τεχνικής «string-edit-distance»

4.2.2 Σημασιολογική Ομοιότητα – WordNet

Προκειμένου να προσδιοριστεί η σημασιολογική ομοιότητα, η σημασία των ονομασιών των στοιχείων των μοντέλων λαμβάνεται υπόψη. Οι τεχνικές σε αυτήν την κατηγορία βασίζονται σε συνώνυμα και άλλες σημασιολογικές σχέσεις που υπάρχουν σε Εγκυκλοπαιδικούς Θησαυρούς (π.χ. WordNet).

Θα πρέπει να σημειωθεί ότι όταν οι προσφερόμενες λύσεις αποκλειστικά μετρούν τη σημασιολογική ομοιότητα των στοιχείων διαφορετικών επιχειρησιακών διαδικασιών, τότε προκύπτει πρόβλημα όταν η δομή της διαδικασίας είναι διαφορετική (π.χ. σειρά των δραστηριοτήτων, εισαγωγή συνδέσμων κ.λπ.) καθώς η σημασιολογική ομοιότητα είναι ίδια. Παρακάτω ενδεικτικά περιγράφεται ένας υπάρχων Εγκυκλοπαιδικός Θησαυρός.

WordNet

Κάθε σύστημα που έχει ως στόχο να επεξεργαστεί φυσικές γλώσσες όπως κάνουν οι άνθρωποι πρέπει να έχει πληροφορίες σχετικά με τις λέξεις και τις έννοιές τους. Αυτές οι πληροφορίες παρέχονται μέσα από λεξικά. Επομένως λεξικά που χρησιμοποιούνται από συστήματα είναι πλέον ευρέως διαθέσιμα. Το WordNet είναι μια ηλεκτρονική λεξιλογική βάση δεδομένων. Τα αγγλικά ουσιαστικά, ρήματα, επίθετα και επιρρήματα οργανώνονται σε σύνολα συνωνύμων, όπου καθένα αντιπροσωπεύει μια έννοια. Σημασιολογικές σχέσεις συνδέουν τα σύνολα συνωνύμων. Στο WordNet ορίζεται το λεξιλόγιο της γλώσσας ως ένα σύνολο W ζευγών (f, s) , όπου f είναι μία συμβολοσειρά πάνω σε ένα πεπερασμένο αλφάβητο και s είναι ένα στοιχείο από ένα δεδομένο σύνολο εννοιών. Σημειώνεται ότι κάθε συμβολοσειρά με μια έννοια, σε μια γλώσσα, ονομάζεται λέξη στη γλώσσα αυτή. Μία λέξη η οποία έχει περισσότερες από μία έννοιες λέγεται πολύσημο, δύο λέξεις που μοιράζονται τουλάχιστον μία κοινή έννοια λέγονται συνώνυμα.

Η χρήση μιας λέξης είναι το σύνολο C , από γλωσσολογικό περιεχόμενο, μέσα στο οποίο η λέξη μπορεί να χρησιμοποιηθεί. Η σύνταξη της γλώσσας χωρίζει το C σε συντακτικές κατηγορίες. Λέξεις που περιέχονται στο υποσύνολο N είναι ουσιαστικά, οι λέξεις που εμφανίζονται στην υποσύνολο V είναι ρήματα κ.λπ.. Μέσα σε κάθε κατηγορία συντακτικού περιεχομένου υπάρχουν και κατηγορίες σημασιολογικού περιεχομένου- το σύνολο των περιεχομένων στο οποίο μια συγκεκριμένη f μπορεί να χρησιμοποιηθεί για να εκφράσει ένα συγκεκριμένο s .

Οι σημασιολογικές σχέσεις στο WordNet επελέγησαν επειδή εφαρμόζονται ευρέως σε όλη την αγγλική και επειδή είναι οικείες - ο χρήστης δεν χρειάζεται να έχει ειδική εκπαίδευση στη γλωσσολογία για να τους καταλάβει. Το WordNet περιλαμβάνει τις ακόλουθες σημασιολογικές σχέσεις (Εικόνα 2):

- Συνωνυμία: είναι βασική σχέση στο WordNet, γιατί το WordNet χρησιμοποιεί σύνολα συνωνύμων (synsets) προκειμένου να παρουσιάσει τις έννοιες των λέξεων. Συνωνυμία είναι μια συμμετρική σχέση μεταξύ μορφών λέξεων.
- Αντώνυμα: είναι επίσης μια συμμετρική σημασιολογική σχέση μεταξύ των μορφών λέξεων, ιδιαίτερα σημαντική στην οργάνωση των νοημάτων των επιθέτων και των επιρρημάτων.
- Υπωνυμία και το αντίστροφο Υπερνημία: είναι μεταβατικές σχέσεις μεταξύ συνωνύμων. Επειδή συνήθως υπάρχει μόνο ένα υπερώνυμο, αυτή η

σημασιολογική σχέση οργανώνει τις έννοιες των ουσιαστικών σε μια ιεραρχική δομή.

- Μερωνυμία και η αντίστροφή της, «holonymy», είναι πολύπλοκες σημασιολογικές σχέσεις.
- «Troponymy» είναι για τα ρήματα όπως είναι η Υπωνυμία για τα ουσιαστικά, αν και οι προκύπτουσες ιεραρχίες είναι πιο ρηχές.
- Συνεπαγωγή: είναι σχέσεις μεταξύ ρημάτων.

Κάθε μία από αυτές τις σημασιολογικές σχέσεις εκπροσωπείται από δείκτες μεταξύ των μορφών των λέξεων ή μεταξύ συνωνύμων.

Semantic Relation	Syntactic Category	Examples
Synonymy (similar)	N, V, Aj, Av	pipe, tube rise, ascend sad, unhappy rapidly, speedily
Antonymy (opposite)	Aj, Av, (N, V)	wet, dry powerful, powerless friendly, unfriendly rapidly, slowly
Hyponymy (subordinate)	N	sugar maple, maple maple, tree tree, plant
Meronymy (part)	N	brim, hat gin, martini ship, fleet
Troponomy (manner)	V	march, walk whisper, speak
Entailment	V	drive, ride divorce, marry
<i>Note: N = Nouns Aj = Adjectives V = Verbs Av = Adverbs</i>		

Εικόνα 2: Σημασιολογικές Σχέσεις στο WordNet

4.3 Δομική Ομοιότητα

Υπενθυμίζεται ότι η αντιστοίχιση Γράφων είναι η διαδικασία που, δεδομένου ενός γραφήματος G που αποτελεί το ερώτημα και μια σειρά από γραφήματα T τα οποία υπάρχουν σε κάποια αποθήκη γραφημάτων, συγκρίνει το γράφημα G με κάθε γράφημα T ελέγχοντας το κατά πόσο είναι παρόμοια ή όχι.

Σημειώνεται ότι στόχος του συγκεκριμένου μαθήματος δεν είναι να εμβαθύνει στην περιοχή της αντιστοίχισης γράφων αλλά να παρουσιάσει συνοπτικά την περιοχή. Για το λόγο αυτό παρουσιάζονται στη συνέχεια μόνο η τεχνική Graph-edit-Distance για τη δομική αντιστοίχιση γράφων και ο αλγόριθμος A^* .

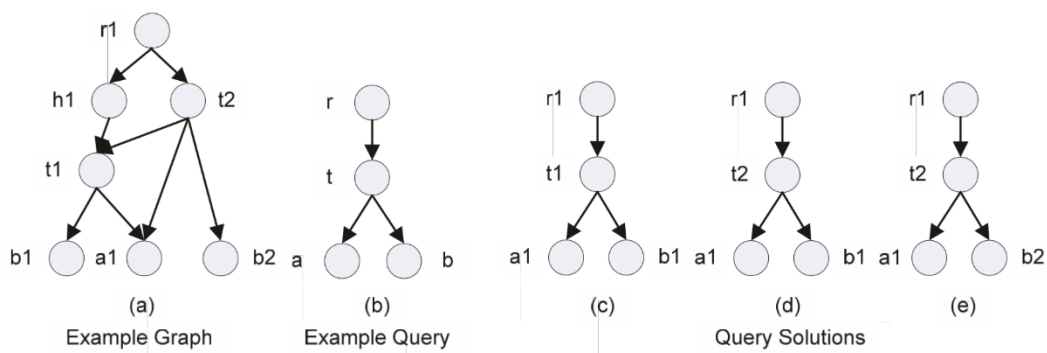
Γενικά οι αλγόριθμοι που χρησιμοποιούνται για την αντιστοίχιση των γράφων μπορούν να χωριστούν σε δύο ευρύτερες κατηγορίες με βάση τα αποτελέσματά τους: αλγόριθμοι που πετυχαίνουν ακριβή αντιστοίχιση γράφων και αλγόριθμοι που βασίζονται στην κατά προσέγγιση αντιστοίχιση γράφων.

Στις περισσότερες επιχειρηματικές εφαρμογές, όπου οι αναζητήσεις σε αποθήκες επιχειρησιακών διαδικασιών δεδομένου ενός ερωτήματος γίνονται επιθυμώντας την ακριβή αντιστοίχιση (πχ. Ισομορφισμός γράφων) είναι πολύ σπάνια εξαιτίας της πολυπλοκότητας των διαδικασιών. Επομένως οι χρήστες είναι πολύ ικανοποιημένοι με τα αποτελέσματα που φαίνονται παρόμοια με τις απαιτήσεις τους.

Ακριβής αντιστοίχιση γράφων (Exact Graph Matching)

Ο στόχος των αλγορίθμων στην τρέχουσα κατηγορία είναι να εξετάσουν κατά πόσον δύο γράφοι G και G' , είτε ως σύνολο είτε ένα μέρος τους, είναι ταυτόσημοι. Με άλλα λόγια δύο παρακείμενοι κόμβοι του G συσχετίζονται με δύο παρακείμενους κόμβους του γραφήματος G' .

Ο πιο διαδεδομένος τύπος ακριβής αντιστοίχισης γράφων είναι ο Ισομορφισμός Γράφων (Εικόνα 3) ο οποίος στοχεύει στην συσχέτιση μεταξύ των κόμβων του G και G' , διατηρώντας τη δομή των ακμών.



Εικόνα 3: Γραφήματα (α), (β) και αποτελέσματα Ισομορφισμού Γράφων (c), (d), (e)

Η πιο δημοφιλής τεχνική για Ισομορφισμό γράφων βασίζεται στην οπισθοδρόμηση σε ένα δέντρο αναζήτησης όπου μια μερική αντιστοίχιση κόμβων επαναληπτικά επεκτείνεται έως

όπου οι περιορισμοί σε σχέση με τη δομή των άκρων παραβιάζονται ή οι ονομασίες των κόμβων / ακμών είναι ασυνεπής, όπου ενεργοποιείται η διαδικασία οπισθοδρόμησης. Ο Ullman [9] παρουσιάζει μια αντιπροσωπευτική εφαρμογή αυτής της τεχνικής.

Προσεγγιστική αντιστοίχιση γράφων (Inexact Graph Matching)

Ο στόχος αυτών των αλγορίθμων είναι να εξετάσουν μέσω της σύγκρισης το κατά πόσον τα δύο γραφήματα που αποτελούν το γράφο του ερωτήματος και το υπό εξέταση γράφο, είναι πανομοιότυπα ή παρόμοια. Οι τεχνικές που χρησιμοποιούνται συνήθως είναι είτε η μερική αντιστοίχιση των κόμβων του ερωτήματος σε κόμβους της υπό εξέταση διαδικασίας ή μέσω της μετατροπής των κόμβων της διαδικασίας με στόχο τη δημιουργία της απόλυτης ταύτισης (Graph Edit Distance). Έτσι, οι γράφοι που είναι κοντά στο υπό εξέταση ερώτημα δεν θα πρέπει να αποκλείονται από το αποτέλεσμα της διαδικασίας της αντιστοίχισης.

Επομένως η πρόκληση και ο στόχος είναι να εντοπιστούν μέσω των προσεγγιστικών αλγορίθμων αντιστοίχισης πόσο οι επιχειρησιακές διαδικασίες διαφέρουν ή μοιράζονται.

Οι αλγόριθμοι που βασίζονται στην προσεγγιστική αντιστοίχιση γράφων μπορούν να διαχωριστούν περαιτέρω βάσει της ποιότητας του αποτελέσματος. Έτσι διακρίνουμε τους αλγόριθμους που βρίσκουν τις βέλτιστες ή ήμι-βέλτιστες λύσεις.

Αλγόριθμοι για βέλτιστες λύσεις

Πιο συγκεκριμένα οι αλγόριθμοι της πρώτης κατηγορίας βρίσκουν πάντα μια λύση με το ελάχιστο κόστος αντιστοιχίας, σε όλο το πεδίο της αναζήτησης, πράγμα που σημαίνει ότι εγγυώνται την ανεύρεση λύσης που ταιριάζει απόλυτα με το γράφο του ερωτήματος, εάν υπάρχει. Ωστόσο το πρόβλημα με αυτήν την κατηγορία αλγορίθμων είναι ότι ο χρόνος απόκρισης τους σε σχέση με τους αλγόριθμους που βασίζονται στην ακριβή αντιστοιχία δεν είναι βελτιωμένος.

Αλγόριθμοι για ημι-βέλτιστες λύσεις

Αντίθετα οι αλγόριθμοι που βρίσκουν ημι-βέλτιστες λύσεις, βρίσκουν πάντα μια λύση με το ελάχιστο κόστος αντιστοιχίας, σε τοπικό επίπεδο που σημαίνει ότι παρόλο που μπορεί να υπάρχει κάποιος γράφος που ταιριάζει απόλυτα με το γράφο του ερωτήματος δεν είναι εγγυημένο ότι θα βρεθεί. Πάραυτα το βασικό πλεονέκτημα αυτής της κατηγορίας αλγορίθμων είναι ότι ο χρόνος απόκρισης είναι μειωμένος σε πολυωνυμικό χρόνο.

Graph Edit Distance και Αλγόριθμος A*

Όπως αναφέρθηκε και παραπάνω ένας διαδομένος τρόπος που χρησιμοποιείται από αλγορίθμους που εστιάζονται στον κατά προσέγγιση αντιστοίχιση γράφων είναι ο λεγόμενος 'Graph-Edit Distance' [6].

Ακολουθούν κάποιοι ορισμοί που βοηθούν στην κατανόηση των εννοιών που χρησιμοποιούνται:

Η συγκεκριμένη τεχνική βασίζεται στον υπολογισμό του ελάχιστου ποσού των στρεβλώσεων που πρέπει να γίνουν σε κάποιο γράφο προκειμένου να μετατραπεί στο γράφο του ερωτήματος. Στο πλαίσιο της επεξεργασίας, με αυτή την τεχνική, οι βασικότερες αλλαγές που μπορούν να εφαρμοστούν για την μετατροπή ενός γραφήματος είναι η αντικατάσταση κόμβου ή ακμής (που σημαίνει ουσιαστικά την αντικατάσταση της ετικέτας τους), η διαγραφή κόμβου ή ακμής, και η εισαγωγή κόμβου ή ακμής. Η επιλογή μιας συγκεκριμένης εργασίας που πρόκειται να χρησιμοποιηθεί κατά την επεξεργασία της συγκεκριμένης τεχνικής εξαρτάται από το πόσο παρόμοιοι είναι οι υπό εξέταση κόμβοι. Αν για παράδειγμα ο κόμβος του ερωτήματος είναι παρόμοιος με κάποιον κόμβο του υποψήφιου γράφου τότε επιλέγεται η εργασία της αντικατάστασης διαφορετικά επιλέγεται η εργασία της εισαγωγής. Επομένως κάθε εργασία κατά τη διάρκεια της επεξεργασίας συνδέεται με κόστος. Στις περισσότερες περιπτώσεις, το κόστος για εργασίες όπως εισαγωγή και διαγραφή είναι σταθερή, ενώ το κόστος της αντικατάστασης εξαρτάται από το βαθμό ομοιότητας των κόμβων (πχ. βάσει συντακτικής ή/και σημασιολογικής ομοιότητας).

Έτσι, έχοντας ένα ζεύγος γραφημάτων: Γράφημα Ερωτήματος G και το υποψήφιο Γράφημα G' , υπάρχει μια σειρά από εργασίες επεξεργασίας $\{e_1, e_2, e_3 \dots\}$ που απαιτούνται προκειμένου να μετατραπεί το ένα γράφημα στο άλλο. Η ακολουθία των εργασιών επεξεργασίας ονομάζεται "μονοπάτι επεξεργασίας" από το G' στο G .

Σε γενικές γραμμές, πολλά μονοπάτια επεξεργασίας μπορεί να υπάρχουν μεταξύ δύο δεδομένων γραφημάτων. Για να αξιολογηθεί ποσοτικά ποιο μονοπάτι επεξεργασίας είναι το καλύτερο, οι λειτουργίες κόστους επεξεργασίας εισάγονται και συνδέονται με κάθε εργασία επεξεργασίας. Το κόστος επεξεργασίας ως τιμή μετράει πόσο ισχυρή είναι κάθε τροποποίηση από κάθε εργασία επεξεργασίας στο γράφημα. Ως εκ τούτου, όταν οι εργασίες επεξεργασίας εφαρμόζονται σε δύο παρόμοια γραφήματα, πρέπει να υπάρχει ένα μονοπάτι επεξεργασίας με εργασίες επεξεργασίας χαμηλού κόστους, ενώ για ανόμοια γραφήματα όλα τα μονοπάτια επεξεργασίας θα πρέπει να συνδέονται με εργασίες επεξεργασίας υψηλού κόστους. Καθώς η συνάρτηση κόστους χρησιμοποιείται για να ευνοήσει μικρές στρεβλώσεις σε σχέση με

μεγάλες τροποποιήσεις του γραφήματος, εξαρτάται σε μεγάλο βαθμό από τις υποκείμενες ονομασίες των κόμβων και ακμών.

Η εν λόγω τεχνική του graph-edit distance παρέχει μεγάλη ευελιξία και επιτρέπει την ενσωμάτωση ειδικής γνώσης που υπάρχει σε διαφορετικούς τύπους γράφων (όπως η ειδική γνώση των δομικών στοιχείων της γλώσσας ορισμού των επιχειρησιακών διαδικασιών), σε κάθε πεδίο, σχετικά με την ομοιότητα των αντικειμένων, του δια μέσου του ορισμού ειδικών λειτουργιών κόστους.

Αξίζει να σημειωθεί ότι οι προτάσεις αλγορίθμων που βασίζονται σε αυτή τη τεχνική αντιστοίχισης γράφων είναι σχεδιασμένοι να βρίσκουν τη βέλτιστη λύση, ενώ η υπολογιστική πολυπλοκότητα τους είναι εκθετική όσον αφορά στον αριθμό των κόμβων των συμμετεχόντων γράφων. Αυτό σημαίνει ότι ο χρόνος απόκρισης και η πολυπλοκότητα του αλγορίθμου μπορεί να είναι τεράστια, ακόμη και για μικρού μεγέθους γράφων.

Η τεχνική του graph-edit distance στηρίζεται στη χρησιμοποίηση αλγορίθμων αναζήτησης δενδρικής μορφής οι οποίοι εξερευνούν το πεδίο όλων των πιθανών αντιστοιχίσεων των κόμβων/ακμών μεταξύ του γράφου του ερωτήματος και των υποψήφιων γράφων. Αυτοί οι αλγόριθμοι συνήθως βασίζονται στον αλγόριθμο A*[1] όπου η βασική ιδέα είναι ότι το πεδίο αναζήτησης παρουσιάζεται ως ένα ταξινομημένο δέντρο του οποίου σημείο εκκίνησης είναι ο κόμβος ρίζας, οι ενδιάμεσοι κόμβοι αναπαριστούν μερικές λύσεις ενώ οι κόμβοι-φύλλα αναπαριστούν ολοκληρωμένες λύσεις, αλλά όχι απαραίτητα και τις βέλτιστες. Κατά την εκτέλεση του αλγόριθμου οι ενδιάμεσοι κόμβοι του δέντρου αναζήτησης κατασκευάζονται δυναμικά και επαναληπτικά με την επέκταση των κόμβων, σε κάθε επανάληψη του αλγόριθμου, βάσει ευρετικών (heuristic) λειτουργιών κόστους.

Πιο συγκεκριμένα ο αλγόριθμος αυτός χρησιμοποιεί δύο συναρτήσεις μέτρησης του κόστους ενός συνόλου αλλαγών p , την συνάρτηση $g(p)$ που μετρά το κόστος μετάβασης από τον αρχικό κόμβο του γραφήματος έως τον τρέχον κόμβο, και την συνάρτηση $h(p)$ που μετρά το κόστος μετάβασης από τον τρέχον κόμβο έως τον τελευταίο κόμβο του γραφήματος που βρίσκεται στο επιλεγμένο μονοπάτι.

Ο αλγόριθμος ταξινομεί τους κόμβους του πρώτου γραφήματος του αποθετηρίου, με βάση την τιμή της συνάρτησης $g(p)$ για κάθε κόμβο και εισάγει σε ένα ανοικτό σετ από σύνολα αλλαγών, τη διαγραφή του πρώτου στην ταξινόμηση κόμβου, καθώς και όλες τις αντικαταστάσεις του από καθένα ξεχωριστά από τους κόμβους του δεύτερου γραφήματος του ερωτήματος.

Στη συνέχεια, και επαναληπτικά, εντοπίζει το επόμενο πιο υποσχόμενο σύνολο αλλαγών, που είναι αυτό που ελαχιστοποιεί το κόστος της A* αναζήτησης, δηλαδή το άθροισμα των συναρτήσεων κόστους $g(p) + h(p)$ για όλες τις αλλαγές που περιέχονται στο ανοικτό σετ από σύνολα αλλαγών. Αν το σύνολο αλλαγών με το μικρότερο κόστος είναι αρκετό ώστε να μετατρέψει το πρώτο γράφημα στο δεύτερο, τότε ο αλγόριθμος έχει εντοπίσει ουσιαστικά την μετρική «Graph edit distance». Αν όχι, τότε η διαδικασία συνεχίζεται εισάγοντας στο σετ από ανοικτά σύνολα αλλαγών, τις ενώσεις του συνόλου αλλαγών με το μικρότερο κόστος με την διαγραφή αλλά και με όλες τις αντικαταστάσεις του επόμενου στην ταξινόμηση κόμβου από το πρώτο γράφημα. Ο αλγόριθμος αυτός, επομένως, είναι σίγουρο ότι θα εντοπίσει το σύνολο αλλαγών που μετατρέπουν το πρώτο γράφημα στο δεύτερο, με το μικρότερο δυνατό κόστος, δηλαδή την τιμή της μετρικής «Graph edit distance».

Στην εικόνα που ακολουθεί, και προέρχεται από το [8], φαίνονται τα βήματα του αλγορίθμου A* σε ψευδογλώσσα.

Algorithm 4. Computation of graph edit distance by A* algorithm

Input: Non-empty graphs $g_1 = (V_1, E_1, \mu_1, \nu_1)$ and $g_2 = (V_2, E_2, \mu_2, \nu_2)$, where $V_1 = \{u_1, \dots, u_{|V_1|}\}$ and $V_2 = \{v_1, \dots, v_{|V_2|}\}$

Output: A minimum-cost edit path from g_1 to g_2
e.g. $p_{min} = \{u_1 \rightarrow v_3, u_2 \rightarrow \varepsilon, \dots, \varepsilon \rightarrow v_6\}$

- 1: Initialize OPEN to the empty set
- 2: For each vertex $w \in V_2$, insert the substitution $\{u_1 \rightarrow w\}$ into OPEN
- 3: Insert the deletion $\{u_1 \rightarrow \varepsilon\}$ into OPEN
- 4: **loop**
- 5: Remove $p_{min} = \arg \min_{p \in \text{OPEN}} \{g(p) + h(p)\}$ from OPEN
- 6: **if** p_{min} is a complete edit path **then**
- 7: Return p_{min} as the solution
- 8: **else**
- 9: Let $p_{min} = \{u_1 \rightarrow v_{i_1}, \dots, u_k \rightarrow v_{i_k}\}$
- 10: **if** $k < |V_1|$ **then**
- 11: For each $w \in V_2 \setminus \{v_{i_1}, \dots, v_{i_k}\}$, insert $p_{min} \cup \{u_{k+1} \rightarrow w\}$ into OPEN
- 12: Insert $p_{min} \cup \{u_{k+1} \rightarrow \varepsilon\}$ into OPEN
- 13: **else**
- 14: Insert $p_{min} \cup \bigcup_{w \in V_2 \setminus \{v_{i_1}, \dots, v_{i_k}\}} \{\varepsilon \rightarrow w\}$ into OPEN
- 15: **end if**
- 16: **end if**
- 17: **end loop**

Εικόνα 4: Αλγόριθμος A*

Θα πρέπει να σημειωθεί ότι στον Αλγόριθμο A*, σε αντίθεση με τους αλγόριθμους που επιτυγχάνουν την ακριβή αντιστοίχιση διαφορετικών γραφών, όλοι οι κόμβοι του ενός γράφου μπορούν να συσχετιστούν με οποιοδήποτε κόμβο του δεύτερου γράφου με αποτέλεσμα να αυξάνεται η πολυπλοκότητα του αλγορίθμου.

4.4 Ομοιότητα Προφίλ Συμπεριφοράς (Behavioural Profile Similarity)

Στην ενότητα αυτή παρουσιάζεται αναλυτικά ένας τρόπος μέτρησης της ομοιότητας συμπεριφοράς ο οποίος βασίζεται στη δημιουργία «προφίλ συμπεριφοράς». Επίσης

επικεντρώνεται στη σειρά εκτέλεσης των δραστηριοτήτων και παραμελεί άλλες λεπτομέρειες της συμπεριφοράς της επιχειρησιακής διαδικασίας.

Η εκτίμηση της ομοιότητας βάσει του προφίλ συμπεριφοράς διαφορετικών διαδικασιών απαιτεί αντιστοιχισμένες δραστηριότητες. Δηλαδή σε δύο μοντέλα διαδικασιών, οι αντιστοιχίες μεταξύ των δραστηριοτήτων πρέπει να είναι καθορισμένες. Αυτές οι αντιστοιχίες χρησιμοποιούνται για την ποσοτικοποίηση της επικάλυψης της συμπεριφοράς και στα δύο μοντέλα.

Ορισμός (Προφίλ Συμπεριφοράς)

P είναι ένα μοντέλο διαδικασίας

Ένα ζεύγος δραστηριοτήτων (x, y)

$\forall (A, P) \exists X$ έχει τις εξής σχέσεις

- Αυστηρής ακολουθίας $\rightsquigarrow p$, iff $x > y$ and $y \not> x$.
- Αποκλειστικότητα $+p$, iff $x \not> y$ and $y \not> x$.
- Διεμπλοκή ακολουθίας $||P$, iff $x > y$ and $y > x$.

$\mathcal{B}P = \{ \rightsquigarrow p, +p, ||p \}$ είναι το προφίλ συμπεριφοράς του μοντέλου P.

Για κάθε ζεύγος (x, y) σε αυστηρή ακολουθία, η αντίστροφη σχέση αυστηρής ακολουθίας περιλαμβάνει το αντίστροφο ζεύγος (y,x).

Ένα μοντέλο διαδικασίας P μοιάζει με ένα άλλο μοντέλο Q σε ορισμένες πτυχές της συμπεριφοράς, αν επικαλύπτονται στο προφίλ της συμπεριφοράς τους $\mathcal{B}P$ και $\mathcal{B}Q$, αντίστοιχα. Μεγαλύτερη επικάλυψη υποδεικνύει πιο παρόμοια μοντέλα να είναι.

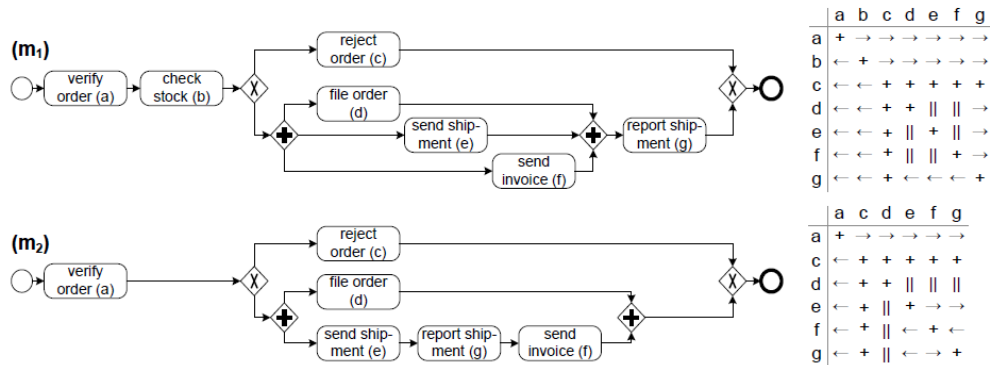
Η Ομοιότητα υπολογίζεται με τη γνωστή μέθοδο Jaccard Coefficient

$$\text{sim}(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

Πρώτον, το μέτρο αυτό μπορεί εύκολα να εφαρμοστεί σε προφίλ συμπεριφοράς, καθώς κάθε σχέση προφίλ συμπεριφοράς είναι ουσιαστικά ένα σύνολο. Δεύτερον μπορεί να μετατραπεί στο μέτρο $d(A,B) = 1 - \text{Sim}(A, B)$ για να καταστεί δυνατή αποδοτική αναζήτηση ομοιότητας [7].

Η Εικόνα 5 δείχνει δύο μοντέλα που αφορούν στη διαδικασία χειρισμού των παραγγελιών (m1, m2), κατά την οποία μια παραγγελία λαμβάνεται, ελέγχεται, και ολοκληρώνεται ή απορρίπτεται. Οι αντιστοιχίες μεταξύ των δραστηριοτήτων που απεικονίζονται με ίδιες ονομασίες στα στοιχεία των δύο μοντέλων, και εκπροσωπούνται μέσω των δεικτών a έως g.

Τα προφίλ συμπεριφοράς ($\mathcal{B} m_1, \mathcal{B} m_2$) απεικονίζεται ως πίνακας πάνω στις δραστηριότητες, π.χ., α και γ είναι σε αυστηρή ακολουθία, αφού σε όλα τα ίχνη, όπου και οι δύο δραστηριότητες εμφανίζονται, η δραστηριότητα α συμβαίνει πριν τη δραστηριότητα γ. Η σχέση αποκλειστικότητας είναι η αυστηρότερη σχέση σε ένα προφίλ συμπεριφοράς, επειδή επιβάλλει την απουσία συνύπαρξης δύο δραστηριοτήτων μέσα σε μία εκτέλεση της διαδικασίας. Η αντίστοιχη ομοιότητα ποσοτικοποιεί, πόσα κοινά ζεύγη υπάρχουν σε δύο μοντέλα διαδικασιών που καλύπτουν την ίδια σχέση αποκλειστικότητας.



Εικόνα 5: Προφίλ Συμπεριφοράς

Ορισμός Ομοιότητας Αποκλειστικότητας

P, Q είναι μοντέλα διαδικασίας και $+p, +q$ οι σχέσεις αποκλειστικότητας των αντίστοιχων προφίλ συμπεριφοράς τους $\mathcal{B}P$ και $\mathcal{B}Q$.

Ομοιότητα Αποκλειστικότητα ορίζεται ως εξής:

$$\text{sim}_+(\mathcal{B}P, \mathcal{B}Q) = \frac{|+P \cap +Q|}{|+P \cup +Q|}$$

Για παράδειγμα οι δραστηριότητες c και d και στα δύο μοντέλα διαδικασιών στην Εικόνα 5. Από το σενάριο, είναι απαγορευτικό να απορρίψει κανείς μια παραγγελία και να αποθηκεύσει την παραγγελία αυτή και να την ολοκληρώσει. Η μόνη διαφορά μεταξύ m_1 και m_2 , όσον αφορά τη σχέση αποκλειστικότητας είναι η απουσία της δραστηριότητας b στο m_2 , επειδή η δραστηριότητα αυτή είναι αποκλειστικά για τη m_2 . Αυτό προσδίδει μια υψηλή ομοιότητα

$$\text{sim}_+(\mathcal{B}m_1, \mathcal{B}m_2) = \frac{14}{15} \approx 0.933$$

Ορισμός Ομοιότητας Αυστηρής ακολουθίας

P, Q είναι μοντέλα διαδικασίας και $\rightsquigarrow p, \rightsquigarrow q$ οι σχέσεις αυστηρής ακολουθίας των αντίστοιχων προφίλ συμπεριφοράς τους $\mathcal{B}P$ και $\mathcal{B}Q$.

Ομοιότητα Αυστηρής ακολουθίας ορίζεται ως εξής:

$$\text{sim}_{\rightsquigarrow}(\mathcal{B}P, \mathcal{B}Q) = \frac{|\rightsquigarrow P \cap \rightsquigarrow Q|}{|\rightsquigarrow P \cup \rightsquigarrow Q|}$$

Λόγω $x \rightsquigarrow y \Leftrightarrow y \rightsquigarrow x$, αρκεί να ενσωματώσει μόνο την σχέση αυστηρής ακολουθίας στον υπολογισμό της ομοιότητας. Η αντίστροφη σχέση σιωπηρά καλύπτεται. Αυτό μπορεί να φανεί στον πίνακα του προφίλ συμπεριφοράς καθώς οι σχέσεις αυστηρής ακολουθίας καθρεφτίζονται κατά μήκος του διαγώνιου άξονα. Για παράδειγμα στο m2 μοντέλο το ζεύγος (f,g) είναι σε αντίστροφη αυστηρή ακολουθία και το ζεύγος (g,f) είναι σε αυστηρή ακολουθία.

Τα μοντέλα M1 και M2 του παραδείγματος δείχνουν μια σημαντική διαφορά στις σχέσεις αυστηρής ακολουθίας, επειδή οι δραστηριότητες e και f βρίσκονται σε διεμπλοκή ακολουθίας στο m1, ενώ στο m2 βρίσκονται σε αυστηρή ακολουθία. Επιπλέον, η δραστηριότητα b λείπει στο m2, το οποίο συνεπάγεται επιπλέον ζεύγη για τη σχέση αυστηρής ακολουθίας στο B του m1. Οι σχέσεις αυστηρής ακολουθίας ανάμεσα σε (a,c), (a,d), (a,e), (a,f), και (a,g) δεν επηρεάζονται ωστόσο από αυτό. Αυτό οδηγεί στην ακόλουθη ομοιότητα αυστηρής ακολουθίας : $\text{sim}_{\rightsquigarrow}(\mathcal{B}m1, \mathcal{B}m2) = \frac{6}{16} = 0.375$

Ορισμός Ομοιότητας Διεμπλοκής ακολουθίας

Διεμπλοκή ακολουθίας είναι η πιο αδύναμη σχέση μεταξύ δύο δραστηριοτήτων, δεδομένου ότι αναφέρει ότι μπορούν μόνο να εκτελεστούν με οποιαδήποτε σειρά σε μία εκτέλεση της διαδικασίας. Έτσι, η ομοιότητα διεμπλοκής ακολουθίας επίσης ανταμείβει τα ζεύγη που αντιστοιχίζονται εάν για παράδειγμα εκτελούνται παράλληλα σε μία διαδικασία και ως μέρος μίας κυκλικής ροής δραστηριοτήτων στην άλλη διαδικασία.

P, Q είναι μοντέλα διαδικασίας και $||p, ||q$ οι σχέσεις διεμπλοκής ακολουθίας των αντίστοιχων προφίλ συμπεριφοράς τους $\mathcal{B}P$ και $\mathcal{B}Q$.

Ομοιότητα διεμπλοκής ακολουθίας ορίζεται ως εξής:

$$\text{sim}_{||}(\mathcal{B}P, \mathcal{B}Q) = \frac{|||P \cap ||Q|}{|||P \cup ||Q|}$$

Τα μοντέλα M1 και M2 του παραδείγματος μοιράζονται διεμπλοκή ακολουθίας για τα ζεύγη (d,e), (e,d), (d,f) και (f,d) και επομένως η ομοιότητά τους καθορίζεται ως εξής $\text{sim}_{||}(\mathcal{B}m1, \mathcal{B}m2) = \frac{4}{8} = 0.5$

Για τις παραπάνω βασικές μετρικές ακολουθείται ένα ενοποιημένο μέτρο της ομοιότητας του προφίλ συμπεριφοράς. Κάθε στοιχειώδης ομοιότητα μεταφράζεται σε μια στοιχειώδη μετρική $d_h(\mathcal{B}P, \mathcal{B}Q) = 1 - \text{Sim}_h(\mathcal{B}P, \mathcal{B}Q)$; όπου $h \in \{+, ||, \rightsquigarrow\}$

Στη συνέχεια, προσθέτουμε αυτές τις αξίες και ορίζουμε βάρη που αντιπροσωπεύουν τον αντίκτυπο της αντίστοιχης μετρική στη συνολική μετρική. Επομένως η ενοποιημένη μετρική υπολογίζεται ως εξής:

$$dh(BP, BQ) = 1 - \sum_h \omega_h \cdot \text{Sim}_h(BP, BQ)$$

$$\omega_h \in \mathbb{R}, 0 < \omega_h < 1 \text{ έτσι ώστε } \sum_h \omega_h = 1$$

5. Εργαλεία

Στην ενότητα αυτή παρουσιάζονται ορισμένες ολοκληρωμένες προσεγγίσεις αντιστοίχισης μοντέλων επιχειρησιακών διαδικασιών, που συμπεριλαμβάνουν τον ορισμό μίας μετρικής για την εύρεση της ομοιότητας δύο γραφημάτων, καθώς και έναν αλγόριθμο ή μέθοδο αντιστοίχισης γράφων για τον μαθηματικό υπολογισμό αυτής της μετρικής. Οι επιχειρήσεις που διατηρούν βάσεις δεδομένων με μοντέλα επιχειρησιακών διαδικασιών σε κάποια μορφή αναπαράστασης, είναι περισσότερο από λογικό να ζητούν την καλύτερη ολοκληρωμένη προσέγγιση αντιστοίχισης μοντέλων επιχειρησιακών διαδικασιών, ώστε να την χρησιμοποιήσουν σε ερωτήματα ομοιότητας κάποιου πρότυπου μοντέλου με όλα τα γραφήματα που διατηρούνται στη βάση δεδομένων τους.

5.1 Τεχνική «Triple-S»

Η πρώτη ολοκληρωμένη προσέγγιση αντιστοίχισης επιχειρησιακών μοντέλων είναι η τεχνική «Triple-S», η οποία είναι μία σχετικά απλή τεχνική χωρίς ιδιαίτερη πολυπλοκότητα. Η τεχνική αυτή συνδυάζει το βαθμό ομοιότητας που υπολογίζεται σε τρία ανεξάρτητα επίπεδα, για να επιτύχει μία καλά τεκμηριωμένη απόφαση σχετικά με την αντιστοίχιση μοντέλων επιχειρησιακών διαδικασιών. Τα τρία επίπεδα που παράγουν το βαθμό (σκορ) ομοιότητας είναι τα εξής:

- α) Συντακτικό επίπεδο - SIMsyn(a,b): Στο επίπεδο αυτό πραγματοποιείται συντακτική ανάλυση των ετικετών των ονομασιών των στοιχείων των μοντέλων, αφού πρώτα ολοκληρωθούν δύο βήματα προεπεξεργασίας, τα οποία είναι αφενός ο διαχωρισμός των λέξεων, των ονομασιών των στοιχείων ετικετών, και αφετέρου η εξάλειψη των σημείων στίξης και των συνδετικών λέξεων από τις ετικέτες (τεχνική stop word elimination). Η συντακτική ανάλυση βασίζεται στον υπολογισμό των αποστάσεων 'Levenshtein' μεταξύ όλων των συνδυασμών λέξεων από τις ονομασίες των στοιχείων a και b. Η τελική συντακτική βαθμολογία που υπολογίζεται είναι η ελάχιστη απόσταση λέξεων, διαιρεμένη με το σύνολο των λέξεων των ονομασιών των δύο στοιχείων, δηλαδή ή ελάχιστη μέση απόσταση σε κάθε λέξη.

β) Σημασιολογικό επίπεδο - SIMsem(a,b): Στο επίπεδο αυτό πραγματοποιούνται τα ίδια βήματα προεπεξεργασίας στις ονομασίες των στοιχείων των μοντέλων, όπως και στο συντακτικό επίπεδο. Ακολούθως, εφαρμόζεται ένας προσεγγιστικός υπολογισμός της σημασιολογικής ομοιότητας μεταξύ των λέξεων δύο στοιχείων a και b, που βασίζεται στο μήκος της απόστασης που έχουν αντίστοιχες έννοιες. Η τελική σημασιολογική βαθμολογία είναι η μέγιστη μέση ομοιότητα, που υπολογίζεται κατ' ανάλογο τρόπο με την τελική συντακτική βαθμολογία.

γ) Δομικό επίπεδο - SIMstruc(a,b): Στο επίπεδο αυτό διερευνάται η ομοιότητα των στοιχείων a και b μέσω μίας σύγκρισης της αναλογικότητας των εισερχόμενων και εξερχόμενων ακμών τους, όπως και της σχετικής τους θέσης σε ολόκληρο το μοντέλο.

Οι τρεις βαθμολογίες που προκύπτουν από τα επίπεδα αυτά συνδυάζονται για να παραχθεί η τελική βαθμολογία SIMtotal(a,b), που αντιπροσωπεύει το βαθμό αντιστοιχίσεως μεταξύ δύο στοιχείων a και b από διαφορετικά μοντέλα επιχειρησιακών διαδικασιών, και που υπολογίζεται από τον ακόλουθο τύπο:

$$\text{SIMtotal}(a,b) = w1 * \text{SIMsyn}(a,b) + w2 * \text{SIMsem}(a,b) + w3 * \text{SIMstruc}(a,b)$$

Στον τύπο αυτό, οι τρεις παράμετροι w1, w2 και w3 ορίζουν το βαθμό που συνεισφέρει κάθε επίπεδο ομοιότητας στην τελική βαθμολογία, ενώ ταυτόχρονα μία τιμή κατωφλίου θ χρησιμοποιείται για να καθορίσει αν θα θεωρηθεί ότι τα στοιχεία a και b αντιστοιχίζονται θετικά μεταξύ τους, στην περίπτωση που $\text{SIMtotal}(a,b) \geq \theta$.

5.2 Τεχνική «RefMod-Mine/NSCM»

Η ολοκληρωμένη προσέγγιση αντιστοιχίσεως επιχειρησιακών μοντέλων αυτή είναι ουσιαστικά μία τεχνική ομαδοποίησης των αντιστοιχίσεων των στοιχείων δύο επιχειρησιακών μοντέλων, η οποία επιτυγχάνεται εκτελώντας διαδοχικά τέσσερα βήματα. Η τεχνική αυτή, αρχικά, πραγματοποιεί μία σημασιολογική ανίχνευση σφαλμάτων, κατά την οποία αναγνωρίζει και διορθώνει τυχόν ελαττώματα των επιχειρησιακών μοντέλων. Στην συνέχεια εισάγει τα μοντέλα σε έναν εργαλείο που ομαδοποιεί τις αντιστοιχίσεις των στοιχείων των μοντέλων τους, χρησιμοποιώντας ένα σημασιολογικό μέτρο ομοιότητας. Το αποτέλεσμα της ομαδοποίησης από το εργαλείο είναι ένα σύνολο από ομάδες δυαδικών αντιστοιχίσεων που προέρχονται από όλα τα μοντέλα εισόδου, και που έχουν παρόμοιο σημασιολογικό μέτρο ομοιότητας. Στις επόμενες παραγράφους παρουσιάζονται ορισμένα ιδιαίτερα τεχνικά χαρακτηριστικά των τεσσάρων βημάτων της τεχνικής «RefMod-Mine/NSCM».

Η κύρια λειτουργία της σημασιολογικής ανίχνευσης σφαλμάτων είναι η αναγνώριση λαθών που έχουν πραγματοποιηθεί κατά την αναπαράσταση μεταβάσεων των μοντέλων επιχειρησιακών διαδικασιών, σε κόμβους των γραφημάτων. Η λανθασμένη μοντελοποίηση του συγκεκριμένου τύπου κόμβων, είναι πιθανόν να οδηγήσει σε παρανοήσεις κατά τη διάρκεια της αντιστοίχισης των κόμβων. Για τον λόγο αυτό η σημασιολογική ανίχνευση σφαλμάτων αναλύει τη μορφή και την σειρά των ουσιαστικών και των ρημάτων της ονομασίας κάθε κόμβου (εφαρμόσιμη μόνο για ετικέτες γραμμένες στην αγγλική γλώσσα), προσπαθώντας να αναγνωρίσει το κατά πόσο κάθε κόμβος συμβολίζει μία μετάβαση. Στην περίπτωση που αποφανθεί ότι ένας κόμβος συμβολίζει μία μετάβαση, τον επισημαίνει ώστε να μην ληφθεί υπόψιν στα επόμενα βήματα της τεχνικής, για να απλοποιηθεί η διαδικασία της αντιστοίχισης των κόμβων.

Σε αντίθεση με άλλες τεχνικές αντιστοίχισης επιχειρησιακών μοντέλων που αντιστοιχίζουν τα στοιχεία μοντέλων δυαδικά, η συγκεκριμένη τεχνική επιλέγει να ομαδοποιήσει κάποιες από τις δυαδικές αντιστοιχίσεις κόμβων σε σύνολα. Κατά την ομαδοποίηση αυτή, συγκρίνονται ανά δύο όλα τα στοιχεία από όλα τα μοντέλα εισόδου, χρησιμοποιώντας ένα σημασιολογικό μέτρο ομοιότητας, και αντιστοιχίζονται σε περίπτωση που είναι παρόμοια σημασιολογικά. Η διαδικασία ομαδοποίησης αρχικά δημιουργεί ομάδες που περιλαμβάνουν μόνο μία δυαδική αντιστοίχιση στοιχείων. Στη συνέχεια εντάσσει μία αντιστοίχιση σε κάποια ομάδα που περιλαμβάνει ήδη μία άλλη αντιστοίχιση, μόνο στην περίπτωση που το σημασιολογικό μέτρο ομοιότητας των στοιχείων της νέας αντιστοίχισης διαφέρει ελάχιστα από το σημασιολογικό μέτρο ομοιότητας των στοιχείων της αντιστοίχισης που προϋπάρχει στην ομάδα (Αυτό ελέγχεται από μία τιμή κατωφλίου που δίνετε σαν παράμετρος στην διαδικασία της ομαδοποίησης). Επίσης, σε περίπτωση που δύο δυαδικές αντιστοιχίσεις στοιχείων από διαφορετικές ομάδες, έχουν παράξει παρόμοια σημασιολογικά μέτρα ομοιότητας, οι ομάδες τους συγχωνεύονται.

Το σημασιολογικό μέτρο ομοιότητας δύο στοιχείων που χρησιμοποιείται κατά την διαδικασία της ομαδοποίησης υπολογίζεται σε τρία βήματα. Στο πρώτο βήμα διασπώνται σε μεμονωμένες λέξεις w_iL οι ονομασίες L των δύο στοιχείων (χωρίς να λαμβάνονται υπόψιν συνδετικές λέξεις), ούτως ώστε $split(L) = \{w_1L, \dots, w_nL\}$. Στο δεύτερο βήμα υπολογίζεται για κάθε λέξη η ρίζα της (τεχνική stemming), $stem(w_iL)$, και συγκρίνονται τα σύνολα με τις ρίζες όλων των λέξεων των δύο ονομασιών.

Στο τρίτο βήμα επίσης ελέγχονται οι ονομασίες των δύο κόμβων για την ύπαρξη αντωνύμων ή λέξεων άρνησης. Σε περίπτωση που υπάρχουν λέξεις άρνησης μόνο στη μία από τις δύο

ετικέτες, τότε αυτομάτως το σημασιολογικό μέτρο ομοιότητας παίρνει τιμή 0, ενώ στην αντίθετη περίπτωση διατηρεί την τιμή $\text{sim}(L1,L2)$.

5.3 Ανάλυση των Εργαλείων και Ευρήματα

Στην ενότητα αυτή παρουσιάζονται κάποιες γενικές παρατηρήσεις που έχουν προκύψει μετά την ανάλυση και αξιολόγηση διαφορετικών λύσεων για την αντιστοίχιση επιχειρησιακών διαδικασιών συμπεριλαμβανομένων των λύσεων που αναφέρθηκαν στην προηγούμενη ενότητα [2].

Μία πρώτη σύγκριση των επτά ολοκληρωμένων λύσεων μπορεί να επιχειρηθεί με κριτήριο αξιολόγησης τα επίπεδα εύρεσης ομοιοτήτων που αξιοποιεί κάθε τεχνική κατά την διάρκεια επιλογής των αντιστοιχίσεων που προτείνει. Στον Πίνακα 1 που ακολουθεί δίνεται μία επισκόπηση για το αν το συντακτικό, το σημασιολογικό ή το δομικό επίπεδο ομοιότητας χρησιμοποιείται από κάποια από τις επτά τεχνικές, στον υπολογισμό του βαθμού ομοιότητας δύο μοντέλων επιχειρησιακών διαδικασιών.

Αριθμός	Τεχνική	Συντακτικό	Σημασιολογικό	Δομικό
1	Triple-S	✓	✓	✓
2	BPG Matching	✓	✓	
3	RefMod/NSCM		✓	
4	RefMod/ESGM	✓	✓	
5	Bag-of-Words	✓	✓	
6	PMLM			✓
7	ICoP	✓	✓	✓

Πίνακας 1: Επίπεδα Εύρεσης Ομοιοτήτων των Επτά Τεχνικών Αντιστοίχισης

Είναι προφανές ότι οι τεχνικές που αξιοποιούν και τις συντακτικές, και τις σημασιολογικές, και τις δομικές πληροφορίες των μοντέλων, κατά τον υπολογισμό του βαθμού ομοιότητάς τους, δηλαδή οι τεχνικές «Triple-S» και «ICoP Framework», θα προτιμώνται από τις επιχειρήσεις και τους οργανισμούς σε περιπτώσεις που τα μοντέλα επιχειρησιακών διαδικασιών τους μπορούν να αντιστοιχιστούν από όλες τις τεχνικές με παρόμοιο ποσοστό ακρίβειας και ανάκλησης. Επίσης διαπιστώθηκε ότι δεν υπάρχει κάποια ολοκληρωμένη τεχνική αντιστοίχισης που να παρέχει τα βέλτιστα αποτελέσματα, σε όλα ανεξαιρέτως τα προβλήματα, εξαιτίας φυσικά και της διαφορετικής φύσης του συνόλου δεδομένων τους.

Επομένως, παραμένει ανοικτή η πρόκληση της κατασκευής ενός πρωτότυπου εργαλείου αντιστοίχισης μοντέλων επιχειρησιακών διαδικασιών, το οποίο θα προσφέρει καλύτερα αποτελέσματα σε σύγκριση με όλες τις υπάρχουσες ολοκληρωμένες τεχνικές, τόσο σε

επίπεδο χρόνου εκτέλεσης και χαρακτηριστικών για την ομοιότητα των μοντέλων που θα εξετάζει, όσο κυρίως σε επίπεδο συμβιβασμού μεταξύ ακρίβειας και ανάκλησης των αποτελεσμάτων που θα παρέχει, και μάλιστα θα επιτυγχάνει αυτά τα καλύτερα αποτελέσματα για οποιοδήποτε πρόβλημα αντιστοίχισης επιχειρησιακών διαδικασιών κληθεί να αντιμετωπίσει.

Αναφορές

- [1] Bunke, H. and Allermann, G., 1983. Inexact graph matching for structural pattern recognition. *Pattern Recognition Letters*, 1(4), pp.245-253
- [2] Cayoglu, U., Dijkman, R., Dumas, M., Fettke, P., García-Bañuelos, L., Hake, P., Klinkmüller, C., Leopold, H., Ludwig, A., Loos, P. and Mendling, J., 2013, August. Report: The process model matching contest 2013. In *International Conference on Business Process Management* (pp. 442-463). Springer International Publishing
- [3] Dijkman, R., 2007, October. A classification of differences between similar BusinessProcesses. In *Enterprise Distributed Object Computing Conference, 2007. EDOC 2007. 11th IEEE International* (pp. 37-37). IEEE
- [4] Dijkman, R., Dumas, M., Van Dongen, B., Käärik, R. and Mendling, J., 2011. Similarity of business process models: Metrics and evaluation. *Information Systems*, 36(2), pp.498-516
- [5] Dumas, M., García-Bañuelos, L. and Dijkman, R.M., 2009. Similarity Search of Business Process Models. *IEEE Data Eng. Bull.*, 32(3), pp.23-28
- [6] Gao, X., B. Xiao, et al. (2010). "A survey of graph edit distance." *Pattern Analysis and applications* 13(1): 113-129
- [7] Lipkus, A.H., 1999. A proof of the triangle inequality for the Tanimoto distance. *Journal of Mathematical Chemistry*, 26(1-3), pp.263-265
- [8] Neuhaus, M., Riesen, K. and Bunke, H., 2006, August. Fast suboptimal algorithms for the computation of graph edit distance. In *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)* (pp. 163-172). Springer Berlin Heidelberg
- [9] Ullmann, J.R., 1976. An algorithm for subgraph isomorphism. *Journal of the ACM (JACM)*, 23(1), pp.31-42