Οικονομικό Πανεπιστήμιο Αθηνών

Μεταπτυχιακό Πρόγραμμα Σπουδών

στην Επιστήμη των Υπολογιστών

Διπλωματική εργασία

Αυτόματη Παραγωγή Συγκρίσεων σε ένα Σύστημα Παραγωγής Κειμένων Φυσικής Γλώσσας

Γιώργος Καρακατσιώτης

Επιβλέπων: Ίων Ανδρουτσόπουλος

Αθήνα, Μάιος 2007

Περιεχόμενα

Περιεχόμενα 2

Περίληψη 4

1. Εισαγωγή 5

1.1 Αντικείμενο της εργασίας 5

1.2 Διάρθρωση της εργασίας 8

1.3 Ευχαριστίες 9

2. Περιγραφή και σύγκριση προηγούμενων μεθόδων 10

2.1 Είδη συγκρίσεων 10

2.1.1 Ευθεία σύγκριση 10

2.1.2 Σύγκριση ειδοποιού διαφοράς τύπων 12

2.1.3 Σύγκριση με συμπλήρωμα συνόλου 13

2.1.4 Σύγκριση αποσαφήνισης 15

2.1.5 Αναφορά σε παρόμοια οντότητα/τύπο 16

2.1.6 Αναλογία 17

2.2 Αλγόριθμος Milosavljevic 18

2.2.1 Σύγκριση αποσαφήνισης και αναφορά σε παρόμοια οντότητα/τύπο 20

2.2.2 Ευθεία σύγκριση 21

2.2.3 Σύγκριση ειδοποιού διαφοράς 21

2.2.4 Σύγκριση με συμπλήρωμα συνόλου 22

2.2.5 Αναλογία 22

2.3 Μέτρα σύγκρισης 23

2.4 Ο αλγόριθμος του Μελέγκογλου 23

2.4.1 Αναπαράσταση οντοτήτων που έχουν ήδη περιγραφεί 24

2.4.2 Βασικές ιδέες του αλγορίθμου 25

2.4.3 Ο αλγόριθμος 28

2.4.4 Πρώτο στάδιο 29

2.4.5 Δεύτερο στάδιο 31

2.4.6 Τρίτο στάδιο 35

3. Το σύστημά μας 37

3.1 Επιλογή αλγορίθμου 37

3.1.1 Ευθεία σύγκριση 38

3.1.2 Σύγκριση ειδοποιού διαφοράς τύπων 39

3.1.3 Σύγκριση με συμπλήρωμα συνόλου 40

3.1.4 Σύγκριση αποσαφήνισης 40

3.1.5 Αναφορά σε παρόμοια οντότητα/τύπο 41

3.1.6 Αναλογία 41

3.2 Τροποποιήσεις του αλγορίθμου Μελέγκογλου 42

3.2.1 Παραλλαγή του αλγορίθμου Μελέγκογλου 44

3.3 Συγκρίσεις με ολόκληρη τη συλλογή 46

4. Πειραματική αξιολόγηση ____ *50*

4.1 Τι θέλουμε να αξιολογήσουμε ____ **50**

4.2 Τρόπος εκτέλεσης των πειραμάτων ____ **50**

4.3 Πειραματικά αποτελέσματα ____ **53**

4.3.1 Συγκρίσεις με όλα τα εκθέματα της αίθουσας ____ 53

4.3.2 Συγκρίσεις με τον αλγόριθμο Μελέγκογλου ____ 61

5 Συμπεράσματα και μελλοντικές κατευθύνσεις ____ *72*

5.1 Συμπεράσματα ____ **72**

5.2 Μελλοντικές επεκτάσεις ____ **73**

Αναφορές ____ *75*

Περίληψη

Σε αυτή την εργασία προσπαθήσαμε να διευρύνουμε τις δυνατότητες μιας προϋπάρχουσας μηχανής παραγωγής κειμένων φυσικής γλώσσας. Η μηχανή αυτή παράγει περιγραφές αντικειμένων στην ελληνική και στην αγγλική γλώσσα από συμβολικές πληροφορίες αποθηκευμένες σε οντολογίες OWL. Στη διάρκεια της εργασίας προστέθηκε η δυνατότητα παραγωγής συγκρίσεων ανάμεσα στην οντότητα που περιγράφεται και στις οντότητες που έχουν ήδη περιγραφεί. Για να το πετύχουμε αυτό, χρησιμοποιήσαμε μια βελτιωμένη μορφή του αλγόριθμου παραγωγής συγκρίσεων που είχε αναπτυχθεί στη διάρκεια του ερευνητικού έργου M-PIRO. Επίσης δημιουργήσαμε μία παραλλαγή του αλγορίθμου, που παράγει συγκρίσεις με όλες τις οντότητες της οντολογίας, εντοπίζοντας χαρακτηριστικά της τρέχουσας οντότητας που είναι μοναδικά στην οντολογία ή που είναι κοινά για όλες τις οντότητες κάποιου υπο-τύπου.

Τέλος, διερευνήσαμε πειραματικά αν οι χρήστες προτιμούν τα κείμενα του νέου συστήματος, που περιέχουν συγκρίσεις, έναντι των αντιστοίχων κειμένων χωρίς συγκρίσεις. Για τις ανάγκες των πειραμάτων, χρησιμοποιήσαμε μία οντολογία που περιλαμβάνει πληροφορίες για μια συλλογή εκθεμάτων ενός ηλεκτρονικού μουσείου. Τα αποτελέσματα των πειραμάτων δείχνουν ότι η χρήση των συγκρίσεων κάνει εν γένει τις περιγραφές των κειμένων πιο αρεστές στους επισκέπτες του μουσείου.

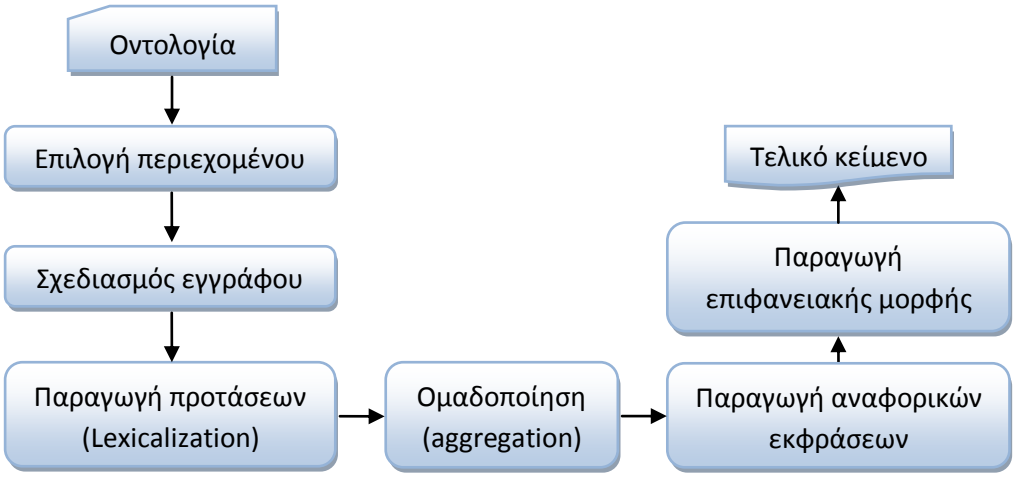
«Η αρχή είναι το ήμισυ του παντός»

1. Εισαγωγή

1.1 Αντικείμενο της εργασίας

Στη διάρκεια προηγούμενης διπλωματικής εργασίας (Γαλάνης, 2006), δημιουργήθηκε μία μηχανή παραγωγής κειμένων φυσικής γλώσσας (Reiter & Dale, 2006) που παράγει περιγραφές οντοτήτων (και τάξεων οντοτήτων) αντλώντας πληροφορίες από οντολογίες γραμμένες σε OWL[1]. Η μηχανή αυτή βασίζεται σε ιδέες του ερευνητικού έργου M-PIRO (Isard κ.ά., 2004, Androutsopoulos κ.ά. 2007) και επεκτάθηκε περαιτέρω στη διάρκεια του ερευνητικού έργου Ξένιος[2]. Παράγει κείμενα στα ελληνικά και στα αγγλικά, αλλά πολύ εύκολα μπορεί να επεκταθεί και σε άλλες γλώσσες.

Η γενική αρχιτεκτονική της μηχανής φαίνεται στο παρακάτω σχήμα:



Αρχικά αντλούνται από την οντολογία οι πληροφορίες (σε μορφή λογικών κατηγορημάτων) που είναι σχετικές με την οντότητα (ή τάξη) που πρόκειται να περιγραφεί, αποφεύγοντας πληροφορίες που έχουν ήδη δοθεί στο χρήστη. Αν οι πληροφορίες είναι περισσότερες από αυτές που επιθυμούμε (π.χ. αν υπάρχουν περιορισμοί ως προς το μέγεθος του κειμένου που θέλουμε να

[1] Βλ. http://www.w3.org/TR/owl-features/
[2] Βλ. http://www.ics.forth.gr/xenios/

παραχθεί), επιλέγονται από αυτές οι πιο ενδιαφέρουσες, βάσει μοντέλων χρηστών. Κατόπιν, στη διάρκεια του σχεδιασμού εγγράφου, αποφασίζεται η σειρά με την οποία θα αναφερθούν οι πληροφορίες στο κείμενο. Στο επόμενο βήμα, την παραγωγή προτάσεων, για κάθε πληροφορία (λογικό κατηγόρημα) παράγεται μία πρόταση φυσικής γλώσσας. Κατόπιν, οι προτάσεις ομαδοποιούνται σε πιο σύνθετες, ώστε να έχει μεγαλύτερη συνοχή το κείμενο. Για παράδειγμα, αντί να πούμε:

- ***Ο υπολογιστής Α έχει τον επεξεργαστή Υ.***
- ***Ο υπολογιστής Α φτιάχνεται στην Κίνα.***

θα μπορούσαμε να ενώσουμε τις δύο προτάσεις σε μία:

Ο υπολογιστής Α έχει τον επεξεργαστή Υ και φτιάχνεται στην Κίνα.

Οι παραγόμενες προτάσεις περιέχουν αρχικά κωδικούς οντοτήτων (από την οντολογία) αντί αναφορικών εκφράσεων φυσικής γλώσσας (π.χ. τον κωδικό της οντότητας στην οποία πρέπει να αναφέρεται το υποκείμενο, αντί για την έκφραση φυσικής γλώσσας που θα αποτελέσει τελικά το υποκείμενο της πρότασης). Στη συνέχεια, κατά την παραγωγή αναφορικών εκφράσεων, προσπαθούμε να αντικαταστήσουμε τους κωδικούς αυτούς με κατάλληλες εκφράσεις φυσικής γλώσσας, χωρίς όμως να επαναλαμβάνουμε συνεχώς τις ίδιες λέξεις (διαφορετικά το κείμενο θα γινόταν μονότονο) και χωρίς οι αναφορές να είναι αμφίσημες. Για παράδειγμα, αν θέλουμε να αναφερθούμε σε έναν κούρο, μπορούμε να χρησιμοποιήσουμε διάφορες εκφράσεις: «αυτό», «αυτό το έκθεμα», «αυτό το άγαλμα», «το προηγούμενο άγαλμα», «ο κούρος» κλπ.

Το τελευταίο στάδιο, η παραγωγή επιφανειακής μορφής, είναι ιδιαίτερα απλό στη μηχανή παραγωγής που χρησιμοποιούμε, γιατί η παραγωγή γίνεται χρησιμοποιώντας σχεδιότυπα (templates) αντί για «βαθύτερες» γλωσσικές παραστάσεις (π.χ. συντακτικά δέντρα) και τα κείμενα είναι ουσιαστικά ήδη έτοιμα από τα προηγούμενα στάδια. Απλά προστίθενται σημεία στίξεως, ετικέτες HTML (αν πρόκειται να παραχθεί ιστοσελίδα), ετικέτες που καθοδηγούν το συνθέτη φωνής (αν πρόκειται να ακολουθήσει σύνθεση φωνής) κλπ.

Οι επεκτάσεις της παρούσας εργασίας αφορούν κυρίως το πρώτο στάδιο, την επιλογή περιεχομένου. Εκτός από τα χαρακτηριστικά της τρέχουσας οντότητας που θα μεταδίδει το κείμενο, στη διάρκεια αυτού του σταδίου αποφασίζουμε πλέον και ποιες συγκρίσεις πρέπει να γίνουν μεταξύ των χαρακτηριστικών της τρέχουσας οντότητας και των αντιστοίχων χαρακτηριστικών άλλων οντοτήτων. Για την ακρίβεια, παράγεται πάντα το πολύ μία σύγκριση ανά περιγραφή οντότητας, γιατί διαφορετικά η περιγραφή γίνεται κουραστική, οπότε το ερώτημα που

τίθεται είναι ποια είναι η «καλύτερη» μεταξύ των συγκρίσεων που είναι δυνατόν να παραχθούν, βάσει ποιων κριτηρίων πρέπει να επιλέγεται, και πότε είναι προτιμότερο να μη γίνεται καμία σύγκριση. Εκτός από επεκτάσεις στο στάδιο επιλογής περιεχομένου, χρειάστηκε να γίνουν υποστηρικτικές τροποποιήσεις και σε άλλα στάδια. Για παράδειγμα, η ύπαρξη συγκρίσεων επηρεάζει το σχεδιασμό του εγγράφου, την παραγωγή των προτάσεων και, σε μερικές περιπτώσεις, την ομαδοποίηση.

Το πρώτο είδος συγκρίσεων που παράγει το νέο σύστημα είναι συγκρίσεις με οντότητες που έχουν περιγραφεί σε προηγούμενα κείμενα. Για το σκοπό αυτό χρησιμοποιείται μια βελτιωμένη μορφή του αντίστοιχου αλγορίθμου που είχε αναπτυχθεί στο M-PIRO από τον κ. Μελέγκογλου, συνεπιβλέποντα της εργασίας. Ο αλγόριθμος αυτός ομαδοποιεί τις οντότητες που είχαν περιγραφεί προηγουμένως με βάση τον τύπο τους (την τάξη τους). Έτσι αντί να προσπαθεί να συγκρίνει την τρέχουσα οντότητα με κάποια άλλη μεμονωμένη, βρίσκει εκείνη την ομάδα (τύπο) προηγούμενων οντοτήτων που είναι καταλληλότερη για σύγκριση, βάσει κριτηρίων που θα περιγραφούν σε επόμενα κεφάλαια. Με αυτόν τον τρόπο παράγονται συγκρίσεις όπως:

> ***Όπως η προηγούμενη λήκυθος, (αυτή η λήκυθος) δημιουργήθηκε κατά την κλασική περίοδο.***
>
> ***Αντίθετα από όλα τα προηγούμενα αγγεία, που ακολουθούσαν τον ερυθρόμορφο ρυθμό, αυτή η λήκυθος ακολουθεί το μελανόμορφο ρυθμό.***

Επεκτείναμε αυτόν τον αλγόριθμο ώστε να είναι δυνατόν να κάνει συγκρίσεις και με την πλειονότητα των οντοτήτων ενός τύπου:

> ***Όπως οι περισσότεροι από τους προηγούμενους κούρους κατασκευάστηκε κατά την κλασική περίοδο.***

Επιπλέον, αναπτύξαμε ένα δεύτερο αλγόριθμο που παράγει συγκρίσεις με το σύνολο των οντοτήτων της οντολογίας, αντί αυτών που έχουν περιγραφεί σε προηγούμενα κείμενα. Έτσι, μπορούμε πια να παράγουμε συγκρίσεις μοναδικότητας για ολόκληρη την οντολογία:

> ***Αυτός είναι ο μόνος κούρος της συλλογής (ή της αίθουσας).***

Καθώς και συγκρίσεις ομοιότητας με όλες τις οντότητες ενός τύπος (ή την πλειονότητα των οντοτήτων ενός τύπου):

Όπως όλοι οι παναθηναϊκοί αμφορείς της συλλογής (ή της αίθουσας), προέρχεται από την Αττική.

Όπως οι περισσότεροι παναθηναϊκοί αμφορείς της συλλογής (ή της αίθουσας), ακολουθεί τον ερυθρόμορφο ρυθμό

Τέλος, αξιολογήσαμε πειραματικά το κατά πόσον οι παραγόμενες συγκρίσεις κάνουν τα κείμενα πιο ελκυστικά στους αναγνώστες. Συγκεκριμένα, δημιουργήσαμε τρεις εικονικές αίθουσες με εκθέματα μιας μουσειακής συλλογής, χρησιμοποιώντας την αντίστοιχη οντολογία του M-PIRO. Οι χρήστες (επισκέπτες της συλλογής) έπρεπε να επιλέξουν ο καθένας 10 εκθέματα μίας αίθουσας, με όποια σειρά ήθελαν. Σε γενικές γραμμές, κάθε φορά που επέλεγαν κάποιο έκθεμα, εμφανίζονταν δύο περιγραφές: μία που δεν περιείχε καμία σύγκριση και μία που περιείχε μία σύγκριση που είχε παραχθεί με έναν από τους δύο παραπάνω αλγορίθμους. Οι δύο περιγραφές ήταν κατά τα άλλα ίδιες. Οι ερωτηθέντες έπρεπε να διαλέξουν ποια περιγραφή τούς άρεσε περισσότερο και προαιρετικά να αιτιολογήσουν την επιλογή τους. Στα δύο πειράματα, που περιγράφονται αναλυτικότερα σε επόμενο κεφάλαιο, έλαβαν συνολικά μέρος 87 και 88 φοιτητές, αντίστοιχα, του Τμήματος Πληροφορικής του Ο.Π.Α. Τα πειραματικά αποτελέσματα έδειξαν ότι οι παραγόμενες συγκρίσεις κάνουν εν γένει πιο ελκυστικά τα κείμενα όταν χρησιμοποιείται ο αλγόριθμος του Μελέγκογλου, ιδιαίτερα στην περίπτωση αρνητικών συγκρίσεων («Αντίθετα από ...»), ενώ οι συγκρίσεις του δεύτερου αλγορίθμου (που συγκρίνει με όλες της οντότητες της οντολογίας) είναι λιγότερο αποδεκτές από τους επισκέπτες.

1.2 Διάρθρωση της εργασίας

Η υπόλοιπη εργασία είναι χωρισμένη σε 4 κεφάλαια:

- Το Κεφάλαιο 2 παρέχει το θεωρητικό υπόβαθρο στο οποίο στηρίζεται η εργασία και χωρίζεται σε δύο μέρη. Στο πρώτο μέρος παρουσιάζονται διάφορα είδη συγκρίσεων, όπως τα έχει παρουσιάσει η Milosavljevic (1999). Στο δεύτερο μέρος παρουσιάζονται δύο αλγόριθμοι παραγωγής συγκρίσεων, ένας που έχει προταθεί από τη Milosavljevic (1999) και εκείνος του Μελέγκογλου.
- Στο Κεφάλαιο 3 περιγράφονται οι λόγοι για τους οποίους επιλέξαμε τον αλγόριθμο του Μελέγκογλου και ο τρόπος με τον οποίο παράγει συγκρίσεις το σύστημά μας. Δεν ασχολούμαστε με το πώς θα παραχθεί το τελικό κείμενο, αλλά μόνο με τις επεμβάσεις που κάναμε στο στάδιο της επιλογής περιεχομένου, όπως ορίσθηκε στην προηγούμενη ενότητα.

- Στο Κεφάλαιο 4 αναλύονται διεξοδικά ο τρόπος με τον οποίο κάναμε τα πειράματά μας, καθώς και τα αποτελέσματά τους.

- Στο Κεφάλαιο 5 παρουσιάζονται πιθανές μελλοντικές επεκτάσεις και βελτιώσεις του συστήματος, σύμφωνα με τα συμπεράσματα που προέκυψαν από την εργασία.

1.3 Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον επιβλέποντα της εργασίας κ. Ίωνα Ανδρουτσόπουλο για την καθοδήγησή του. Επίσης, θα ήθελα να ευχαριστήσω τον κ. Αλέξανδρο Μελέγκογλου, συν-επιβλέποντα της εργασίας, που με βοήθησε στην καλύτερη κατανόηση του αλγορίθμου του και για τα γόνιμα σχόλιά του. Ακόμα, θα ήθελα να ευχαριστήσω όλους τους φοιτητές που πήραν μέρος στα πειράματα και την οικογένειά μου που με ανέχτηκε και με στήριξε κατά τη διάρκεια όλης της προσπάθειας. Ένα μεγάλο ευχαριστώ στο Δημήτρη Γαλάνη, που κάθε φορά που του ζητούσα νέα έκδοση του συστήματός του, μου την έστελνε άμεσα, και με βοηθούσε στα σημεία που είχα πρόβλημα με τον κώδικά του. Τέλος, ευχαριστώ τον κ. Παναγιώτη Κωνσταντόπουλο, που αποδέχτηκε τον ρόλο του δεύτερου αξιολογητή.

«Όποιος δεν ξέρει ιστορία,
είναι καταδικασμένος να την επαναλάβει»
Harry Truman

2. Περιγραφή και σύγκριση προηγούμενων μεθόδων

2.1 Είδη συγκρίσεων

Σύμφωνα με την Milosavljevic (1997a, 1997b, 1999, Milosavljevic & Dale, 1996) μπορούμε να διακρίνουμε 6 διαφορετικούς τύπους συγκρίσεων (εντός παρενθέσεων δίνονται οι όροι που χρησιμοποιεί η Milosavljevic):

- Ευθεία σύγκριση (Direct Comparison)
- Σύγκριση ειδοποιού διαφοράς τύπων (Significant Type Comparison)
- Σύγκριση με συμπλήρωμα συνόλου (Set Complement Comparison)
- Σύγκριση αποσαφήνισης (Clarificatory Comparison)
- Αναφορά σε παρόμοια οντότητα/τύπο (Like-entity Comparison)
- Αναλογία (Illustrative Comparison)

Οι συγκρίσεις αυτές προέκυψαν μετά από μελέτη άρθρων της εγκυκλοπαίδειας Encarta 95[3]. Εξετάζουμε στη συνέχεια αναλυτικότερα καθέναν από τους παραπάνω τύπους, βασιζόμενοι στην ανάλυση της Milosavljevic.

2.1.1 Ευθεία σύγκριση

Έστω ότι κάποιος θέλει να αγοράσει ένα φορητό υπολογιστή. Για να επιλέξει αυτόν που ικανοποιεί καλύτερα τις ανάγκες του, θα πρέπει να συγκρίνει τα διάφορα μοντέλα μεταξύ τους.

[3] http://www.microsoft.com/products/encarta/default.mspx

Μοντέλο Α

Φορητός υπολογιστής με επεξεργαστή Χ και 512MB μνήμης DDR2 533MHz, με κάρτα γραφικών Υ και αυτονομία 90'. Χωρητικότητα σκληρού δίσκου 100 GB.

Μοντέλο Β

Φορητός υπολογιστής με 1024MB μνήμης DDR 400MHz και επεξεργαστή Ζ, με κάρτα γραφικών ενσωματωμένη και σκληρό δίσκο 120 GB. Αυτονομία 150'.

Διαβάζοντας τις παραπάνω περιγραφές δεν θα βοηθηθεί ιδιαίτερα. Αυτό που απαιτείται είναι μία σύγκριση των χαρακτηριστικών που έχει το κάθε μοντέλο:

	Μοντέλο Α	Μοντέλο Β
Επεξεργαστής	Χ	Ζ
Ποσότητα Μνήμης	512	1024
Τύπος Μνήμης	DDR2 533MHz	DDR 400MHz
Κάρτα γραφικών	Υ	Ενσωματωμένη
Σκληρός δίσκος	100	120
Αυτονομία	90'	150'

Με αυτόν τον τρόπο ο αγοραστής θα μπορέσει ευκολότερα να καταλάβει τα ιδιαίτερα χαρακτηριστικά κάθε μοντέλου και θα σχηματίσει πιο άρτια εικόνα για το κάθε μοντέλο, χωρίς να τα συγχέει μεταξύ τους.

Μία τέτοια περιγραφή φυσικά μπορεί να γίνει και με κείμενο φυσικής γλώσσας αντί για πίνακες. Οι πίνακες είναι πιο ευανάγνωστοι, ειδικά στις περιπτώσεις στις οποίες υπάρχουν πολλά χαρακτηριστικά, αλλά μπορεί να μην είναι δυνατή η χρήση πινάκων, για παράδειγμα όταν η περιγραφή πρέπει να εκφωνηθεί. Κάθε γραμμή του πίνακα μπορεί να θεωρηθεί ότι αποτελεί και από μόνη της μία σύγκριση. Για παράδειγμα, η δεύτερη γραμμή θα μπορούσε να εκφραστεί σε φυσική γλώσσα ως εξής: «Το μοντέλο Β έχει 1024 MB RAM, ενώ το Α έχει 512 MB RAM».

Γενικά η μία-προς-μία σύγκριση των χαρακτηριστικών βοηθά κυρίως να απαντηθούν ερωτήσεις της μορφής «Ποιες είναι οι ομοιότητες και οι διαφορές ανάμεσα στα Χ και Υ;» ή «Σύγκρινε τα Χ και Υ». Συνήθως όταν ο χρήστης επιλέγει να συγκρίνει δύο συγκεκριμένες οντότητες (από ένα ευρύτερο σύνολο), απαιτεί μια λεπτομερή περιγραφή τους και επιπλέον θέλει να μάθει ποιες

ιδιότητες είναι κοινές και ποιες είναι ιδιαίτερες για κάθε οντότητα. Στην περίπτωση αυτή μιλάμε για *Ευθεία Σύγκριση*.

Η Ευθεία Σύγκριση μπορεί να γίνει και μεταξύ πολλών οντοτήτων, αλλά σε κείμενα φυσικής γλώσσας γίνεται συνήθως μεταξύ μόνο δύο.

Στις ευθείες συγκρίσεις κάθε οντότητα-μέλος της σύγκρισης είναι εξίσου σημαντικό και μας ενδιαφέρει να πληροφορηθούμε πλήρως για όλες τους τις ιδιότητες. Αντίθετα στα υπόλοιπα είδη συγκρίσεων, που θα αναλύσουμε στη συνέχεια, συνήθως υπάρχει μία κυρίαρχη οντότητα (αυτή στην οποία εστιαζόμαστε) και κάποιες άλλες που μας βοηθούν να κατανοήσουμε καλύτερα την κυρίαρχη οντότητα, ενώ η σύγκριση ενδέχεται να μην αναφέρει όλες τις ιδιότητες των συγκρινόμενων οντοτήτων. Τα κείμενα που παράγονται από την ευθεία σύγκριση είναι συνήθως μακροσκελή και αυτοτελή, ενώ τα κείμενα των άλλων ειδών συγκρίσεων είναι συχνά συντομότερα και αποτελούν τμήματα μεγαλύτερων κειμένων.

Επίσης, μια ευθεία σύγκριση γίνεται συνήθως μεταξύ παρόμοιων οντοτήτων. Για παράδειγμα, μπορεί να γίνει μεταξύ ενός φορητού και ενός σταθερού υπολογιστή, που είναι και οι δύο υπολογιστές. Όσο πιο διαφορετικοί είναι οι τύποι των συγκρινόμενων οντοτήτων, τόσο πιο περίεργη φαίνεται μια ευθεία σύγκριση (π.χ. υπολογιστής και καφετιέρα). Σημειώνουμε, ακόμη, ότι ευθείες συγκρίσεις είναι δυνατόν να γίνουν και μεταξύ *τύπων* (κατηγοριών, τάξεων) οντοτήτων, όχι μόνο μεταξύ μεμονωμένων οντοτήτων. Για παράδειγμα, μια ευθεία σύγκριση είναι δυνατόν να συγκρίνει ένα-προς-ένα τα (τυπικά) χαρακτηριστικά των επιτραπέζιων και των φορητών υπολογιστών.

2.1.2 Σύγκριση ειδοποιού διαφοράς τύπων

> ***Οι φορητοί υπολογιστές μπορούν να χωρισθούν σε δύο κατηγορίες: σε αυτούς που έχουν μικρό μέγεθος και μπορούν εύκολα να μετακινηθούν και σε αυτούς που είναι βαρύτεροι και προορίζονται για την αντικατάσταση σταθερών συστημάτων.***

Στο παραπάνω παράδειγμα μπορούμε να δούμε μία σύγκριση ανάμεσα σε διαφορετικές υπο-κατηγορίες (υπο-τύπους, υπο-τάξεις) φορητών υπολογιστών. Στόχος του αποσπάσματος είναι να μας ενημερώσει για το πλήθος των διαφορετικών υπο-κατηγοριών φορητών υπολογιστών, καθώς και για την ειδοποιό διαφορά τους, το χαρακτηριστικό που τους κάνει να ξεχωρίζουν μεταξύ τους (στο παράδειγμά μας, το πόσο εύκολα μπορούν να μετακινηθούν). Αντίστοιχες συγκρίσεις μπορούν να γίνουν και για περισσότερους από δύο υπο-τύπους οντοτήτων (π.χ. οι διαφορετικές εκδόσεις ενός προγράμματος: basic, professional, enterprise).

Σημαντικό στοιχείο αυτού του είδους των συγκρίσεων είναι ότι όλοι οι υπο-τύποι οντοτήτων που αναφέρονται είναι εξίσου σημαντικοί. Ωστόσο ο λόγος που χρησιμοποιείται η σύγκριση είναι η καλύτερη περιγραφή του υπερ-τύπου (στο παράδειγμά μας, των φορητών υπολογιστών) και για να γίνει σαφής ο τρόπος με τον οποίον ξεχωρίζουν οι υπο-τύποι του. Άλλωστε μια τέτοια σύγκριση συναντάται συνήθως μέσα σε έναν ορισμό του υπερ-τύπου.

Γενικά, η *σύγκριση ειδοποιού διαφοράς τύπων* πραγματοποιείται συνήθως ανάμεσα σε τύπους οντοτήτων (και όχι μεμονωμένες οντότητες), και μάλιστα ανάμεσα σε υπο-τύπους ενός υπερ-τύπου οντοτήτων, έτσι ώστε να επισημανθούν οι υπο-τύποι του υπερ-τύπου και οι διαφορές τους.

Σε αντίθεση με την ευθεία σύγκριση, στη σύγκριση ειδοποιού διαφοράς τύπων εξετάζονται οι ομοιότητες και οι διαφορές μόνο των πιο σημαντικών χαρακτηριστικών. Δεν μας ενδιαφέρει να περιγράψουμε πλήρως τους υπο-τύπους, αλλά θέλουμε να δώσουμε ένα κριτήριο το οποίο να μπορεί να τους ξεχωρίσει εύκολα. Μία εξήγηση που δίνεται από τη Milosavljevic είναι ότι τέτοιου είδους συγκρίσεις υπάρχουν συνήθως μέσα σε κείμενα που σκοπό έχουν τον ορισμό κυρίως του υπερ-τύπου και όχι των υπο-τύπων του. Η ευθεία σύγκριση δεν θα εξυπηρετούσε το στόχο ενός τέτοιου κειμένου, αφού θα αναλωνόταν στην καταγραφή των διαφορών των υπο-τύπων και δεν θα επικεντρωνόταν στον ίδιο τον υπερ-τύπο. (Αντίθετα, η ευθεία σύγκριση θα ήταν επιθυμητή σε περιγραφές που εστιάζονται στους υπο-τύπους.) Ακόμα, πολλές φορές οι μικροδιαφορές δεν προσφέρουν ιδιαίτερα χρήσιμες πληροφορίες στον αναγνώστη, ειδικά όταν αυτός δεν έχει πολλές γνώσεις επί του θέματος. Αντίθετα, τον μπερδεύουν και τον αποπροσανατολίζουν. Καθώς το πλήθος των πληροφοριών που δέχεται αυξάνεται, δεν μπορεί να τις συγκρατήσει, με αποτέλεσμα να κουράζεται και τελικά να σταματά να προσπαθεί να τις κατανοήσει. Ως προς αυτό, οι συγκρίσεις ειδοποιού διαφοράς τύπων πλεονεκτούν έναντι των ευθέων συγκρίσεων μεταξύ τύπων οντοτήτων, όπου συνήθως αναφέρονται όλα τα (τυπικά) χαρακτηριστικά.

2.1.3 Σύγκριση με συμπλήρωμα συνόλου

> ***Η Java, σε αντίθεση με τις υπόλοιπες γλώσσες προγραμματισμού εκείνης της εποχής, μπορούσε να εκτελεσθεί σε διαφορετικά συστήματα χωρίς να απαιτείται επαναμεταγλώττιση.***

Εδώ παρατηρούμε ότι συγκρίνεται η Java με τις υπόλοιπες γλώσσες προγραμματισμού, δηλαδή με τα υπόλοιπα μέλη του συνόλου των γλωσσών προγραμματισμού, και παρουσιάζεται το ιδιαίτερο χαρακτηριστικό της, αυτό που την κάνει να ξεχωρίζει από τις υπόλοιπες.

> ***Η Java, όπως οι περισσότερες γλώσσες προγραμματισμού, πρέπει να μεταγλωττιστεί για τη δημιουργία των εκτελέσιμων αρχείων της.***

Σε αντίθεση με το προηγούμενο παράδειγμα, εδώ παρουσιάζεται ένα χαρακτηριστικό της Java το οποίο το μοιράζεται με τις υπόλοιπες γλώσσες. Βλέπουμε, λοιπόν, μία σύγκριση ανάμεσα σε μία οντότητα και στα υπόλοιπα μέλη ενός συνόλου στο οποίο ανήκει η οντότητα αυτή. Εναλλακτικά, μπορούμε να θεωρήσουμε ότι γίνεται μια σύγκριση μεταξύ μιας οντότητας-στόχου (Java) και όλων των άλλων οντοτήτων του ιδίου τύπου (γλώσσες προγραμματισμού). Σκοπός της σύγκρισης είναι η καλύτερη περιγραφή της οντότητας-στόχου. Η σύγκριση μας γνωστοποιεί έναν τύπο οντοτήτων στον οποίο ανήκει η οντότητα-στόχος, σημειώνοντας όμως τις ομοιότητες και τις διαφορές της από τις υπόλοιπες οντότητες του τύπου, έτσι ώστε να αποφευχθεί ο κίνδυνος σύγχυσής τους. Ωστόσο θα πρέπει να σημειώσουμε ότι αρκετές φορές η σύγκριση δεν γίνεται με το σύνολο των μελών του τύπου, αλλά με την πλειοψηφία τους («περισσότερες γλώσσες προγραμματισμού»). Επίσης, συχνά, παρατηρείται η σύγκριση να γίνεται μόνο με τους σημαντικότερους αντιπροσώπους του τύπου στον οποίο ανήκει η οντότητα-στόχος, χωρίς να αναφέρεται ο ίδιος ο τύπος:

> ***Η C++, όπως η C# και η Java, υποστηρίζει τη δημιουργία αντικειμένων.***

Εδώ θεωρούμε ότι η C# και η Java αποτελούν τους κυριότερους αντιπροσώπους των αντικειμενοστρεφών γλωσσών προγραμματισμού. Διαφορετικά η σύγκριση θα μπορούσε να ήταν:

> ***Η C++, όπως και κάθε αντικειμενοστρεφής γλώσσα, υποστηρίζει τη δημιουργία αντικειμένων.***

Γενικά στις *συγκρίσεις με συμπλήρωμα συνόλου* ο σκοπός μας είναι διαφορετικός από ό,τι στις συγκρίσεις ειδοποιού διαφοράς. Προσπαθούμε να περιγράψουμε κάποια οντότητα σε σχέση με τις υπόλοιπες του ιδίου τύπου. Αντίθετα, στις συγκρίσεις ειδοποιού διαφοράς εστιαζόμαστε στην περιγραφή ενός υπερ-τύπου και αναφέρουμε (ως μέρος της περιγραφής του υπερ-τύπου) τα χαρακτηριστικά εκείνα που κάνουν τους υπο-τύπους του να ξεχωρίζουν μεταξύ τους.

Σημειώνουμε ότι σε μια παραλλαγή των συγκρίσεων με συμπλήρωμα συνόλου, αντί για οντότητα-στόχο είναι δυνατόν να έχουμε έναν τύπο οντοτήτων ως στόχο. Για παράδειγμα:

> ***Οι αντικειμενοστρεφείς γλώσσες προγραμματισμού, όπως και όλες οι γλώσσες προγραμματισμού, ...***

Στην περίπτωση αυτή έχουμε μια σύγκριση μεταξύ (α) ενός τύπου οντοτήτων (αντικειμενοστρεφείς γλώσσες προγραμματισμού) που είναι υπο-τύπος ενός γενικότερου τύπου (γλώσσες προγραμματισμού) και (β) όλων των άλλων οντοτήτων του γενικότερου τύπου.

2.1.4 Σύγκριση αποσαφήνισης

Συχνά υπάρχουν οντότητες ή τύποι οντοτήτων με πολύ μεγάλη συνάφεια, με αποτέλεσμα να τις/τους μπερδεύουμε μεταξύ τους. Για αυτό αρκετές φορές όταν προσπαθούμε να περιγράψουμε μία τέτοια οντότητα ή τύπο, αναφέρουμε τις διαφορές της/του με τις παρόμοιες οντότητες ή τύπους, ώστε ο αναγνώστης να μην τις/τους συγχέει.

> ***Τα Windows XP είναι η νέα έκδοση του λειτουργικού συστήματος Windows. Σε αντίθεση με τα Windows Millennium, που βασίζονται στον πυρήνα των Windows 9x, τα XP βασίζονται στον πυρήνα των Windows NT.***

Στο παραπάνω παράδειγμα δίνεται μία περιγραφή του λειτουργικού συστήματος Windows XP. Λόγω των πολλών εκδόσεων που προηγήθηκαν, πιθανώς κάποιος να θεωρήσει ότι και αυτά τα Windows βασίζονται στον ίδιο πυρήνα με τα 9x. Για αυτό αναφέρεται ρητώς ότι ο πυρήνας τους είναι αυτός των NT, ώστε να αποφευχθεί κάθε πιθανότητα παρανόησης.

Γενικά οι *συγκρίσεις αποσαφήνισης* γίνονται μέσα στα πλαίσια της περιγραφής μίας οντότητας ή ενός τύπου οντοτήτων. Τα δύο μέρη που συγκρίνονται είναι:

- η οντότητα ή ο τύπος που περιγράφεται και
- κάποια άλλη οντότητα ή τύπος, που έχει μεγάλο βαθμό ομοιότητας με την πρώτη οντότητα ή τύπο.

Σκοπός της σύγκρισης είναι να τονιστούν οι διαφορές μεταξύ των δύο μερών για να απαλειφθεί ο κίνδυνος σύγχυσης. Η δυσκολία στη δημιουργία τέτοιων συγκρίσεων έγκειται στο ότι δεν υπάρχει κάποιο απλό και απολύτως ακριβές κριτήριο το οποίο να ανιχνεύει πότε δύο οντότητες ή τύποι έχουν αρκετά μεγάλη ομοιότητα ώστε να υπάρχει κίνδυνος σύγχυσης. Ένα κριτήριο θα μπορούσε να είναι το αν είναι του ιδίου ή παρεμφερούς τύπου. Όμως κάτι τέτοιο, αν και φαίνεται να ανιχνεύει αρκετές περιπτώσεις πιθανής σύγχυσης (π.χ. πορτοκάλια και μανταρίνια, που είναι και τα δύο εσπεριδοειδή), δεν εντοπίζει όλες τις περιπτώσεις όπου είναι δυνατόν να χρειάζεται μια σύγκριση αποσαφήνισης. Για παράδειγμα, η φάλαινα είναι θηλαστικό και απέχει αρκετά στην οντολογία των έμβιων ζώων από τον «καρχαρία φάλαινα» που είναι ψάρι, ενώ φαίνεται αρκετά πιθανό να γίνει σύγχυση μεταξύ τους.

Η Milosavljevic, τέλος, παρατηρεί ότι είναι απίθανο να χρειαστεί μια σύγκριση αποσαφήνισης μεταξύ μεμονωμένης οντότητας (π.χ. ένας συγκεκριμένος άνθρωπος) και ολόκληρου τύπου οντοτήτων (π.χ. πίθηκος). Στην περίπτωση αυτή, η σύγκριση αποσαφήνισης θα γινόταν μεταξύ δύο τύπων οντοτήτων (άνθρωποι και πίθηκοι).

2.1.5 Αναφορά σε παρόμοια οντότητα/τύπο

> ***Το bandy είναι ένα παιχνίδι πρόγονος του hockey και του hockey σε πάγο. Ο στόχος του παιχνιδιού είναι σχεδόν ίδιος με αυτόν του hockey. Οι παίκτες φορούν παγοπέδιλα και τα μπαστούνια που κρατούν μοιάζουν με αυτά του hockey. Η μπάλα είναι λίγο μικρότερη από τη μπάλα του hockey και λίγο μεγαλύτερη από αυτή του golf. Το παιχνίδι χωρίζεται σε δύο 45-λεπτα ημίχρονα. Οι κανόνες είναι και πάλι παρόμοιοι με του hockey. Κανείς δεν πρέπει να αγγίξει την μπάλα με το χέρι, εκτός από τον τερματοφύλακα, ο οποίος δεν έχει μπαστούνι. Το δε μπαστούνι δεν πρέπει να το σηκώσουν οι παίκτες ψηλότερα από το ύψος του ώμου.***
>
> ***(ελεύθερη μετάφραση από απόσπασμα της εγκυκλοπαίδειας Encarta 95)***

Για να ορίσει το παιχνίδι bandy, ο συγγραφέας του άρθρου χρησιμοποιεί ένα γνωστότερο παιχνίδι, το hockey, ως γνωστικό υπόβαθρο και εστιάζεται στις διαφορές του bandy από το hockey.

Γενικά η *αναφορά σε παρόμοια οντότητα* προσπαθεί να εκμεταλλευτεί την ήδη υπάρχουσα γνώση του αναγνώστη επί κάποιου θέματος. Αναφέρει μια γνωστή οντότητα, με την οποία μοιάζει πολύ η νέα, και στη συνέχεια διασαφηνίζει τις διαφορές της νέας οντότητας με τη γνωστή. Με αυτόν τον τρόπο ο αναγνώστης αφομοιώνει ευκολότερα τις πληροφορίες για τη νέα οντότητα και δεν τη συγχέει με την παλιά. Με αντίστοιχο τρόπο είναι δυνατόν να οριστεί ένας νέος τύπος οντοτήτων με *αναφορά σε παρόμοιο τύπο*: αναφέρεται ένας γνωστός τύπος οντοτήτων, που μοιάζει πολύ με το νέο, και διασαφηνίζονται οι διαφορές του νέου τύπου από τον παλαιό (π.χ. «Οι οθόνες TFT είναι παρόμοιες με τις LCD, αλλά...»).

Η αναφορά σε παρόμοια οντότητα/τύπο μοιάζει πάρα πολύ με τη σύγκριση αποσαφήνισης. Ωστόσο, η διαφορά τους συνίσταται στο τι προσπαθεί να καταφέρει η κάθε μία. Με την αναφορά σε παρόμοια οντότητα/τύπο προσπαθούμε να χρησιμοποιήσουμε τις γνώσεις του αναγνώστη πάνω σε ένα θέμα. Φυσικά για να λειτουργήσει ικανοποιητικά η αναφορά, θα πρέπει ο αναγνώστης να γνωρίζει καλά την οντότητα (αντίστοιχα τύπο) στην οποία αναφερόμαστε. Αντίθετα, στη σύγκριση αποσαφήνισης δεν μας απασχολεί αν ο χρήστης γνωρίζει τη δεύτερη οντότητα (αντίστοιχα τύπο), αλλά προσπαθούμε να τον προστατεύσουμε από το να τις μπερδέψει.

Επίσης αρκετά δύσκολο είναι το πώς θα επιλέξουμε την παρόμοια οντότητα. Ποια θα είναι τα κριτήριά μας ώστε να αποφασίσουμε ότι μία οντότητα είναι κατάλληλη, ώστε να γίνει «υπόβαθρο» για την παρουσίαση μιας άλλης; Αντίστοιχες δυσκολίες υπήρχαν στη σύγκριση αποσαφήνισης.

Σε ένα σύστημα που παράγει αυτόματα περιγραφές οντοτήτων ή τύπων οντοτήτων, η αναφορά σε παρόμοια γνωστή οντότητα ή τύπο διευκολύνεται αν το σύστημα «θυμάται» τις προηγούμενες περιγραφές που έχει παρουσιάσει στο χρήστη, ώστε να μπορεί να αναφερθεί σε οντότητες/τύπους που περιέχονταν σε εκείνες τις περιγραφές, ελπίζοντας φυσικά ότι ο αναγνώστης συγκράτησε τις πληροφορίες των προηγούμενων περιγραφών. Στερεότυπα χρηστών μπορούν επίσης να χρησιμοποιηθούν, προκειμένου να κάνει το σύστημα εύλογες υποθέσεις σχετικά με τις οντότητες/τύπους που γνωρίζει ήδη ένας χρήστης (π.χ. ότι το baseball είναι γνωστό σε Αμερικανούς χρήστες).

2.1.6 Αναλογία

> ***Το ερπετό, ένας Πλησιόσαυρος, ήταν «μεγάλο σαν λεωφορείο, με δόντια μεγαλύτερα από αγγούρια [...] σε ένα κεφάλι που θα μπορούσε να καταπιεί ολόκληρο άνθρωπο» αναφέρει ανακοίνωση του Μουσείου Φυσικής Ιστορίας του πανεπιστημίου. Το συνολικό του μήκος ίσως υπερβαίνει τα δέκα μέτρα.***
>
> ***Βασισμένο σε απόσπασμα άρθρου του in.gr.***[4]

Για να γίνει αισθητό το μέγεθός του, ο Πλησιόσαυρος (είδος δεινοσαύρου) συγκρίνεται με το μέγεθος ενός λεωφορείου, ενώ τα δόντια του με το μέγεθος ενός αγγουριού. Επίσης αναφέρεται ότι το στόμα του ήταν μεγαλύτερο από έναν άνθρωπο. Ο συγγραφέας προσπαθεί να δώσει, δηλαδή, ένα μέτρο σύγκρισης, ώστε ο αναγνώστης να αντιληφθεί καλύτερα τις διαστάσεις του τεράστιου δεινοσαύρου.

Οι *συγκρίσεις αναλογίας* χρησιμοποιούν κάποιο χαρακτηριστικό μίας γνωστής οντότητας προκειμένου να δώσουν ένα μέτρο σύγκρισης για το αντίστοιχο χαρακτηριστικό της οντότητας που θέλουμε να περιγράψουμε. Με αυτόν τον τρόπο ο αναγνώστης μπορεί να αντιληφθεί ευκολότερα το χαρακτηριστικό της νέας οντότητας (χρώμα, μέγεθος, χρησιμότητα κ.ά.). Η γνωστή οντότητα δεν χρειάζεται να έχει κάποια ιδιαίτερη σχέση με τη νέα, αρκεί να τη γνωρίζει ο αναγνώστης και να μοιράζεται κάποιο χαρακτηριστικό (π.χ. ύψος) με την οντότητα που θέλουμε να περιγράψουμε (το

[4] http://www.in.gr/news/article.asp?lngEntityID=744278&lngDtrID=252

λεωφορείο και ο δεινόσαυρος δεν έχουν καμία ιδιαίτερη σχέση μεταξύ τους). Συγκρίσεις αναλογίας μπορούν να γίνουν επίσης μεταξύ τύπων οντοτήτων.

Μία άλλη σύγκριση αναλογίας είναι η ακόλουθη:

> ***Το άγαλμα που βλέπετε έχει το ύψος ενός ενήλικα άνδρα.***

Ένα πρόβλημα που δημιουργείται σε αυτή την περίπτωση είναι το τι θεωρούμε ως ύψος ενός ενήλικα άνδρα. Σε χώρες όπου το μέσο ύψος των κατοίκων είναι μικρό, μία τέτοια σύγκριση θα μπορούσε να εκληφθεί ότι σημαίνει πως το ύψος του αγάλματος είναι μικρότερο από ό,τι σε άλλες χώρες (π.χ. 1,65 αντί 1,80 μέτρα). Γενικά η χρήση συγκρίσεων αναλογίας πρέπει να χρησιμοποιείται με αρκετή προσοχή, ώστε να μη δημιουργούνται τέτοιου είδους αμφισημίες.

Και πάλι εδώ μπορούμε να χρησιμοποιήσουμε τη γνώση του συστήματος για το τι έχει περιγράψει μέχρι εκείνη τη στιγμή. Ωστόσο, δεν αρκεί ο αναγνώστης να έχει ακούσει κάποιες γενικές πληροφορίες για την οντότητα (αντίστοιχα τύπο) που χρησιμοποιούμε ως μέτρο σύγκρισης. Θα πρέπει να τη γνωρίζει σε βάθος, να νιώθει ιδιαίτερα εξοικειωμένος μαζί της. Για παράδειγμα, αν κάποιος έχει επισκεφθεί ένα μουσείο και έχει δει μόνο μία λήκυθο, τύπο εκθεμάτων που έστω ότι δεν γνώριζε προηγουμένως, δεν πρέπει η λήκυθος (ως τύπος οντοτήτων) να χρησιμοποιηθεί ως μέτρο σύγκρισης, αφού ο επισκέπτης δεν γνωρίζει αρκετά αυτόν τον τύπο. Ωστόσο αν έχει περάσει από ολόκληρη πτέρυγα με ληκύθους και στη συνέχεια συναντήσει κάποιο έκθεμα άλλου είδους, ίσως μια σύγκριση αναλογίας με τις ληκύθους να είναι επιτυχημένη.

2.2 Αλγόριθμος Milosavljevic

Στη συνέχεια παρουσιάζεται ο γενικός αλγόριθμος που προτείνεται από την Milosavljevic (1999) σε μορφή ψευδοκώδικα. Ο αλγόριθμος θεωρεί ότι έχει ζητηθεί από το σύστημα να παραγάγει μια περιγραφή μιας συγκεκριμένης οντότητας ή τύπου οντοτήτων και ότι η σύγκριση που θα παραχθεί θα ενσωματωθεί στην περιγραφή αυτή. Στον ψευδοκώδικα παρακάτω έχουμε κάνει ορισμένες τροποποιήσεις, σε σχέση με τον ψευδοκώδικα της Milosavljevic, ώστε να ακολουθείται η ορολογία των προηγουμένων ενοτήτων. Επίσης, έχουμε αφαιρέσει ένα τμήμα που αναφερόταν σε γενικευμένες οντότητες (generic entities), επειδή δεν υποστηρίζονται από τις οντολογίες OWL.

```
Χρησιμοποίησε ένα μέτρο σύγκρισης και βρες μία δεύτερη οντότητα (αν
περιγράφουμε οντότητα) ή ένα δεύτερο τύπο οντοτήτων (αν περιγράφουμε τύπο
οντοτήτων) που μπορεί να συγχυστεί με την οντότητα ή τύπο οντοτήτων που
περιγράφουμε.

Εάν η ομοιότητα είναι πολύ υψηλή (μεγαλύτερη από ένα κατώφλι) τότε
```

κάνε ευθεία σύγκριση των δύο οντοτήτων ή τύπων.

Αλλιώς

Εάν ο αναγνώστης γνωρίζει τη δεύτερη οντότητα/τύπο **τότε**

χρησιμοποίησε αναφορά σε παρόμοια οντότητα/τύπο.

Αλλιώς

Εάν είναι πιθανόν να συναντήσει τη δεύτερη οντότητα/τύπο αργότερα και να την μπερδέψει με την περιγραφόμενη οντότητα/τύπο **τότε**

χρησιμοποίησε σύγκριση αποσαφήνισης.

Τέλος αν

Τέλος αν

Τέλος αν

Εάν περιγράφουμε τύπο οντοτήτων που έχει κάποιο σημαντικό χαρακτηριστικό που τον κάνει να ξεχωρίζει **τότε**

Εάν περιγράφουμε τύπο που έχει υπο-τύπους **τότε**

χρησιμοποίησε μία ή το πολύ δύο συγκρίσεις ειδοποιού διαφοράς μεταξύ αυτών των υπο-τύπων και ως προς αυτό το χαρακτηριστικό.

Τέλος αν

Τέλος αν

Εάν περιγράφουμε οντότητα που ανήκει σε ένα γενικότερο τύπο και η περιγραφόμενη οντότητα διαφέρει ως προς τα χαρακτηριστικά της από τις υπόλοιπες οντότητες που ανήκουν στο γενικότερο τύπο **τότε**

να αναφερθούν (με σύγκριση συμπληρώματος συνόλου) μία έως τρεις τέτοιες διαφορές που αποδεικνύουν την μοναδικότητα της περιγραφόμενης οντότητας σε σχέση με τις υπόλοιπες οντότητες του γενικότερου τύπου.

Τέλος αν

Εάν περιγράφουμε τύπο οντοτήτων που είναι υπο-τύπος ενός γενικότερου τύπου και ο περιγραφόμενος τύπος διαφέρει από όλους τους άλλους (άμεσους) υπο-τύπους του γενικότερου τύπου **τότε**

να αναφερθούν (με σύγκριση συμπληρώματος συνόλου) μία έως τρεις τέτοιες διαφορές που αποδεικνύουν την μοναδικότητα του περιγραφόμενου τύπου σε σχέση με τους υπόλοιπους υπο-τύπους του γενικότερου τύπου.

Τέλος αν

Για κάθε ιδιότητα Α της οντότητας που περιγράφουμε, η οποία δεν χρησιμοποιείται σε κάποια από τις προηγούμενες συγκρίσεις και είναι δύσκολο να περιγραφεί (άρα και να γίνει αντιληπτή από τον αναγνώστη) **επανάλαβε**

Εάν υπάρχει κάποια άλλη οντότητα που την γνωρίζει ο αναγνώστης και που μοιράζεται αυτή την ιδιότητα και μάλιστα έχουν παρόμοια τιμή **τότε**

πρέπει να χρησιμοποιηθεί σύγκριση αναλογίας με την άλλη οντότητα για την ιδιότητα Α.

Τέλος αν

Εάν η ιδιότητα είναι μετρήσιμη (ή γενικότερα συγκρίσιμη με κάποιο τρόπο) και υπάρχει κάποια άλλη οντότητα που την γνωρίζει ο αναγνώστης και που μοιράζεται αυτή την ιδιότητα **τότε**

μπορεί να χρησιμοποιηθεί το χαρακτηριστικό (η τιμή της ιδιότητας) της δεύτερης οντότητας ως μονάδα μέτρησης/σύγκρισης για να περιγραφεί καλύτερα η ιδιότητα της περιγραφόμενης οντότητας με σύγκριση αναλογίας.

Τέλος αν

Τέλος επανάληψης

Σημειώνουμε ότι στις οντολογίες OWL ο τελευταίος βρόχος μπορεί να εφαρμοστεί και στην περίπτωση που περιγράφουμε έναν τύπο οντοτήτων. Στην περίπτωση αυτή, εφαρμόζεται σε περιορισμούς τιμών των ιδιοτήτων του τύπου.

Στη συνέχεια θα αναφέρουμε τους αλγορίθμους που προτείνει η Milosavljevic για κάθε μία από τις κατηγορίες των συγκρίσεων.

2.2.1 Σύγκριση αποσαφήνισης και αναφορά σε παρόμοια οντότητα/τύπο

Όπως έχουμε ήδη αναφέρει, αυτού του τύπου οι συγκρίσεις μοιάζουν πάρα πολύ στη μορφή τους. Η βασική τους διαφορά είναι ο σκοπός που προσπαθούν να επιτύχουν. Για αυτό ο αλγόριθμός τους είναι κοινός. Επίσης, καθώς τα χαρακτηριστικά στα οποία μπορούν να μοιάζουν δύο οντότητες/τύποι είναι πάρα πολλά και αρκετά από αυτά δεν παρουσιάζουν ιδιαίτερο ενδιαφέρον, η Milosavljevic προτείνει μία λίστα από κατηγορίες χαρακτηριστικών που έχει νόημα να αντιπαραταχθούν (σχήμα, εμφάνιση, θέση στην ταξινομία κ.ά.)

Για κάθε κοινό χαρακτηριστικό των δύο οντοτήτων/τύπων που υπάρχει στη λίστα των χαρακτηριστικών που έχει νόημα να αντιπαραταχθούν **επανάλαβε**

Υπολόγισε το βαθμό ομοιότητας του χαρακτηριστικού για τις δύο οντότητες/τύπους με βάση κάποιο κριτήριο ομοιότητας.

Τέλος επανάληψης

Επίλεξε εκείνα τα χαρακτηριστικά από τα παραπάνω που μοιάζουν περισσότερο ανάμεσα στις δύο οντότητες/τύπους.

Για κάθε χαρακτηριστικό Α που βρήκες στο προηγούμενο βήμα **επανάλαβε**
    Εάν υπάρχει κάποιο άλλο χαρακτηριστικό Β το οποίο συσχετίζεται με το χαρακτηριστικό Α (με βάση κάποιο κριτήριο)και οι δύο οντότητες/τύποι που συγκρίνουμε διαφέρουν ως προς το χαρακτηριστικό Β **τότε**
        Το χαρακτηριστικό Α δεν μπορεί να χρησιμοποιηθεί ως μέλος κάποιας σύγκρισης.
    Τέλος αν
Τέλος επανάληψης

Επίλεξε το πολύ τρία χαρακτηριστικά από αυτά που έχουν απομείνει. Δημιούργησε μια σύγκριση (αποσαφήνισης ή αναφοράς σε παρόμοια οντότητα/τύπο, ανάλογα με το στόχο της κλήσης του αλγορίθμου αυτού) ως προς τα χαρακτηριστικά που επέλεξες (η σειρά επιλογής των χαρακτηριστικών και παρουσίασης τους εξαρτάται από τη σειρά προτεραιότητας των χαρακτηριστικών στη λίστα).

Εάν κάποιο χαρακτηριστικό της λίστας που περιέχει τα χαρακτηριστικά εκείνα που έχει νόημα να αντιπαραταχθούν υπάρχει μόνο στη δεύτερη οντότητα/τύπο, αλλά όχι στην οντότητα/τύπο που περιγράφουμε **τότε**
    Αυτό το χαρακτηριστικό πιθανόν να προκαλέσει σύγχυση στον αναγνώστη. Δημιούργησε σύγκριση (αποσαφήνισης ή αναφοράς σε παρόμοια οντότητα/τύπο, ανάλογα με το στόχο) για αυτό το χαρακτηριστικό.
Τέλος αν

2.2.2 Ευθεία σύγκριση

Για να επιτευχθεί η ευθεία σύγκριση αρκεί να αντιπαραβάλουμε ένα–ένα τα χαρακτηριστικά των δύο οντοτήτων ή τύπων. Ωστόσο θα ήταν προτιμότερο αρχικά να παρουσιάζονται οι σημαντικότερες ομοιότητες και διαφορές των οντοτήτων ή τύπων και στη συνέχεια οι επουσιώδεις. Για να το πετύχουμε αυτό, μπορούμε να χρησιμοποιήσουμε, όπως και στον προηγούμενο αλγόριθμο, μια λίστα με τα σημαντικότερα χαρακτηριστικά και η ευθεία σύγκριση να εστιάζεται αρχικά στα χαρακτηριστικά της λίστας, πριν προχωρήσει στα υπόλοιπα.

2.2.3 Σύγκριση ειδοποιού διαφοράς

Σε αυτόν τον τύπο σύγκρισης δεν μας ενδιαφέρουν οι ομοιότητες, παρά μόνο οι διαφορές των υπο-τύπων του τύπου οντοτήτων που περιγράφουμε. Θέλουμε να βρούμε εκείνο το χαρακτηριστικό που διαχωρίζει τους υπο-τύπους. Επιπλέον, θέλουμε αυτό το χαρακτηριστικό να είναι ιδιαίτερης σημασίας (αφού αν δύο υπο-τύποι διαφέρουν σε κάποια μικρή λεπτομέρεια, ίσως αυτό να μην είναι αρκετό για να τους ξεχωρίσουμε) και η τιμή που παίρνει στους υπο-τύπους να διαφέρει σημαντικά (αν η τιμή του χαρακτηριστικού είναι παρόμοια, τότε η διαφορά δεν μπορεί να θεωρηθεί ως ειδοποιός).

Έτσι ο αλγόριθμος που προτείνεται είναι:

Αρχικά φτιάχνουμε ταξινομημένη λίστα με τις σημαντικότερες ιδιότητες του υπερ-τύπου που περιγράφουμε (κατά φθίνουσα σειρά σημαντικότητας) για το συγκεκριμένο χρήστη.

Με βάση κάποιο μέτρο σύγκρισης, επιλέγουμε μία ιδιότητα του υπερ-τύπου οντοτήτων που περιγράφουμε έτσι ώστε να ανήκει στην παραπάνω λίστα (και κατά προτίμηση στην αρχή της) και να έχει σημαντικά διαφορετικές τιμές σε κάποιους από τους υπο-τύπους που θέλουμε να διαχωρίσουμε.

Επιλέγουμε μόνο υπο-τύπους που αναδεικνύουν την διαφορά (διαφέρουν, δηλαδή, σημαντικά ως προς την τιμή της επιλεγείσας ιδιότητας) και πληροφορούμε τον αναγνώστη για τους υπο-τύπους αυτούς και τις διαφορετικές τιμές που έχουν στην επιλεγείσα ιδιότητα.

2.2.4 Σύγκριση με συμπλήρωμα συνόλου

Εδώ προσπαθούμε να βρούμε (α) τις διαφορές της οντότητας-στόχου από τις υπόλοιπες οντότητες του ιδίου τύπου ή (β) τις διαφορές του τύπου-στόχου από τους τύπους που είναι αδέλφια του. Ένας τρόπος που προτείνει η Milosavljevic για να το επιτύχουμε αυτό είναι να κατασκευάσουμε μία ή περισσότερες βοηθητικές ψευδο-οντότητες ή ψευδο-τύπους που θα αποτελούν «αντιπροσώπους» των οντοτήτων ή τύπων με τους οποίους θέλουμε να συγκρίνουμε. Η σύγκριση μπορεί να γίνει κατόπιν με τους αντιπροσώπους, που θα είναι λιγότεροι. Οι αντιπρόσωποι μπορούν να δημιουργηθούν είτε κατασκευάζοντας ψευδο-οντότητες/τύπους που θα έχουν ως χαρακτηριστικά τούς μέσους όρους των χαρακτηριστικών των οντοτήτων/τύπων με τους οποίους θέλουμε να συγκρίνουμε, είτε χρησιμοποιώντας μεθόδους συσταδοποίησης (clustering). Και στις δύο περιπτώσεις χρειαζόμαστε να ξέρουμε το εύρος τιμών που μπορεί να πάρει κάθε χαρακτηριστικό. Κατά τη δημιουργία της συγκρίσεως θα πρέπει, επίσης, να εστιαστούμε στα χαρακτηριστικά του στόχου που δεν διαφέρουν απλά από εκείνα των αντιπροσώπων, αλλά διαφέρουν αισθητά. Τέλος, θα πρέπει να δίνουμε ιδιαίτερη βαρύτητα στα μοναδικά χαρακτηριστικά του στόχου, δηλαδή τα χαρακτηριστικά εκείνα τα οποία τα διαθέτει μόνο η οντότητα/υπο-τύπος που περιγράφουμε (ή έστω τα χαρακτηριστικά αυτά να τα συναντάμε σε λίγες οντότητες/υπο-τύπους του συνόλου με το οποίο συγκρίνουμε) και την κάνουν να ξεχωρίζει.

Σημειώνουμε ότι το συγκεκριμένο είδος σύγκρισης δεν υλοποιήθηκε από τη Milosavljevic.

2.2.5 Αναλογία

Στην αναλογία συγκρίνουμε μία οντότητα με μία άλλη (η οποία μας είναι ήδη γνωστή και χρησιμοποιείται ως μέτρο σύγκρισης) ως προς μία ιδιότητα. Η (γνωστή) οντότητα με την οποία θα

συγκρίνουμε και η ιδιότητα στην οποία θα βασιστεί η σύγκριση μάς δίνονται από το γενικό αλγόριθμο που περιγράψαμε παραπάνω.

2.3 Μέτρα σύγκρισης

Όπως έχουμε ήδη αναφέρει, είναι αναγκαίο να υπάρχει ένα μέτρο σύγκρισης που να μπορεί να εντοπίσει οντότητες (ή τύπους οντοτήτων) που είναι ίδιες, μοιάζουν ή διαφέρουν σημαντικά σε κάποια χαρακτηριστικά.

Η Milosavljevic χρησιμοποιεί τον ακόλουθο μαθηματικό τύπο για να συγκρίνει δύο χαρακτηριστικά X:A και X:B, όπου X μία ιδιότητα και A, B οι τιμές της σε δύο οντότητες ή τύπους οντοτήτων.

$$Ομοιότητα(X\colon A,\ X\colon B) = \frac{|A-B|}{max_{\ p,q\ in\ range\ (X)}|p-q|}$$

Έτσι μπορούμε να υπολογίσουμε την κανονικοποιημένη ομοιότητα ανάμεσα σε δύο διαφορετικές τιμές (A και B) του χαρακτηριστικού X. Στον αριθμητή έχουμε την απόλυτη τιμή της διαφοράς τους, ενώ στον παρονομαστή τη μεγαλύτερη δυνατή διαφορά που μπορεί να υπάρξει για αυτό το χαρακτηριστικό.

Έχοντας λοιπόν ένα μέτρο σύγκρισης, μπορούμε να ορίσουμε τάξεις ομοιότητας. Αν ο αριθμός που μας επιστρέφει ο τύπος είναι κοντά στην μονάδα, τότε αυτό σημαίνει ότι οι δύο οντότητες διαφέρουν σημαντικά. Αν πάλι ο τύπος μάς επιστρέφει τιμή κοντά στο μηδέν, αυτό σημαίνει ότι τα χαρακτηριστικά είναι σχεδόν όμοια. Έτσι μπορούμε να φτιάξουμε επίπεδα ομοιότητας (όμοια, παρόμοια, μικρές διαφορές, μεγάλες διαφορές) χρησιμοποιώντας κάποια εύρη επιστρεφόμενων τιμών. Το ίδιο μπορεί να γίνει συχνά και με τα ποιοτικά χαρακτηριστικά, μετατρέποντάς τα πρώτα σε αριθμητικές τιμές.

2.4 Ο αλγόριθμος του Μελέγκογλου

Στη συνέχεια θα περιγράψουμε τη μέθοδο που αναπτύχθηκε από το Μελέγκογλου στη διάρκεια του έργου IST M-PIRO (Isard κ.ά. 2003, Androutsopoulos κ.ά. 2007).[5] Σκοπός της είναι να βρει ποιες από τις προηγούμενες οντότητες, που έχει ήδη περιγράψει το σύστημα, είναι οι καταλληλότερες, ώστε να συγκριθούν με τη νέα οντότητα που εξετάζει αυτή τη στιγμή ο χρήστης, και βάσει ποιας ιδιότητας. Μπορεί να παράγει τόσο συγκρίσεις ομοιότητας (π.χ. «Όπως οι κάνθαροι που είδατε, έτσι και αυτή η λήκυθος κατασκευάστηκε στη διάρκεια της κλασικής

[5] Συμπεριλαμβάνουμε στην περιγραφή βελτιώσεις που έγιναν στη μέθοδο από τον κ. Μελέγκογλου με την ευκαιρία της παρούσας εργασίας.

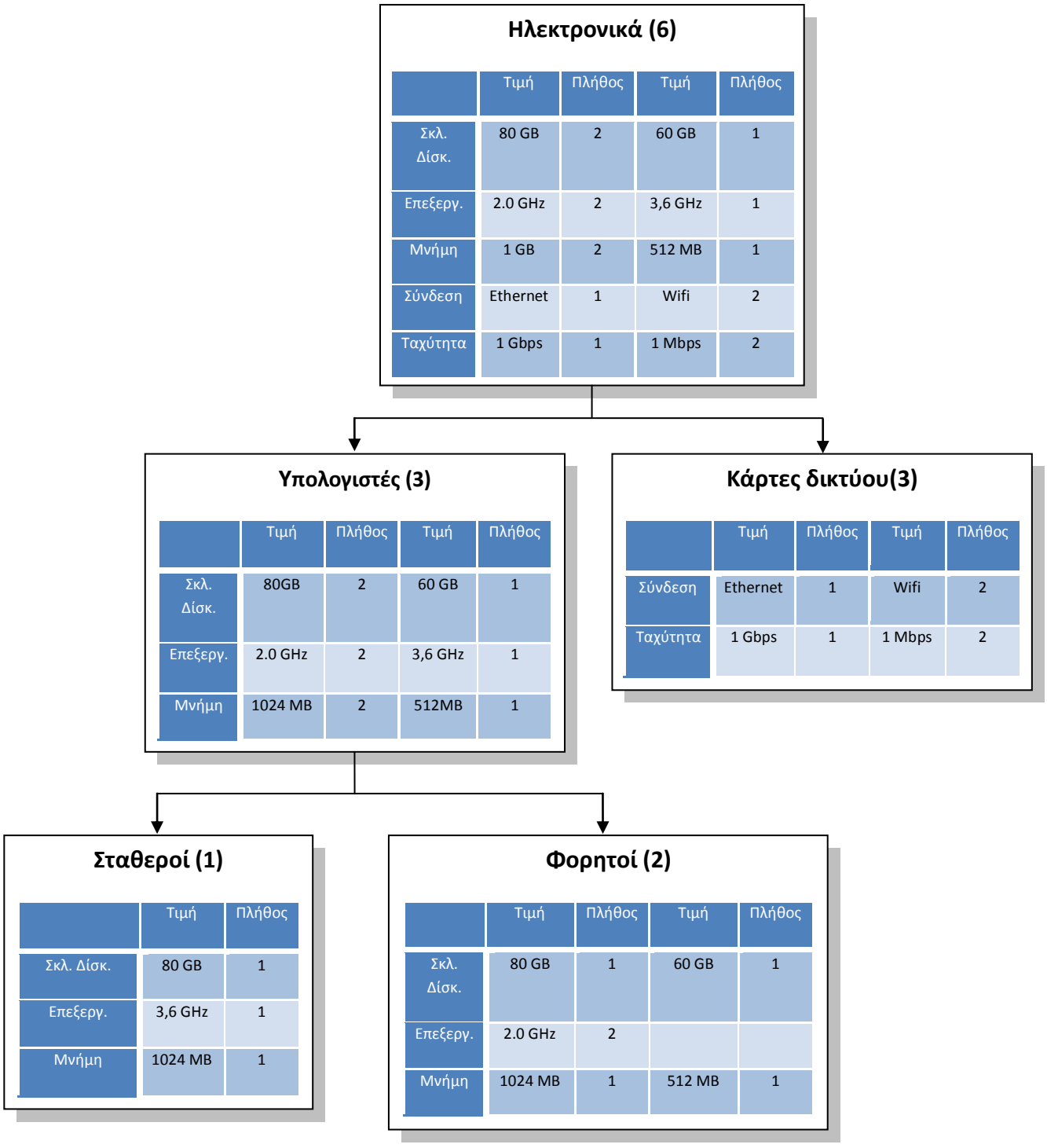
περιόδου.») όσο και διαφοράς (π.χ. «Αντίθετα από τις ληκύθους που είδατε, οι οποίες ακολουθούσαν τον ερυθρόμορφο ρυθμό, αυτός ο αμφορέας...»).

Η μέθοδος έχει ως βάση την παρατήρηση ότι οι αλγόριθμοι που επιλέγουν τις προηγούμενες οντότητες, με τις οποίες θα γίνουν οι συγκρίσεις, χρησιμοποιώντας αποκλειστικά μετρήσεις του βαθμού ομοιότητας μεταξύ των οντοτήτων, συχνά επιβαρύνουν τις περιγραφές με δυσνόητες και περίπλοκες αναφορικές φράσεις (π.χ. «Αντίθετα από την ερυθρόμορφη λήκυθο που βρέθηκε στην Αθήνα, ...»). Οι εκφράσεις αυτές είναι απαραίτητες για να καταστήσουν σαφές με ποια (ή ποιες) προηγούμενη οντότητα γίνεται κάθε φορά η σύγκριση. Σύμφωνα, λοιπόν, με το Μελέγκογλου, η ομοιότητα ανάμεσα στη νέα και στις ήδη γνωστές οντότητες δεν πρέπει να αποτελεί το μοναδικό κριτήριο επιλογής. Επιπρόσθετα, πρέπει να δίνεται προτεραιότητα στην αναζήτηση γνωστών οντοτήτων που είτε ατομικά είτε ομαδικά δύνανται να συγκριθούν με την οντότητα της νέας περιγραφής με απλή αναφορά στο όνομα του τύπου τους (π.χ. «Σε αντίθεση με τις ληκύθους που είδατε...»). Για να το πετύχει αυτό, η μέθοδος χρησιμοποιεί μια ταξινομική αναπαράσταση των περασμένων οντοτήτων, η οποία περιγράφεται στη συνέχεια.

2.4.1 Αναπαράσταση οντοτήτων που έχουν ήδη περιγραφεί

Για την παραγωγή συγκρίσεων, ο αλγόριθμος δημιουργεί μια δυναμική ταξινομία η οποία ανά πάσα στιγμή περιλαμβάνει τους τύπους και τους υπερ-τύπους των οντοτήτων που έχουν ήδη περιγραφεί στο χρήστη. Σε κάθε τύπο καταγράφεται ο αριθμός των οντοτήτων του που έχουν περιγραφεί, οι ιδιότητες του τύπου και οι διαφορετικές τιμές που έχει συγκεντρώσει η κάθε ιδιότητα στις προηγούμενες οντότητες. Επίσης καταγράφεται και ο αριθμός των οντοτήτων που παρουσιάζουν την κάθε τιμή.

Στο παρακάτω παράδειγμα περιγράφονται δύο φορητοί υπολογιστές, ένας σταθερός, καθώς και τρεις κάρτες δικτύου. Οι ιδιότητες που εξετάζουμε για τους υπολογιστές (σταθερούς και φορητούς) είναι η χωρητικότητα των σκληρών δίσκων τους, η ταχύτητα του επεξεργαστή τους και το μέγεθος της μνήμης τους. Για τις κάρτες δικτύου ελέγχουμε ποιο πρωτόκολλο χρησιμοποιούν, καθώς και την ταχύτητα που επιτυγχάνουν. Ο κόμβος με τίτλο «ηλεκτρονικά» χρησιμοποιείται ως κοινός υπερ-τύπος των υπολογιστών και των καρτών δικτύων.



2.4.2 Βασικές ιδέες του αλγορίθμου

Η αναζήτηση για την καταλληλότερη προηγούμενη οντότητα ή τύπο (ουσιαστικά ομάδα προηγούμενων οντοτήτων) με την οποία θα γίνει η σύγκριση ξεκινά από τον ειδικότερο κόμβο της δυναμικής ταξινομίας στον οποίο ανήκει και η οντότητα που περιγράφουμε. Στο παραπάνω παράδειγμα, αν θέλαμε να περιγράψουμε ένα σταθερό υπολογιστή, θα ξεκινάγαμε από τον κόμβο «Σταθεροί». Όλοι οι κόμβοι κάτω από αυτόν τον κόμβο (μαζί με αυτόν) αποτελούν το αρχικό πεδίο αναζήτησης, στο οποίο αναζητούμε την καταλληλότερη οντότητα ή τύπο με την οποία θα γίνει η

σύγκριση. Το αρχικό πεδίο αναζήτησης, δηλαδή, είναι ένα υπο-δέντρο της δυναμικής ταξινομίας, με ρίζα τον ειδικότερο κόμβο στον οποίο ανήκει η οντότητα που περιγράφουμε.[6] Κατά την εκτέλεση του αλγορίθμου, το πεδίο αναζήτησης σταδιακά επεκτείνεται, χρησιμοποιώντας σε κάθε επέκταση ως ρίζα του πεδίου αναζήτησης τον πατέρα της προηγούμενης ρίζας, μέχρι να φτάσουμε σε κάποια κόμβο που δεν επιτρέπεται να γενικευθεί περαιτέρω, δηλαδή του οποίου δεν επιτρέπεται να χρησιμοποιηθεί ο πατέρας ως νέα ρίζα. Οι κόμβοι που δεν επιτρέπεται να γενικευθούν ορίζονται κατά τη διαμόρφωση (configuration) του αλγορίθμου για μια νέα εφαρμογή, ώστε να αποφεύγονται συγκρίσεις με υπερβολικά γενικούς τύπους.

Ο αλγόριθμος έχει στηριχθεί στα παρακάτω πέντε κριτήρια. Τα τρία πρώτα είναι θεμελιώδη. Από αυτά προκύπτουν οι οντότητες ή τύποι που είναι κατάλληλοι για σύγκριση, μαζί με τα χαρακτηριστικά για τα οποία μπορούν να συγκριθούν με την τρέχουσα οντότητα. Τα επόμενα δύο λειτουργούν ως φίλτρα, δηλαδή προσπαθούν να επιλέξουν την καλύτερη σύγκριση μεταξύ αυτών που επιτρέπουν τα θεμελιώδη κριτήρια. Αυτά τα δύο κριτήρια-φίλτρα δεν είναι απαραίτητα, με την έννοια ότι οι συγκρίσεις που επιτρέπουν τα θεμελιώδη κριτήρια είναι γενικά αποδεκτές, αλλά ίσως όχι όλες στον ίδιο βαθμό. Τα κριτήρια-φίλτρα μπορούν να αντικατασταθούν από άλλα ή να συμπληρωθούν με άλλα, ανάλογα με την εφαρμογή. Σημειώνουμε ότι ουσιαστικά αναζητούμε πάντα έναν τύπο οντοτήτων με τον οποίο θα γίνει η σύγκριση. Απλά αν επιλεγεί ένας τύπος του οποίου έχουμε συναντήσει μόνο μία οντότητα, η σύγκριση γίνεται με την οντότητα εκείνη.

1. **Η τιμή της ιδιότητας θα πρέπει να είναι κοινή για όλες τις οντότητες του τύπου:** Για να παραχθεί μία σύγκριση με κάποιον τύπο για μια ιδιότητά του, θα πρέπει η τιμή αυτής της ιδιότητας να είναι η ίδια σε όλες τις οντότητες του τύπου που έχουμε συναντήσει. Διαφορετικά δεν θα μπορούσαμε να χρησιμοποιήσουμε το όνομα του τύπου για να αναφερθούμε στις οντότητές του.

   Έστω ότι ο αναγνώστης έχει συναντήσει μόνο δύο φορητούς υπολογιστές, και θέλουμε να τους συγκρίνουμε με ένα σταθερό υπολογιστή που εξετάζει τώρα. Αν οι δύο φορητοί υπολογιστές έχουν επεξεργαστή Χ και ο σταθερός έχει τον ίδιο επεξεργαστή, τότε μία σύγκριση που θα μπορούσε να παραχθεί είναι:

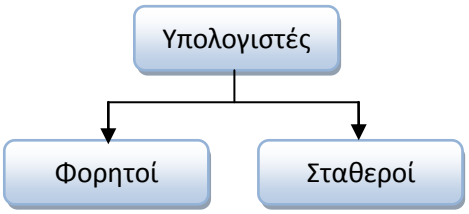
   Όπως όλοι οι φορητοί υπολογιστές που είδατε, αυτός ο σταθερός υπολογιστής έχει τον επεξεργαστή Χ.

   Αντίθετα, αν ένας από τους φορητούς είχε άλλον επεξεργαστή, τότε η σύγκριση δεν θα μπορούσε να γίνει αναφέροντας μόνο τον τύπο «φορητοί υπολογιστές». Θα χρειαζόταν

[6] Χάριν απλότητας, θεωρούμε ότι η ταξινομία έχει τη μορφή δέντρου, οπότε δεν επιτρέπεται πολλαπλή κληρονομικότητα.

μια πιο περίπλοκη αναφορική έκφραση (π.χ. «Όπως εκείνος ο φορητός υπολογιστής που είδατε που είχε κατασκευαστεί στη Γερμανία...»), κάτι που προσπαθούμε να αποφύγουμε.

Το πεδίο αναζήτησης σε αυτό το παράδειγμα είναι αρχικά οι σταθεροί υπολογιστές. Στη συνέχεια επεκτείνεται στο υποδέντρο με ρίζα τους υπολογιστές, το οποίο συμπεριλαμβάνει τους φορητούς υπολογιστές. Θεωρούμε ότι ο κόμβος των υπολογιστών δεν επιτρέπεται να γενικευθεί περαιτέρω.



2. **Η μοναδικότητα του χαρακτηριστικού:** Ο αλγόριθμος εξετάζει όλους τους τύπους του πεδίου αναζήτησης, προσπαθώντας να βρει έναν τύπο, έστω Α, που να διαθέτει κάποιο χαρακτηριστικό (ζεύγος ιδιότητας-τιμής) το οποίο να μην εμφανίζεται και σε άλλον τύπο, έστω Β, του πεδίου αναζήτησης που δεν είναι υπο-τύπος του Α. Αν η σύγκριση γινόταν με τον τύπο Α και το χαρακτηριστικό του Α στο οποίο βασίζεται η σύγκριση υπήρχε και στον Β, τότε ο αναγνώστης θα αναρωτιόταν γιατί συγκρίναμε με τον Α και όχι με τον Β. Η αναγκαιότητα αυτού του κριτηρίου θα γίνει καθαρότερη στο παράδειγμα χρήσης του αλγορίθμου που ακολουθεί σε επόμενες σελίδες.
3. **Συγκρίσεις αντίθεσης έχουν νόημα όταν αντικρούουν μια κατάσταση που τείνει να παγιωθεί:** Ενδέχεται οι προηγούμενες οντότητες που περιγράφηκαν να μοιράζονται κάποιο χαρακτηριστικό (ζεύγος ιδιότητας-τιμής) και να παγιώνουν στη σκέψη του αναγνώστη μία κατάσταση. (Για παράδειγμα, μπορεί όλοι οι υπολογιστές που περιγράψαμε να έχουν τον ίδιο επεξεργαστή Χ και ο αναγνώστης να αρχίζει να υποθέτει ότι όλοι οι υπολογιστές, τουλάχιστον του συγκεκριμένου καταστήματος, έχουν τον ίδιο επεξεργαστή Χ.) Έτσι, όταν περιγράφεται μία οντότητα του ιδίου τύπου με διαφορετική τιμή στη συγκεκριμένη ιδιότητα, έχει νόημα να τονιστεί η μοναδικότητά της ως προς τις προηγούμενες (π.χ. «Σε αντίθεση με όλους τους προηγούμενους υπολογιστές, που είχαν τον επεξεργαστή Χ, αυτός ο υπολογιστής έχει τον επεξεργαστή Υ.»). Διαφορετικά αν δεν παραβιάζεται μια κατάσταση που τείνει να παγιωθεί, οι συγκρίσεις αντίθεσης δείχνουν συχνά τυχαίες, οπότε και κάνουν τον αναγνώστη να αναρωτιέται τι σκοπό εξυπηρετούν.
4. **Επιλογή υπο-τύπου που περιγράφει το χαρακτηριστικό στο χαμηλότερο δυνατό επίπεδο της ταξινομίας:** Στις περιπτώσεις που το επιλεγέν χαρακτηριστικό της

σύγκρισης εμφανίζεται τόσο σε έναν τύπο Α όσο και σε υπο-τύπους του (αποκλείεται να εμφανίζεται σε τύπους Β που δεν είναι υπο-τύποι του Α, λόγω του δεύτερου κριτηρίου), τότε προτιμάται ο υπο-τύπος του Α που βρίσκεται χαμηλότερα στην ταξινομία και που περιλαμβάνει το ίδιο πλήθος οντοτήτων με τον Α (διαφορετικά θα υπήρχε και άλλος υπο-τύπος στο ίδιο επίπεδο της ταξινομίας που θα παρουσίαζε το ίδιο χαρακτηριστικό, κάτι που αντιβαίνει στο δεύτερο κριτήριο). Έτσι οι συγκρίσεις τείνουν να είναι πιο ακριβείς. Αντί να αναφέρονται σε γενικούς τύπους όπως *βάζα*, *εκθέματα*, προτιμώνται ειδικότεροι τύποι όπως *αμφορείς* ή *νομίσματα*. χρησιμοποιούνται οι πιο συγκεκριμένοι υπο-τύποι που επιτυγχάνουν την ίδια σύγκριση (αφού περιλαμβάνουν τις ίδιες επιμέρους οντότητες).

5. **Συγκρίσεις ομοιότητας μεταξύ «κοντινών» οντοτήτων-τύπων:** Μεταξύ όλων των συγκρίσεων που επιτρέπουν τα προηγούμενα κριτήρια, προτιμώνται οι συγκρίσεις με τύπους που είναι υπερ-τύποι, υπο-τύποι ή ίδιοι με τον τύπο της οντότητας που περιγράφουμε. Θεωρείται, δηλαδή, ότι οι συγκρίσεις με τύπους που έχουν κάποια σχέση προγόνου, απογόνου ή ταυτότητας με τον τύπο της τρέχουσας οντότητας είναι πιο πετυχημένες από άλλες.

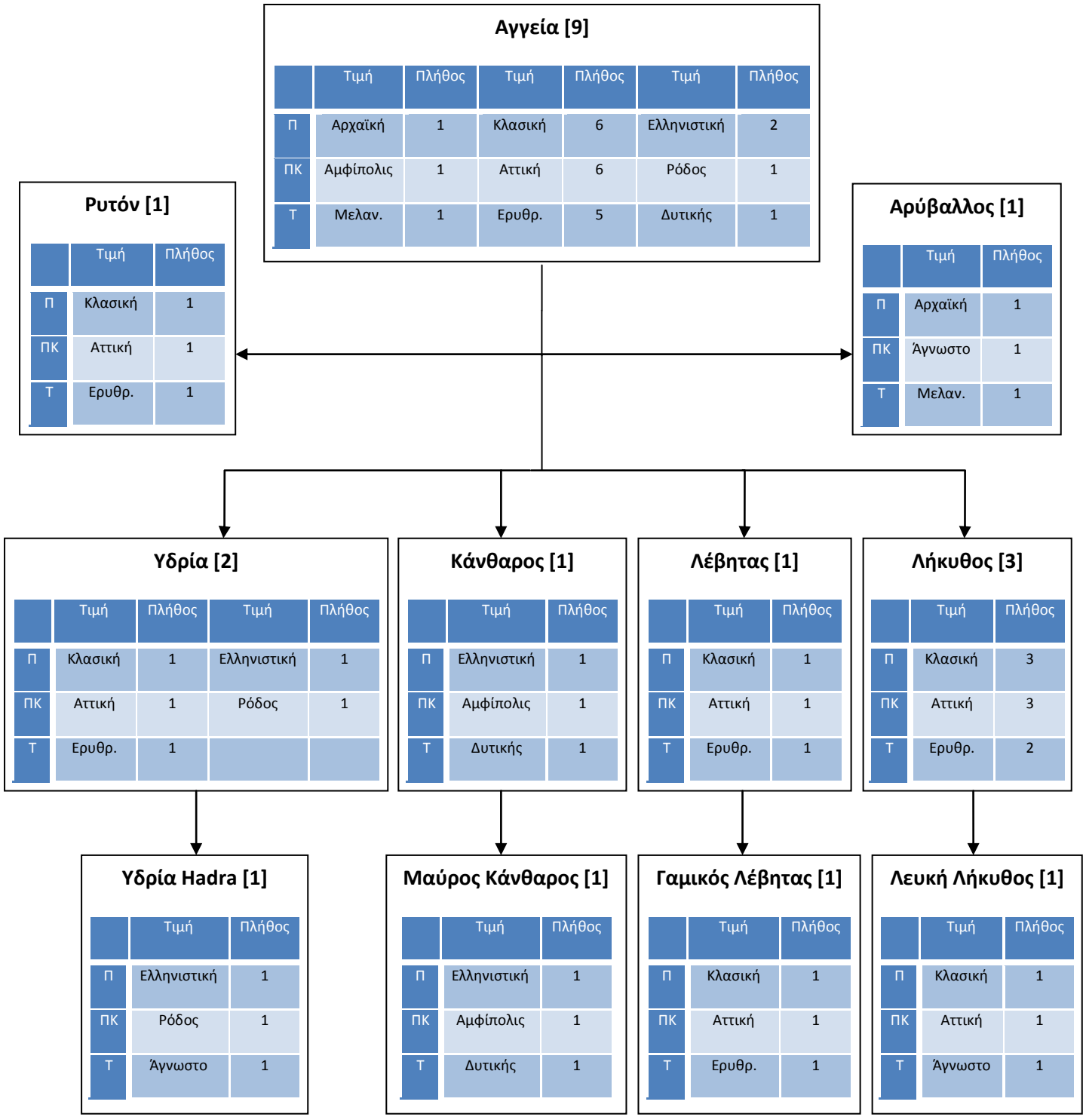
2.4.3 Ο αλγόριθμος

Έστω ότι έχουμε ήδη περιγράψει τις παρακάτω οντότητες. Τα χαρακτηριστικά των οντοτήτων είναι: περίοδος (Π), Περιοχή Κατασκευής (ΠΚ) και Τεχνοτροπία (Τ).

Τύπος	Περίοδος που ανήκει	Περιοχή που κατασκευάστηκε	Τεχνοτροπία
Υδρία	Κλασική	Αττική	Ερυθρόμορφη
Ρυτόν	Κλασική	Αττική	Ερυθρόμορφη
Γαμικός Λέβητας	Κλασική	Αττική	Ερυθρόμορφη
Υδρία Hadra	Ελληνιστική	Ρόδος	Άγνωστο
Μαύρος Κάνθαρος	Ελληνιστική	Αμφίπολις	Δυτικής
Αρύβαλλος	Αρχαϊκή	Άγνωστο	Μελανόμορφη
Λευκή Λήκυθος	Κλασική	Αττική	Άγνωστο
Λήκυθος	Κλασική	Αττική	Ερυθρόμορφη
Λήκυθος	Κλασική	Αττική	Ερυθρόμορφη

Το νέο έκθεμα που θέλουμε να περιγράψουμε είναι μία μελανόμορφη λήκυθος, της αρχαϊκής περιόδου, η οποία προέρχεται από την Αττική.

Η δυναμική ταξινομία που συνοψίζει τις οντότητες που έχουμε ήδη περιγράψει δίνεται στο παρακάτω σχήμα:



2.4.4 Πρώτο στάδιο

Στο στάδιο αυτό ανταποκρινόμαστε στο τρίτο κριτήριο που αναφέραμε. Δημιουργούμε ένα αντίγραφο της δυναμικής ταξινομίας, με το οποίο και εργαζόμαστε σε αυτό και τα επόμενα βήματα, και αφαιρούμε από το αντίγραφο τα χαρακτηριστικά (ζεύγη ιδιοτήτων-τιμών) που δεν

συμφωνούν με την οντότητα που περιγράφουμε, εκτός κι αν πρόκειται για χαρακτηριστικά που αντικρούουν μια κατάσταση που τείνει να παγιωθεί, ένα «μοτίβο». Ο λόγος είναι ότι τα χαρακτηριστικά που δεν συμφωνούν με την περιγραφόμενη οντότητα μπορούν να οδηγήσουν μόνο σε συγκρίσεις αντίθεσης και το τρίτο κριτήριο επιτρέπει συγκρίσεις αντίθεσης μόνο όταν αντικρούεται ένα μοτίβο. Για να θεωρήσουμε ότι έχει δημιουργηθεί κάποιο μοτίβο για ένα χαρακτηριστικό της ταξινομίας, απαιτούμε το χαρακτηριστικό να υπάρχει σε δύο ή περισσότερες από τις προηγούμενες οντότητες, αριθμός που θα μπορούσε να αυξηθεί ανάλογα με την εφαρμογής.

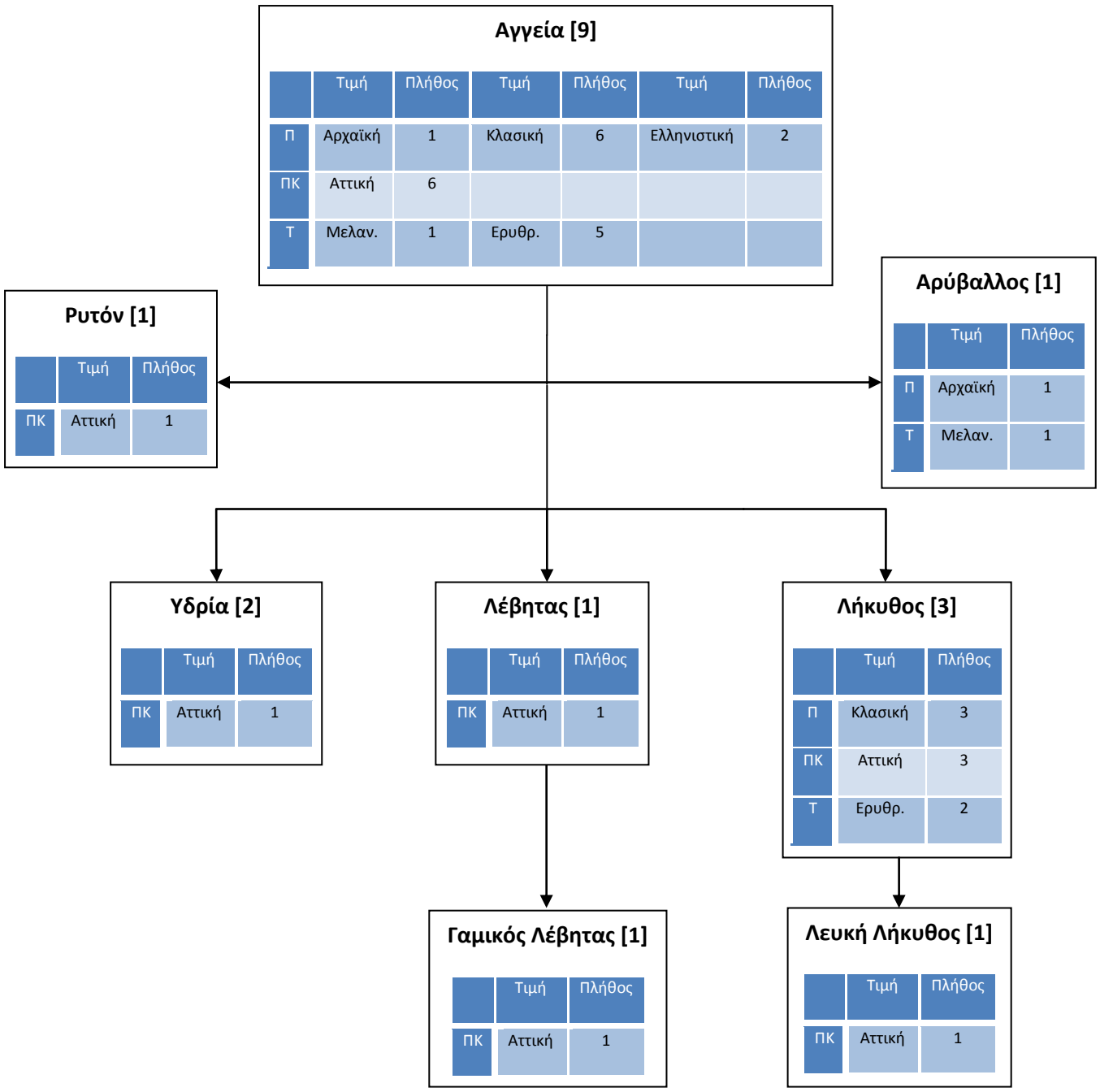
Πιο συγκεκριμένα, ο αλγόριθμος εξετάζει κάθε κόμβο του δέντρου και αφαιρεί από τον κόμβο κάθε χαρακτηριστικό που δεν υπάρχει (δεν υπάρχει η ίδια ιδιότητα με την ίδια τιμή) στην περιγραφόμενη οντότητα, εκτός αν:

- το χαρακτηριστικό εμφανίζεται σε περισσότερες από μία προηγούμενες οντότητες του κόμβου και ταυτόχρονα
- ο τύπος του κόμβου είναι ίδιος με τον τύπο της περιγραφόμενης οντότητας ή υπέρ-τύπος της (ώστε να είναι δυνατόν η περιγραφόμενη οντότητα να καταστρέψει ένα μοτίβο του κόμβου).

Παρακάτω δίνεται ο ψευδο-κώδικας αυτού του σταδίου:

```
Για κάθε κόμβο (του αντιγράφου) της δυναμικής ταξινομίας επανάλαβε
    Για κάθε χαρακτηριστικό του κόμβου επανάλαβε
        Εάν το χαρακτηριστικό δεν υπάρχει στην περιγραφόμενη οντότητα
        τότε
            Εάν το χαρακτηριστικό υπάρχει μόνο σε μία οντότητα του
            κόμβου ή εάν ο τύπος του κόμβου δεν είναι πρόγονος του
            τύπου της περιγραφόμενης οντότητας, ούτε είναι ο ίδιος
            με τον τύπο της περιγραφόμενης οντότητας τότε
                Αφαίρεσε το χαρακτηριστικό.
            Τέλος αν
        Τέλος αν
    Τέλος επανάληψης
Τέλος επανάληψης
```

Έτσι η δυναμική ταξινομία του παραδείγματος μετά το πρώτο βήμα γίνεται:



2.4.5 Δεύτερο στάδιο

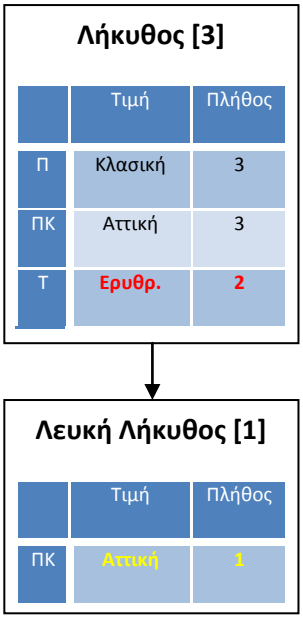
Σε αυτό το στάδιο ανταποκρινόμαστε στο πρώτο, το δεύτερο και το τέταρτο κριτήριο που αναφέραμε. Συγκεκριμένα, ο αλγόριθμος ξεκινά από το στενότερο πεδίο αναζήτησης, που περιλαμβάνει μόνο τον κόμβο της δυναμικής ταξινομίας ο οποίος αντιστοιχεί στον τύπο της περιγραφόμενης οντότητας, και σταδιακά επεκτείνει το πεδίο αναζήτησης, όπως εξηγήθηκε παραπάνω. Στην περίπτωση του παραδείγματός μας, το πεδίο αναζήτησης περιλαμβάνει αρχικά μόνο τον κόμβο «Λήκυθος» και κατόπιν γίνεται το δέντρο με ρίζα «Αγγεία». (Στην πραγματικότητα υπάρχουν και γενικότεροι κόμβοι από το «Αγγεία», αλλά έχουμε ορίσει ότι τα «Αγγεία» δεν

επιτρέπεται να γενικευθούν περαιτέρω.) Σε κάθε επέκταση του πεδίου αναζήτησης, εξετάζουμε κάθε χαρακτηριστικό της ρίζας του πεδίου και κάνουμε τους εξής ελέγχους:

- **Ελέγχουμε αν το χαρακτηριστικό (ζεύγος ιδιότητα-τιμή) είναι κοινό για όλες τις οντότητες της ρίζας:** Ουσιαστικά εξετάζουμε αν μπορούμε να αναφερθούμε σε όλες τις οντότητες της ρίζας με το όνομα του τύπου τους, αν κάνουμε σύγκριση ως προς αυτό το χαρακτηριστικό (κριτήριο 1). Στο παράδειγμά μας, η ρίζα του πεδίου αναζήτησης είναι αρχικά οι λήκυθοι. Έστω ότι το χαρακτηριστικό που εξετάζουμε είναι το [ΠΚ: Αττική]. Και οι τρεις λήκυθοι που έχουμε δει κατασκευάστηκαν στην Αττική, άρα ο έλεγχος επιτυγχάνει (π.χ. μπορούμε να πούμε «*Όπως όλες οι προηγούμενες λήκυθοι, αυτό το έκθεμα κατασκευάστηκε στην Αττική.»).*

    - **Αν ο προηγούμενος έλεγχος επιτύχει, εξετάζουμε επιπλέον αν το ίδιο χαρακτηριστικό υπάρχει και σε απόγονο της ρίζας για το ίδιο πλήθος οντοτήτων** (κριτήρια 2 και 4). Αν ναι, κρατάμε στη δυναμική ταξινομία το χαρακτηριστικό του απογόνου, ενώ το χαρακτηριστικό της ρίζας αφαιρείται. Διαφορετικά, κρατάμε στη δυναμική ταξινομία το χαρακτηριστικό της ρίζας και το χαρακτηριστικό του απογόνου αφαιρείται. Στο παράδειγμά μας, αν και η μοναδική λευκή λήκυθος που έχουμε δει είναι κι αυτή κατασκευασμένη στην Αττική, το πλήθος του χαρακτηριστικού [ΠΚ: Αττική] είναι διαφορετικό στον κόμβο «Λευκή Λήκυθος» (ένα, αντί για δύο στον κόμβο «Λήκυθος»), οπότε κρατάμε το χαρακτηριστικό [ΠΚ: Αττική] του «Λήκυθος». Για να προτιμήσουμε το χαρακτηριστικό του ειδικότερου κόμβου αντί του ίδιου χαρακτηριστικού της ρίζας (κριτήριο 4), πρέπει το πλήθος του χαρακτηριστικού στον ειδικότερο κόμβο να είναι το ίδιο με το πλήθος στη ρίζα, γιατί αν δεν είναι, τότε υπάρχει και κάποιος άλλος κόμβος του πεδίου αναζήτησης που παρουσιάζει το ίδιο χαρακτηριστικό (κάτι που αντιβαίνει στο κριτήριο 2).

Στο παρακάτω σχήμα με κόκκινο χρώμα έχουν σημειωθεί τα χαρακτηριστικά που δεν είναι κοινά για όλες τις οντότητες της αρχικής ρίζας του πεδίου αναζήτησης («Λήκυθος»), ενώ με κίτρινο, εκείνα που παραβιάζουν τον δεύτερο έλεγχο (ίδιο πλήθος χαρακτηριστικού).



Τα χαρακτηριστικά που προκρίνονται για να χρησιμοποιηθούν σε μία πιθανή σύγκριση είναι:

Λήκυθος--[ΠΚ: Αττική] (Όλες οι λήκυθοι που είδαμε, προέρχονταν από την Αττική.)

Λήκυθος--[Π: Κλασική] (Όλες οι λήκυθοι που είδαμε, ήταν κατασκευασμένες την κλασσική περίοδο.)

- **Αν ο πρώτος έλεγχος αποτύχει, δηλαδή αν το χαρακτηριστικό δεν είναι κοινό για όλες τις οντότητες της ρίζας:** Στην περίπτωση αυτή δεν μπορούμε να χρησιμοποιήσουμε τη ρίζα του πεδίου αναζήτησης για να κάνουμε κάποια σύγκριση. Ωστόσο, ενδέχεται να υπάρχει κάποιος απόγονός της που να μπορεί να χρησιμοποιηθεί. Έτσι, ελέγχουμε κάθε απόγονο-κόμβο της ρίζας ως προς το αν έχει κάποιο χαρακτηριστικό που το έχουν όλες οι οντότητες του απογόνου-κόμβου και δεν το έχει κανένας άλλος κόμβος του πεδίου αναζήτησης. Στο παράδειγμά μας κάτι τέτοιο ισχύει μόνο για το χαρακτηριστικό [Π: αρχαϊκή περίοδος] και το χαρακτηριστικό [Τ: μελανόμορφος] του κόμβου «Αρύβαλλος», όταν η ρίζα του πεδίου αναζήτησης έχει γίνει «Αγγεία».

  Πρέπει να σημειώσουμε ότι στην περίπτωση των διαφορετικών τιμών δεν εξετάζουμε τους υπο-τύπους που έχουμε ήδη εξετάσει σε προηγούμενες επαναλήψεις αφού τα χαρακτηριστικά τους είναι έγκυρα για κάποιο στενότερο πεδίο αναζήτησης.

Έτσι τα επιπλέον χαρακτηριστικά που προκρίνονται για να χρησιμοποιηθούν σε μία πιθανή σύγκριση είναι:

Αρύβαλλος--[Τ: μελανόμοροφος] (Ο μοναδικός αρύβαλλος που είδαμε είναι μελανόμορφης τεχνοτροπίας) και

Αρύβαλλος--[Τ: μελανόμοροφος] (Ο μοναδικός αρύβαλλος που είδαμε κατασκευάστηκε κατά την αρχαϊκή περίοδο.)

Ακολουθεί ο ψευδοκώδικας αυτού του σταδίου.

```
Βρες σε ποιον κόμβο (τύπο) της δυναμικής ταξινομίας (όπως έχει διαμορφωθεί
από το προηγούμενο στάδιο) αντιστοιχεί η οντότητα που περιγράφουμε.

Αρχή Επανάληψης
    Για κάθε χαρακτηριστικό του κόμβου επανάλαβε
        Εάν το χαρακτηριστικό είναι κοινό για όλες τις οντότητες που
        ανήκουν στον κόμβο τότε
            Για κάθε παιδί του κόμβου επανάλαβε
                Εάν (συνθήκη Σ) το χαρακτηριστικό υπάρχει σε όλες
                τις οντότητες που ανήκουν στο παιδί ΚΑΙ υπάρχει σε
                τόσες οντότητες όσες και στον κόμβο-πατέρα τότε
                    Ψάξε αναδρομικά και τα παιδιά του παιδιού,
                    μέχρι να μην ισχύει η συνθήκη (Σ) σε κανένα
                    από τα νέα παιδιά ή να φτάσουμε σε φύλλο.
                Τέλος αν
            Τέλος επανάληψης
            Εάν η συνθήκη (Σ) ικανοποιήθηκε σε κάποιον απόγονο του
            κόμβου τότε
                Σβήσε το χαρακτηριστικό από τον κόμβο και κράτα το
                μόνο στον πιο ειδικό απόγονο όπου ικανοποιήθηκε η
                συνθήκη (Σ).
            Τέλος αν
        Αλλιώς
            Για κάθε παιδί του κόμβου (που δεν έχουμε ήδη εξετάσει)
            επανάλαβε
                Εάν (συνθήκη Σ) το χαρακτηριστικό υπάρχει σε όλες
                τις οντότητες που ανήκουν στο παιδί ΚΑΙ υπάρχει σε
                τόσες οντότητες όσες και στον κόμβο-πατέρα τότε
                    Ψάξε αναδρομικά και τα παιδιά του παιδιού
                    μέχρι να μην ισχύει η συνθήκη (Σ) σε κανένα
                    από τα νέα παιδιά ή να φτάσουμε σε φύλλο.
                Τέλος αν
            Τέλος επανάληψης
```

Σβήσε το χαρακτηριστικό από τον κόμβο και κράτα το μόνο στον πιο ειδικό απόγονο όπου ικανοποιήθηκε η συνθήκη (Σ).

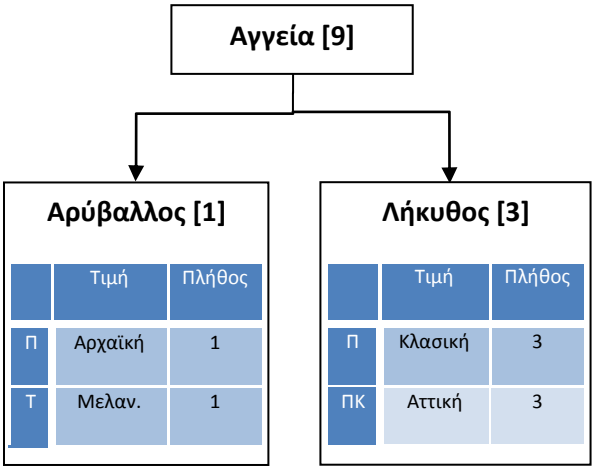
Τέλος αν

Τέλος επανάληψης

Πάμε στον υψηλότερο κόμβο (νέα ρίζα του πεδίου αναζήτησης)

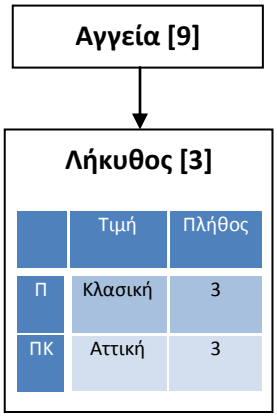
Μέχρις ότου να μην υπάρχει υψηλότερος κόμβος ή να μην επιτρέπεται να χρησιμοποιηθεί ως ρίζα του πεδίου αναζήτησης.

Τελικά η δυναμική ταξινομία μετά το πέρας και του δεύτερου σταδίου του αλγορίθμου γίνεται:



2.4.6 Τρίτο στάδιο

Σύμφωνα με το πέμπτο κριτήριο, αν μπορεί να γίνει σύγκριση ομοιότητας με περισσότερους του ενός τύπους (κόμβους της δυναμικής ταξινομίας), τότε θα πρέπει να προτιμώνται εκείνοι οι τύποι (κόμβοι) που είναι υπερ-τύποι, υπο-τύποι ή ίδιοι με τους τύπους της περιγραφόμενης οντότητας. Στο παράδειγμά μας, αυτό έχει ως αποτέλεσμα η δυναμική ταξινομία να γίνει:



Τελικά η δυναμική ταξινομία περιέχει μόνο έναν κόμβο (τύπο) για σύγκριση με δύο χαρακτηριστικά: την περίοδο και την τοποθεσία. Αφού δεν έχουμε κάποιο άλλο κριτήριο, η

επιλογή του χαρακτηριστικού που θα χρησιμοποιήσουμε για τη σύγκριση θα γίνει τυχαία. Η τελική σύγκριση, λοιπόν, θα είναι της μορφής:

> *Σε αντίθεση με τις προηγούμενες ληκύθους, που είχαν κατασκευαστεί την κλασική περίοδο, αυτή η λήκυθος είναι της αρχαϊκής περιόδου.*

ή

> *Και αυτή η λήκυθος, όπως και οι προηγούμενες, προέρχεται από την περιοχή της Αττικής.*

“An ounce of action is worth a ton of theory”
Friedrich Engels

3. Το σύστημά μας

3.1 Επιλογή αλγορίθμου

Μετά τη βιβλιογραφική μελέτη του προηγούμενου κεφαλαίου, αποφασίσαμε ότι ο βασικός άξονας των συγκρίσεών μας θα είναι ο αλγόριθμος Μελέγκογλου. Ο λόγος που απορρίψαμε τον αλγόριθμο Milosavljevic είναι ότι προσπαθεί να κάνει συγκρίσεις μεταξύ δύο μεμονωμένων οντοτήτων και όχι ολόκληρων ομάδων. Αυτό σημαίνει ότι δε θα μπορούσε να παραγάγει συγκρίσεις της μορφής:

> ***Όπως όλα τα προηγούμενα αγγεία…***

Επίσης για να αναφερθεί σε μία οντότητα, ο αλγόριθμος Milosavljevic καταφεύγει σε χρήση μακροσκελών αναφορικών εκφράσεων.

Για παράδειγμα, αν έχει ήδη περιγράψει τρεις υπολογιστές με χαρακτηριστικά:

	Υπολογιστής Α	Υπολογιστής Β	Υπολογιστής Γ
Επεξεργαστής	X	Y	Y
Σκληρός δίσκος	60 GB	60 GB	60 GB
Ποσότητα Μνήμης	1024 MB	1024 MB	2048 MB
Δίκτυο	Ethernet	Ethernet	Ethernet

και στη συνέχεια προσπαθήσει να περιγράψει έναν νέο (έστω Δ) με χαρακτηριστικά:

	Υπολογιστής Δ
Επεξεργαστής	X
Σκληρός δίσκος	60 GB
Ποσότητα Μνήμης	2048 MB
Δίκτυο	WiFi

θα παραγάγει μία σύγκριση της μορφής:

> ***Όπως ο υπολογιστής με τον επεξεργαστή Y και το σκληρό δίσκο χωρητικότητας 60 GB, αυτός ο υπολογιστής έχει 2048 MB μνήμης.***

Αντίθετα, ο αλγόριθμος Μελέγκογλου αναφέρεται σε προηγούμενες οντότητες μόνο μέσω των τύπων τους, μόνο όποτε αυτό είναι δυνατόν, χωρίς μακροσκελείς αναφορικές εκφράσεις, παράγοντας συγκρίσεις όπως η ακόλουθη:

> ***Αντίθετα με τους προηγούμενους υπολογιστές, που είχαν δικτύωση Ethernet, αυτός ο υπολογιστής υποστηρίζει WiFi.***

Ο δεύτερος λόγος που καταλήξαμε στον αλγόριθμο Μελέγκογλου είναι ότι ουσιαστικά παράγει τις περισσότερες από τις κατηγορίες συγκρίσεων που όρισε η Milosavljevic, εξαιρώντας κατηγορίες συγκρίσεων που δεν είναι επιθυμητές σε ένα σύστημα περιηγήσεων σαν αυτό στο οποίο επικεντρωνόμαστε.

3.1.1 Ευθεία σύγκριση

Ο τύπος αυτός των συγκρίσεων δεν υποστηρίζεται από τον αλγόριθμο Μελέγκογλου. Ωστόσο, αυτού του είδους οι συγκρίσεις δεν είναι επιθυμητές στις εφαρμογές που μελετάμε, όπου ο χρήστης επιλέγει κάθε φορά μία οντότητα (π.χ. έκθεμα) που θέλει να περιγραφεί και όχι ένα ζεύγος οντοτήτων μεταξύ των οποίων θέλει να γίνει εξαντλητική σύγκριση. Ο επισκέπτης, δηλαδή, εστιάζει κάθε φορά την προσοχή του σε μία μόνο οντότητα και δεν ζητά εξαντλητικές συγκρίσεις μεταξύ δύο συγκεκριμένων οντοτήτων. Ο αλγόριθμος Milosavljevic παράγει, βέβαια, ευθείες, εξαντλητικές συγκρίσεις ακόμα και σε αυτό το σενάριο, όταν βρίσκει κάποια οντότητα που μοιάζει πάρα πολύ με την οντότητα που εξετάζει ο επισκέπτης. Αλλά σε αυτή την περίπτωση, δεν φαίνεται καλή επιλογή η χρήση ευθείας, εξαντλητικής σύγκρισης, γιατί αφού οι δύο οντότητες μοιάζουν πάρα πολύ, τα περισσότερα χαρακτηριστικά τους θα είναι ίδια. Πιο εύλογη θα φαινόταν σε αυτή

την περίπτωση η χρήση σύγκρισης αποσαφήνισης ή αναφοράς σε παρόμοια οντότητα/τύπο ή τύπο (βλ. παρακάτω).

3.1.2 Σύγκριση ειδοποιού διαφοράς τύπων

Στις συγκρίσεις ειδοποιού διαφοράς, εξετάζουμε έναν τύπο και αναφέρουμε τα ιδιαίτερα χαρακτηριστικά που κάνουν τους υπο-τύπους του να ξεχωρίζουν μεταξύ τους. Όμως σε ένα σύστημα που παράγει περιγραφές για κάποια συγκεκριμένη κάθε φορά οντότητα, ποτέ δε θα ζητηθεί η περιγραφή ενός ολόκληρου τύπου. Άρα μία τέτοια σύγκριση δεν μας απασχολεί άμεσα.

Υπάρχει, ωστόσο, ένα είδος συγκρίσεων που συγγενεύει με τις συγκρίσεις ειδοποιού διαφοράς τύπων και το οποίο θα ήταν χρήσιμο στο σύστημα με το οποίο ασχολείται η παρούσα εργασία. Για παράδειγμα, την πρώτη φορά που εξετάζει μια λήκυθο ο επισκέπτης, το σύστημα της εργασίας παράγει ένα κείμενο της μορφής: «Αυτή είναι μια λήκυθος, ένα είδος αγγείου. Αυτή η λήκυθος κατασκευάστηκε τον 5ο π.Χ. αιώνα και...». Ας υποθέσουμε ότι αμέσως μετά ο επισκέπτης συναντά τον πρώτο του αμφορέα και ας υποθέσουμε ότι η έννοια του αμφορέα θεωρείται αρχικά άγνωστη (σύμφωνα με το μοντέλο χρήστη). Θα παραγόταν ένα κείμενο της μορφής: «Αυτός είναι ένας αμφορέας, ένα είδος αγγείου. Κατασκευάστηκε...», ενώ θα ήταν καλύτερα αν το κείμενο γινόταν: «Αυτός είναι ένας αμφορέας, ένα είδος αγγείου. Αντίθετα από τις ληκύθους, το άλλο είδος αγγείου που συναντήσατε νωρίτερα, οι αμφορείς... , ενώ οι λήκυθοι... Αυτός ο αμφορέας κατασκευάστηκε...». Αν ο επισκέπτης είχε συναντήσει προηγουμένως και ένα άλλο είδος αγγείου, έστω τα λυχνάρια, θα έπρεπε να παραχθεί κάτι σαν: «Αυτός είναι ένας αμφορέας, ένα είδος αγγείου. Αντίθετα από τις ληκύθους και τα λυχνάρια, τα άλλα είδη αγγείων που συναντήσατε νωρίτερα, οι αμφορείς..., ενώ οι λήκυθοι... και τα λυχνάρια... Αυτός ο αμφορέας...». Ουσιαστικά, δηλαδή, θα θέλαμε να κάνουμε μία σύγκριση ειδοποιού διαφοράς μεταξύ των υπο-τύπων του τύπου της περιγραφόμενης οντότητας τους οποίους έχει συναντήσει ήδη ο χρήστης. Αυτού του είδους οι συγκρίσεις, όμως, δεν εξετάστηκαν περαιτέρω στη διάρκεια αυτής της εργασίας, λόγω των χρονικών της περιορισμών. Ένα πρόβλημα είναι ότι οι συγκρίσεις γίνονται πολύ κουραστικές όταν ο χρήστης έχει συναντήσει ήδη πολλούς υπο-τύπους του τύπου της περιγραφόμενης οντότητας, όπως διαφαίνεται ήδη από το προηγούμενο παράδειγμα.

Οι συγκρίσεις της παραπάνω μορφής θα μπορούσαν να βασίζονται στους περιορισμούς των τιμών των ιδιοτήτων που θέτει η οντολογία OWL στους διάφορους υπο-τύπους (υπο-τάξεις) οντοτήτων. Για παράδειγμα, οι περιορισμοί τιμών της οντολογίας μπορεί να δείχνουν ότι όλοι οι κούροι (υπο-τύπος των αγαλμάτων) κατασκευάστηκαν κατά την αρχαϊκή περίοδο, ενώ τα αυτοκρατορικά πορτρέτα (άλλος υπο-τύπος αγαλμάτων) κατασκευάστηκαν κατά τη ρωμαϊκή περίοδο.

3.1.3 Σύγκριση με συμπλήρωμα συνόλου

Σε αυτόν τον τύπο συγκρίσεων προσπαθούμε να συγκρίνουμε την περιγραφόμενη οντότητα με όλες τις υπόλοιπες οντότητες του ίδιου τύπου που υπάρχουν στην οντολογία. Αντίθετα, ο αλγόριθμος Μελέγκογλου συγκρίνει μόνο με προηγούμενες οντότητες που έχουμε δει και όχι απαραιτήτως με οντότητες του ιδίου τύπου με την περιγραφόμενη οντότητα. Είναι, όμως, εύκολο να συμπληρωθεί με έναν άλλο αλγόριθμο που παράγει συγκρίσεις με όλες τις οντότητες του ιδίου τύπου που υπάρχουν στην οντολογία. Ο συμπληρωματικός αυτός αλγόριθμος, που αναπτύχθηκε κατά τη διάρκεια της εργασίας, περιγράφεται σε επόμενη ενότητα. Μπορεί να παράγει συγκρίσεις μοναδικότητας (η περιγραφόμενη οντότητα έχει ένα χαρακτηριστικό που δεν υπάρχει σε καμία άλλη οντότητα του ιδίου τύπου ή υπερ-τύπου της) καθώς και καθολικής ομοιότητας (η περιγραφόμενη οντότητα έχει ένα χαρακτηριστικό που είναι κοινό για όλες τις οντότητες του τύπου της ή κάποιου υπερ-τύπου). Επειδή η καθολική ομοιότητα είναι αρκετά δύσκολο να επιτευχθεί, ειδικά όταν υπάρχουν πολλές οντότητες του ίδιου τύπου, ο συμπληρωματικός αλγόριθμος μπορεί να παράγει και συγκρίσεις με την πλειοψηφία (αντί το σύνολο) των οντοτήτων του ιδίου τύπου (ή υπερ-τύπου).

> ***Αυτός είναι ο μόνος φορητός υπολογιστή (του καταστήματος) με επεξεργαστή X.***
>
> ***(παράδειγμα μοναδικότητας)***
>
> ***Όπως όλοι οι φορητοί υπολογιστές (του καταστήματος), ο σκληρός του δίσκος έχει ταχύτητα 5.200 rpm.***
>
> ***(παράδειγμα καθολικής ομοιότητας)***
>
> ***Όπως οι περισσότεροι φορητοί υπολογιστές (του καταστήματος), το chip γραφικών αυτού του φορητού υπολογιστή χρησιμοποιεί τη μνήμη RAM.***
>
> ***Παράδειγμα ομοιότητας με τις περισσότερες οντότητες)***

3.1.4 Σύγκριση αποσαφήνισης

Σε αυτόν τον τύπο συγκρίσεων προσπαθούμε να συγκρίνουμε την τρέχουσα οντότητα με κάποια άλλη πολύ παρόμοια, την οποία όμως δεν έχουμε δει ακόμα. Σκοπός της σύγκρισης είναι να αποφύγουμε να μπερδέψει ο χρήστης τις δύο οντότητες. Κάτι τέτοιο ωστόσο φαίνεται αφύσικο σε μία εφαρμογή περιήγησης, αφού προσθέτει στην τρέχουσα περιγραφή αναφορές σε οντότητες (εκθέματα) που ο χρήστης δεν έχει ακόμα συναντήσει. Αντί αυτού, είναι προτιμότερο να

συμπεριληφθεί η παρόμοια οντότητα στη λίστα των υπερσυνδέσμων προς προτεινόμενες σχετικές οντότητες που συστήματα όπως του M-PIRO προσθέτουν στο τέλος της τρέχουσας περιγραφής.

3.1.5 Αναφορά σε παρόμοια οντότητα/τύπο

Όπως ήδη αναφέραμε, οι συγκρίσεις αναφοράς σε παρόμοια οντότητα συχνά χρησιμοποιούν μακροσκελείς αναφορικές εκφράσεις για να καταφέρνουν να αναφερθούν μονοσήμαντα σε μια μεμονωμένη παρόμοια οντότητα του παρελθόντος. Έτσι, όμως, αποσπούν την προσοχή του επισκέπτη από την τρέχουσα οντότητα (π.χ. «Όπως η ερυθρόμορφη λήκυθος που βρέθηκε στην Αττική και κατασκευάστηκε τον 5ο αιώνα π.Χ., αυτός ο αμφορέας...»). Στις εφαρμογές δε που εξετάζουμε, το πρόβλημα οξύνεται ακόμα περισσότερο όσο αυξάνεται το πλήθος των εκθεμάτων που έχουμε ήδη περιγράψει, αφού οι αναφορικές εκφράσεις συχνά απαιτείται να γίνουν πιο μακροσκελείς για να είναι μονοσήμαντες. Ένα συναφές πρόβλημα είναι ότι ο επισκέπτης ενδέχεται να μη θυμάται το προηγούμενο έκθεμα στο οποίο αναφερόμαστε ή να μη θυμάται αρκετές λεπτομέρειες για αυτό, ιδιαίτερα αν το είχε εξετάσει πολλή ώρα πριν, με αποτέλεσμα να μην μπορεί να το ανασύρει από τη μνήμη του. Επίσης, συχνά υπάρχουν ομοιότητες με πολλά από τα εκθέματα που ήδη έχει δει, οπότε δημιουργείται στο χρήστη η απορία γιατί το τρέχον έκθεμα συγκρίνεται με ένα μόνο από εκείνα ή/και γιατί δεν συγκρίνεται με κάποιο άλλο. Ο αλγόριθμος Μελέγκογλου αντιμετωπίζει αυτά τα προβλήματα με το να συγκρίνει το τρέχον έκθεμα με ολόκληρες ομάδες προηγούμενων εκθεμάτων, οι οποίες επιλέγονται έτσι, ώστε να μπορούμε να αναφερθούμε σε ολόκληρη την ομάδα σύντομα, μέσω του τύπου των μελών της (π.χ. «Όπως όλοι οι υπολογιστές που είδατε νωρίτερα, ...»).

3.1.6 Αναλογία

Σε αυτόν τον τύπο συγκρίσεων προσπαθούμε να εκμεταλλευτούμε χαρακτηριστικά πολύ γνωστών οντοτήτων (ή τύπων) ως μέτρα σύγκρισης για την περιγραφή των χαρακτηριστικών της τρέχουσας οντότητας (για παράδειγμα «όσο μία γάτα»). Για να πετύχουμε όμως κάτι τέτοιο, θα πρέπει οι πολύ γνωστές αυτές οντότητες να περιλαμβάνονται στην οντολογία της εφαρμογής (κάτι που συνήθως δεν συμβαίνει) ή να χρησιμοποιήσουμε μια πρόσθετη οντολογία που να περιέχει πολύ συνηθισμένες οντότητες (στο παράδειγμά μας τη γάτα). Στη διάρκεια της εργασίας δεν ασχοληθήκαμε με συγκρίσεις αναλογίας, επειδή θέλαμε να αποφύγουμε τις επεμβάσεις στην οντολογία της εφαρμογής και τη χρήση πρόσθετων οντολογιών. Η παραγωγή συγκρίσεων αναλογίας θα μπορούσε, πάντως, να αποτελεί έναν επιπλέον, συμπληρωματικό αλγόριθμο (όπως συμβαίνει με τις συγκρίσεις μοναδικότητας και καθολικής ομοιότητας). Πολύ εύκολα θα μπορούσαμε να χρησιμοποιήσουμε τον αντίστοιχο υπο-αλγόριθμο της Milosavljevic.

3.2 Τροποποιήσεις του αλγορίθμου Μελέγκογλου

Όπως ήδη αναφέραμε, ο βασικός αλγόριθμος που χρησιμοποιείται στο σύστημα της εργασίας για τη δημιουργία των συγκρίσεων είναι ο αλγόριθμος Μελέγκογλου.

Ένα πρόβλημα που παρατηρήσαμε ήταν ότι στο τέλος του τρίτου σταδίου του αλγορίθμου παράγονταν συχνά και συγκρίσεις για ιδιότητες όχι ιδιαίτερα σημαντικές. Για παράδειγμα:

> ***Όπως όλα τα προηγούμενα νομίσματα, σήμερα αυτό το τετράδραχμο βρίσκεται στο Νομισματικό Μουσείο.***

Για να βελτιώσουμε το ενδιαφέρον των συγκρίσεων, επιτρέψαμε οι συγκρίσεις να βασίζονται μόνο σε ιδιότητες που εισάγονται χειρωνακτικά σε μία λίστα ως ιδιαίτερα ενδιαφέρουσες, κάτι που είχε γίνει και στο έργο M-PIRO. Στην οντολογία του M-PIRO που χρησιμοποιήσαμε, η οποία περιγράφει τα εκθέματα μιας μουσειακής συλλογής, η λίστα αυτή ορίστηκε να περιέχει μόνο τις ιδιότητες:

i. Περιοχή προέλευσης
ii. Υλικό κατασκευής
iii. Χρονολογική περίοδος δημιουργίας
iv. Τεχνοτροπία ζωγραφικής.

Επίσης, αν στην αρχή του τρίτου σταδίου υπάρχουν πολλές υποψήφιες συγκρίσεις (που θα ικανοποιούν σίγουρα τα πρώτα τέσσερα κριτήρια της ενότητας), επιλέγουμε κατά σειρά προτεραιότητας:

- σύγκριση αντίθεσης
- ομαδική σύγκριση ομοιότητας (σύγκριση με περισσότερα του ενός προηγούμενα εκθέματα, αναφέροντας τον τύπο τους) που ικανοποιεί και το πέμπτο κριτήριο,
- ατομική σύγκριση ομοιότητας (σύγκριση με μεμονωμένο προηγούμενο έκθεμα, αναφέροντας τον τύπο του) που ικανοποιεί και το πέμπτο κριτήριο,
- σύγκριση ομοιότητας που δεν ικανοποιεί το πέμπτο κριτήριο (δίνοντας προτεραιότητα στις ομαδικές συγκρίσεις).

Η προτεραιότητα δόθηκε στις συγκρίσεις αντίθεσης, επειδή η εμπειρία μας έδειχνε ότι αυτές είναι συνήθως οι πιο ενδιαφέρουσες, κάτι που επαληθεύτηκε από τα πειραματικά αποτελέσματα. Σε περίπτωση που υπάρχουν περισσότερες από μία υποψήφιες συγκρίσεις αντίθεσης, επιλέγουμε τυχαία μία από αυτές.

Κατόπιν προτεραιότητα δόθηκε στις συγκρίσεις ομοιότητας (ομαδικές και ατομικές) που ικανοποιούν και το πέμπτο κριτήριο, έναντι εκείνων που δεν το ικανοποιούν, όπως προβλεπόταν

ήδη από τον αλγόριθμο. (Οι συγκρίσεις αντίθεσης ικανοποιούν πάντα και το τέταρτο κριτήριο.) Μεταξύ ομαδικών και ατομικών συγκρίσεων που ικανοποιούν το πέμπτο κριτήριο, προτεραιότητα δόθηκε στις ομαδικές, για λόγους ποικιλίας, επειδή είναι γενικά σπανιότερες.

Τέλος, στην περίπτωση που όλες οι διαθέσιμες συγκρίσεις είναι συγκρίσεις ομοιότητας (ομαδικές ή ατομικές) που δεν ικανοποιούν το πέμπτο κριτήριο, οι συγκρίσεις ταξινομούνται (ομαδικές και ατομικές μαζί) ως προς την απόσταση από το έκθεμα ή την ομάδα εκθεμάτων με το/την οποία συγκρίνουν. Για παράδειγμα, στη σύγκριση «Όπως η λήκυθος που είδατε...», αν η λήκυθος ήταν το προ-προηγούμενο έκθεμα, η απόσταση είναι 2. (Στην περίπτωση των ομαδικών συγκρίσεων, χρησιμοποιούμε την απόσταση από το πιο πρόσφατο έκθεμα της ομάδας.) Στη συνέχεια επιλέγουμε τυχαία μία από τις συγκρίσεις με απόσταση μικρότερη του 5. Αν υπάρχουν και ομαδικές συγκρίσεις με απόσταση μικρότερη του 5, προτιμώνται έναντι των ατομικών για λόγους ποικιλίας.

Το σκεπτικό για τη χρήση της απόστασης όταν δεν ικανοποιείται το πέμπτο κριτήριο είναι το ακόλουθο. Οι συγκρίσεις ομοιότητας που ικανοποιούν το πέμπτο κριτήριο, δηλαδή οι συγκρίσεις όπου ο τύπος της περιγραφόμενης οντότητας είναι ο ίδιος ή υπερ-τύπος ή υπο-τύπος του τύπου με τον οποίο γίνεται η σύγκριση, δημιουργούν στον αναγνώστη την εντύπωση ότι γίνονται για να ενισχύσουν ένα «μοτίβο» (όπως στην περίπτωση των αρνητικών συγκρίσεων) που αφορά το γενικότερο εμπλεκόμενο τύπο. Για παράδειγμα, η σύγκριση «Όπως όλα τα προηγούμενα αγγεία που είδατε, αυτή η λύκηθος ακολουθεί τον ερυθρόμορφο ρυθμό.» φαίνεται να ενισχύει το «μοτίβο» ότι όλα τα αγγεία (η λύκηθοι είναι υπο-τύπος αγγείων) ακολουθούν τον ερυθρόμορφο ρυθμό. Αυτό προσφέρει στον αναγνώστη μια εξήγηση της σκοπιμότητας της σύγκρισης (η σύγκριση γίνεται για να ενισχυθεί το μοτίβο). Αντίθετα, όταν δεν ικανοποιείται το πέμπτο κριτήριο (π.χ. «Όπως όλες οι υδρίες που είδατε, αυτή η λήκυθος ακολουθεί τον ερυθρόμορφο ρυθμό.»), η σύγκριση δεν φαίνεται να ενισχύει ένα τέτοιο μοτίβο, οπότε ο αναγνώστης δεν κατανοεί τόσο εύκολα τη σκοπιμότητα της σύγκρισης, κυρίως το γιατί η σύγκριση γίνεται με το συγκεκριμένο τύπο (στο παράδειγμά μας, τις υδρίες) και όχι με άλλον. Στην περίπτωση αυτή, προτιμούμε συγκρίσεις με μικρή απόσταση, επειδή η μικρή απόσταση (πρόσφατο έκθεμα) παρέχει μια εναλλακτική αιτιολόγηση του γιατί επελέγη ο συγκεκριμένος τύπος: τα αντίστοιχα εκθέματα (οι υδρίες) είναι πολύ πρόσφατα στη μνήμη του χρήστη και φαίνεται εύλογο το τρέχον έκθεμα να συγκριθεί κυρίως με τα πρόσφατα εκθέματα.

Τέλος, σημειώνουμε ότι όταν μπορούν να γίνουν πολλές συγκρίσεις ομοιότητας με τον ίδιο τύπο για διαφορετικά χαρακτηριστικά, οι συγκρίσεις αυτές ομαδοποιούνται σε μία, όπως στο παρακάτω παράδειγμα:

> ***Όπως η άλλη λήκυθος (ή οι άλλες λήκυθοι) που εξετάσατε, αυτή η λήκυθος δημιουργήθηκε κατά τη διάρκεια της κλασικής περιόδου, ακολουθεί τον ερυθρόμορφο ρυθμό και προέρχεται και αυτή από την Αττική.***

Αντιθέτως, οι συγκρίσεις αντίθεσης περιορίζονται πάντα σε ένα μόνο χαρακτηριστικό, γιατί φαίνεται συνήθως κουραστικό να αναφέρει κανείς πολλές διαφορές:

> ***Αντίθετα από τις άλλες ληκύθους που εξετάσατε, που είχαν δημιουργηθεί στη διάρκεια της αρχαϊκής περιόδου, ακολουθούσαν το μελανόμορφο ρυθμό και προέρχονταν από τη Σπάρτη, αυτή η λήκυθος δημιουργήθηκε κατά τη διάρκεια της κλασικής περιόδου, ακολουθεί τον ερυθρόμορφο ρυθμό και προέρχεται από την Αττική.***

3.2.1 Παραλλαγή του αλγορίθμου Μελέγκογλου

Όπως έχουμε ήδη αναφέρει, ο αλγόριθμος Μελέγκογλου εντοπίζει ομοιότητες ή διαφορές με όλες τις προηγούμενες οντότητες που ανήκουν σε κάποιον τύπο. Προσπαθήσαμε να τον γενικεύσουμε, ώστε να επιστρέφει και συγκρίσεις με τις περισσότερες προηγούμενες οντότητες ενός τύπου. Για παράδειγμα, αν όλα τα αγγεία που έχουμε δει μέχρι τώρα είναι από την Αττική, και το έκθεμα που βλέπουμε αυτή τη στιγμή είναι κι αυτό από την Αττική, τότε ο αλγόριθμος Μελέγκογλου θα δημιουργούσε μία σύγκριση όπως η παρακάτω:

> ***Όπως όλα τα αγγεία που είδατε νωρίτερα, αυτό το άγαλμα προέρχεται από την Αττική.***

Αν όμως υπήρχε έστω κι ένα προηγούμενο αγγείο που δεν ήταν από την Αττική, ή απλά δεν ξέραμε (ίσως από ανθρώπινη παράλειψη κατά τη δημιουργία της βάσης γνώσης) από πού προέρχεται, τότε ο αλγόριθμος δε θα μπορούσε να δημιουργήσει τη σύγκριση. Ωστόσο μια σύγκριση:

> ***Όπως τα περισσότερα αγγεία που είδατε νωρίτερα, αυτό το άγαλμα προέρχεται από την Αττική.***

θα ήταν επιτυχημένη.

Σε αυτό το τμήμα της εργασίας δεν αλλάξαμε τον ίδιο τον αλγόριθμο Μελέγκογλου, αλλά τον τρόπο που παράγεται η δυναμική ταξινομία, πάνω στην οποία βασίζεται ο αλγόριθμος. Συγκεκριμένα, φτιάχνουμε μια νέα ταξινομία, που αποτελεί αντίγραφο της αρχικής (περιλαμβάνει,

δηλαδή, όλες τις οντότητες τις οποίες έχουμε περιγράψει μέχρι εκείνη τη στιγμή). Στη συνέχεια, για κάθε ιδιότητα, σε κάθε κόμβο της ταξινομίας, κρατάμε μόνο εκείνο το χαρακτηριστικό που εμφανίζεται τουλάχιστον στο 80% των οντοτήτων που ανήκουν στο συγκεκριμένο κόμβο, ενημερώνοντας (μειώνοντας) παράλληλα το συνολικό αριθμό οντοτήτων του κόμβου. Για παράδειγμα, αν ο κόμβος «Αγγεία» ήταν:

Αγγεία [9]

	Τιμή	Πλήθος	Τιμή	Πλήθος	Τιμή	Πλήθος
Π	Αρχαϊκή	1	Κλασσική	8		
ΠΚ	Αμφίπολις	1	Αττική	6	Ρόδος	2
Τ	Μελαν.	1	Ερυθρ.	5		

στο αντίγραφο της ταξινομίας ο ίδιος κόμβος θα ήταν:

Αγγεία [2]

	Τιμή	Πλήθος
Π	Κλασσική	2
ΠΚ		
Τ		

Επίσης, επειδή ο αλγόριθμος Μελέγκογλου δίνει ιδιαίτερη βαρύτητα στο πλήθος των εκθεμάτων που υπάρχουν σε κάθε κόμβο, καθώς και στο πόσα εκθέματα έχουν το συγκεκριμένο χαρακτηριστικό, στο νέο δένδρο το πλήθος τους είναι 2 (αρκεί οποιοσδήποτε αριθμός μεγαλύτερος του 1 και να είναι κοινός για όλους, ώστε να λειτουργήσουν σωστά οι μηχανισμοί του αλγορίθμου στο δεύτερο στάδιο).

Το κατώφλι του 80% μπορεί να αλλάξει (ίσως είναι προτιμότερο να αλλάζει, ανάλογα με το πόσες οντότητες ανήκουν στο συγκεκριμένο τύπο). Στις δοκιμές μας, πάντως η τιμή 80% λειτουργούσε ικανοποιητικά, αφού εξασφαλίζει ότι για να παραχθεί μια τέτοια σύγκριση θα πρέπει στον κόμβο να υπάρχουν τουλάχιστον πέντε οντότητες (με λιγότερα από πέντε οντότητες, δεν μπορεί να επιτευχθεί το κατώφλι του 80%). Σημειώνουμε ότι η παραλλαγή αυτή του αλγορίθμου καλείται μόνο όταν ο αρχικός αλγόριθμος Μελέγκογλου δεν καταφέρει να επιστρέψει καμία σύγκριση.

3.3 Συγκρίσεις με ολόκληρη τη συλλογή

Όπως έχουμε αναφέρει, ο αλγόριθμος Μελέγκογλου δημιουργεί συγκρίσεις ανάμεσα στην τρέχουσα οντότητα και στις προηγούμενες, που έχουμε ήδη περιγράψει. Όμως δεν μπορεί να δημιουργήσει συγκρίσεις με το σύνολο των οντοτήτων της οντολογίας. Για αυτό, συμπληρωματικά, δημιουργήσαμε ένα δεύτερο αλγόριθμο, που εκτελείται ανεξάρτητα. Σκοπός του είναι να εντοπίσει εκείνα τα χαρακτηριστικά που κάνουν την περιγραφόμενη οντότητα μοναδική στην οντολογία, καθώς κι εκείνα τα χαρακτηριστικά της που είναι κοινά με όλες της οντότητες του τύπου της (ή ενός υπερ-τύπου της).

Για να το πετύχουμε αυτό, χρησιμοποιούμε πάλι μια ταξινομία, όπως στον αλγόριθμο Μελέγκογλου, στην οποία όμως υπάρχουν όλες οι οντότητες της οντολογίας. Αρχικά εντοπίζουμε τον ειδικότερο κόμβο (τύπο) της ταξινομίας στον οποίο ανήκει η περιγραφόμενη οντότητα και κρατάμε μόνο το μονοπάτι από την ρίζα της ταξινομίας μέχρι και το συγκεκριμένο κόμβο. Κατόπιν, ανάλογα με το είδος της σύγκρισης που επιθυμούμε, κάνουμε ένα από τα παρακάτω:

Σύγκριση μοναδικότητας: Αφαιρούμε από κάθε κόμβο του μονοπατιού όλα εκείνα τα χαρακτηριστικά που εμφανίζονται σε περισσότερες από μία οντότητες. Έπειτα ελέγχουμε αν κάποιο από τα εναπομείναντα χαρακτηριστικά το παρουσιάζει η οντότητα που περιγράφουμε. Αν το παρουσιάζει, τότε το κρατάμε (μαζί με τον αντίστοιχο κόμβο του μονοπατιού) ως υποψήφια σύγκριση μοναδικότητας (π.χ. «Αυτή είναι η μόνη λήκυθος της συλλογής που ακολουθεί το μελανόμορφο ρυθμό.» ή «Αυτό είναι το μόνο αγγείο της συλλογής που ακολουθεί το μελανόμορφο ρυθμό.»). Το δεύτερο στοιχείο που ελέγχουμε είναι αν η περιγραφόμενη οντότητα είναι η μόνη οντότητα του τύπου της στην οντολογία (π.χ. «Αυτή είναι η μόνη υδρία της συλλογής»). Αν είναι, τότε δημιουργούμε μία ακόμη υποψήφια σύγκριση μοναδικότητας, που αναφέρεται στον τύπο της περιγραφόμενης οντότητας.

Σύγκριση καθολικής ομοιότητας: Για να είναι δυνατή μία τέτοια σύγκριση (π.χ. «Όπως όλα τα αγγεία της συλλογής, αυτή η λήκυθος κατασκευάστηκε κατά την κλασική περίοδο.») θα πρέπει όλες οι οντότητες που ανήκουν σε έναν κόμβο να παρουσιάζουν το ίδιο χαρακτηριστικό και ο κόμβος να έχει περισσότερες από μία οντότητες. Έτσι αφαιρούμε κόμβους του μονοπατιού που δεν περιέχουν τουλάχιστον μία οντότητα και σε κάθε κόμβο του μονοπατιού, αφαιρούμε όλα τα χαρακτηριστικά που δεν τα παρουσιάζουν όλες οι οντότητες του κόμβου. Εναλλακτικά, ιδιαίτερα σε μεγάλες συλλογές εκθεμάτων, δεν απαιτούμε να μοιράζονται όλες οι οντότητες του κόμβου το χαρακτηριστικό, αλλά η πλειονότητά τους (στα πειράματά μας, το ποσοστό αυτό ήταν 80%). Μία τέτοια σύγκριση (με πλειονότητα) επιλέγεται μόνο όταν δεν μπορεί να παραχθεί καμία σύγκριση καθολικής

ομοιότητας διαφορετικά (χωρίς πλειονότητα). Τέλος, στις συγκρίσεις καθολικής ομοιότητας δεν χρησιμοποιείται ποτέ ο ίδιος τύπος και ιδιότητα που έχει χρησιμοποιηθεί και σε προηγούμενη σύγκριση καθολικής ομοιότητας. Για παράδειγμα, αν αναφέρουμε σε ένα έκθεμα «Όπως όλα τα αγγεία της συλλογής, αυτή η λήκυθος κατασκευάστηκε κατά την κλασική περίοδο.» (μεταδώσαμε την πληροφορία ότι όλα τα αγγεία της συλλογής είναι της κλασικής περιόδου), δεν κάνουμε σε κανένα άλλο έκθεμα σύγκριση της μορφής «Όπως όλα τα αγγεία της συλλογής, αυτός/αυτή/αυτό ο/η/το/ ... κατασκευάστηκε κατά την κλασική περίοδο» (δεν ξαναμεταδίδουμε την ίδια πληροφορία).

Και σε αυτόν τον αλγόριθμο έχουμε τον περιορισμό ότι η σύγκριση πρέπει να γίνει ως προς μια ιδιαίτερα ενδιαφέρουσα ιδιότητα (που υπάρχει στη σχετική λίστα), όπως και στον αλγόριθμο Μελέγκογλου.

Αγγεία [9]

	Τιμή	Πλήθος	Τιμή	Πλήθος
Π	Αρχαϊκή	9		
ΠΚ	Αμφίπολις	1	Αττική	8
Τ	Μελαν.	4	Ερυθρ.	5

Στο παραπάνω παράδειγμα, αν θέλουμε να περιγράψουμε ένα μελανόμορφο παναθηναϊκό αμφορέα της αρχαϊκής περιόδου, ο οποίος προέρχεται από την Αμφίπολη, θα μπορούσαμε να παραγάγουμε τις παρακάτω συγκρίσεις:

Όπως όλα τα αγγεία της συλλογής, αυτός ο παναθηναϊκός αμφορέας κατασκευάστηκε κατά την αρχαϊκή περίοδο.

Είναι το μόνο αγγείο της συλλογής που προέρχεται από την Αμφίπολη.

Για την τεχνοτροπία ζωγραφικής δε θα μπορούσε να δημιουργηθεί σύγκριση, αφού δε φαίνεται να υπάρχει κάποιο κυρίαρχο χαρακτηριστικό.

Ακόμα, πρέπει να σημειώσουμε ότι κάποιες ιδιότητες δεν ορίζονται σε όλους τους τύπους. Για παράδειγμα, η τεχνοτροπία ζωγραφικής ορίζεται στα αγγεία (και τους υπο-τύπους τους) και δε συναντάται στα αγάλματα. Για αυτό δεν αφήνουμε τέτοιες ιδιότητες να «ανέβουν» ψηλότερα στην ταξινομία. Δε θα είχε δηλαδή νόημα να πούμε ότι:

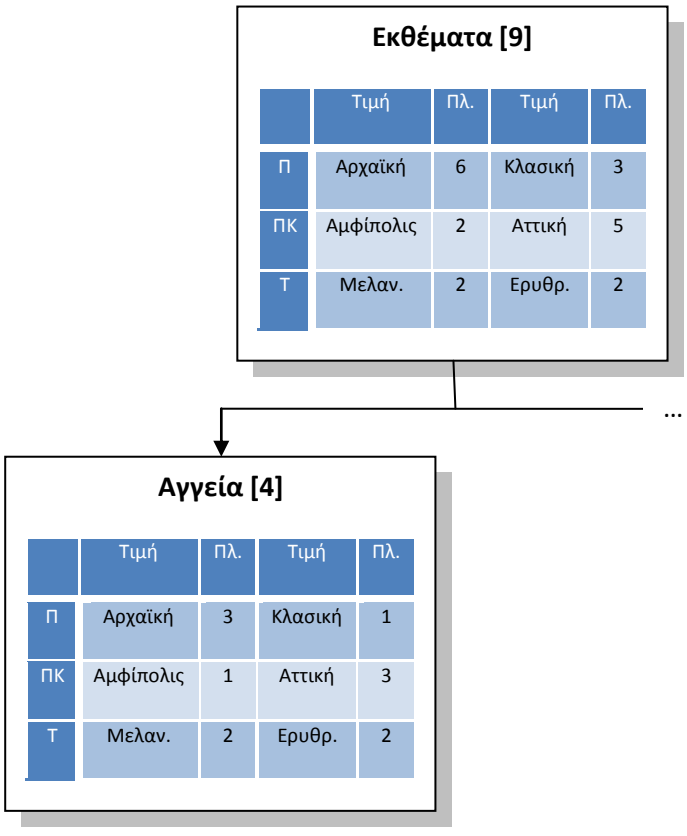
> ***Αυτό είναι το μόνο έκθεμα της συλλογής που ακολουθεί τον κορινθιακό ρυθμό***

αφού ο ρυθμός είναι χαρακτηριστικό μόνο των αγγείων κι όχι όλων των εκθεμάτων. Άρα, θα ήταν προτιμότερη μία σύγκριση όπως:

> ***Αυτό είναι το μόνο αγγείο της συλλογής που ακολουθεί τον κορινθιακό ρυθμό***

Το πεδίο ορισμού κάθε ιδιότητας δηλώνεται ρητώς στις οντολογίες OWL. Αντί να χρησιμοποιήσουμε αυτό τον ορισμό, όμως, προτιμάμε να χρησιμοποιήσουμε το «ενεργό πεδίο ορισμού» κάθε ιδιότητας για το μονοπάτι που εξετάζουμε. Για να εντοπίσουμε το ενεργό πεδίο ορισμού κάποιας ιδιότητας σε ένα μονοπάτι, αρκεί να βρούμε εκείνο τον κόμβο (τύπο) του μονοπατιού ο οποίος βρίσκεται ψηλότερα και ο οποίος διαθέτει έστω και μία παραπάνω οντότητα με αυτή την ιδιότητα από τον κόμβο-παιδί του.

Στο παρακάτω παράδειγμα, βλέπουμε ότι η μετάβαση από τον κόμβο *Αγγεία* στον κόμβο *Εκθέματα*, δεν επηρεάζει το πλήθος των οντοτήτων που έχουν την ιδιότητα *Τεχνοτροπία ζωγραφικής*. Άρα, το ενεργό πεδίο ορισμού της ιδιότητας αυτής είναι τα *Αγγεία* και θα πρέπει να γίνει σύγκριση ως προς τα αγγεία της συλλογής κι όχι ως προς όλα τα εκθεμάτά της.



Επίσης, επειδή υπάρχει το ενδεχόμενο κάποια ιδιότητα να εμφανίζει πολλές μοναδικές τιμές σε έναν τύπο οντοτήτων (π.χ. ένα μουσείο να έχει έναν πίνακα για κάθε μεγάλο ζωγράφο), με αποτέλεσμα η σύγκριση μοναδικότητας ως προς εκείνη την ιδιότητα και εκείνο τον τύπο να μην έχει ιδιαίτερη αξία (στο παράδειγμά μας, όλοι οι πίνακες είναι μοναδικοί ως προς το ζωγράφο), αποφασίσαμε να επιτρέπουμε σε μία ιδιότητα να χρησιμοποιηθεί σε συγκρίσεις μοναδικότητας ενός τύπου μόνο όταν έχει το πολύ δύο μοναδικές τιμές στις οντότητες εκείνου του τύπου.

"No amount of experimentation
can ever prove me right;
a single experiment can prove me wrong"
Albert Einstein

4. Πειραματική αξιολόγηση

4.1 Τι θέλουμε να αξιολογήσουμε

Πραγματοποιήσαμε μία σειρά πειραμάτων, για να αξιολογήσουμε την ποιότητα των συγκρίσεων που παράγει το σύστημά μας. Συγκεκριμένα, θέλαμε να δούμε αν οι αναγνώστες βρίσκουν τα κείμενα που περιέχουν συγκρίσεις πιο ελκυστικά από τα ίδια κείμενα χωρίς συγκρίσεις. Ένα κείμενο που είναι πιο ελκυστικό, ενδέχεται να είναι και πιο επιμορφωτικό, επειδή προσελκύει περισσότερο την προσοχή του αναγνώστη. Ωστόσο, λόγω περιορισμών χρόνου, δεν εξετάσαμε τις επιπτώσεις της χρήσης των συγκρίσεων στην συγκράτηση των πληροφοριών από τους αναγνώστες. Δηλαδή, δεν προσπαθήσαμε να διερευνήσουμε αν η χρήση των συγκρίσεων βοηθά τους χρήστες να θυμούνται περισσότερες από τις πληροφορίες που αναφέρουν τα κείμενα. Επίσης, δοκιμάσαμε να εντοπίσουμε αν κάποιο από τα είδη συγκρίσεων που παράγει το σύστημα θεωρείται καλύτερο από τους αναγνώστες.

4.2 Τρόπος εκτέλεσης των πειραμάτων

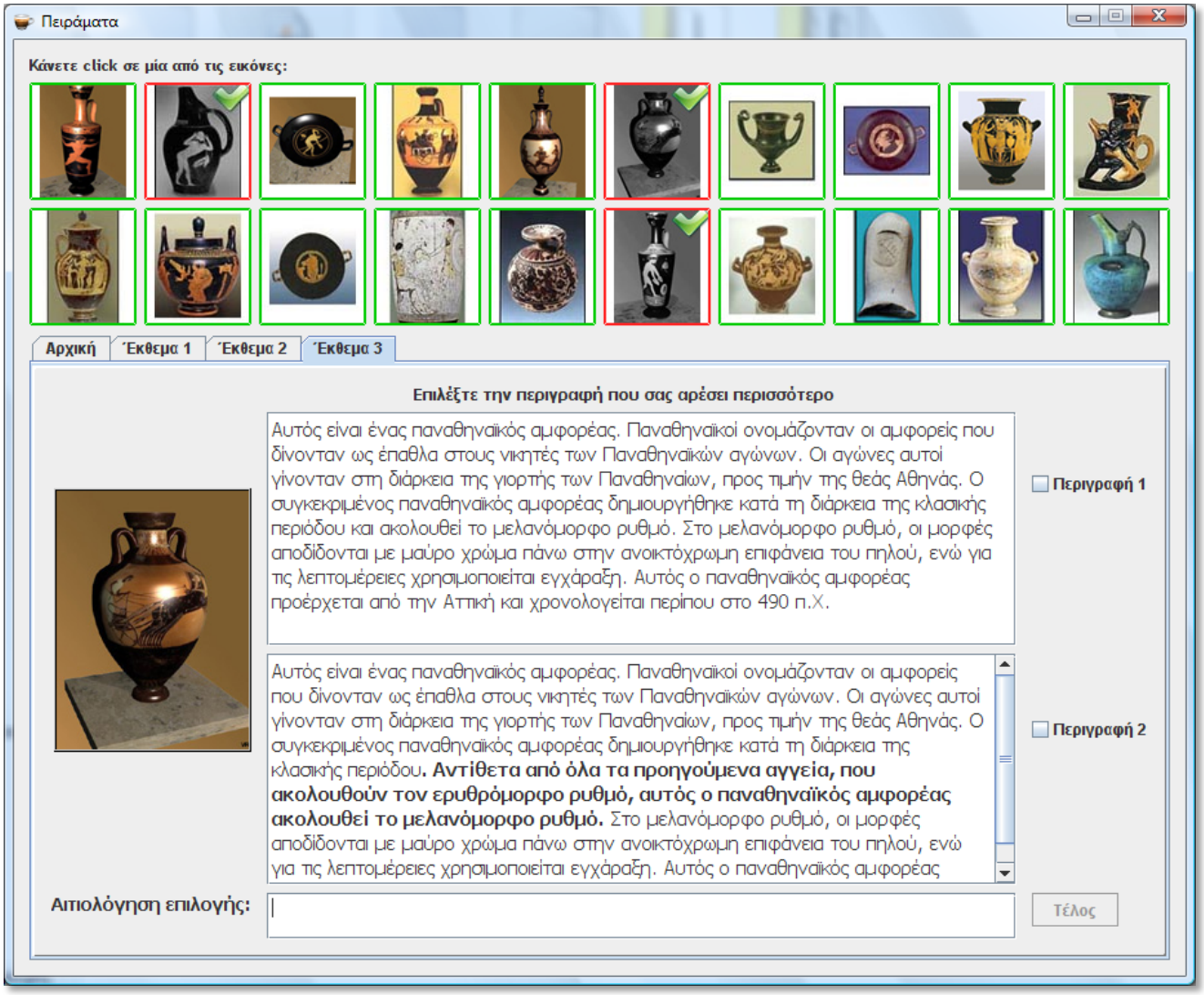
Για τις ανάγκες των πειραμάτων χρησιμοποιήσαμε τα εκθέματα της μουσειακής συλλογής του M-PIRO και την αντίστοιχη οντολογία (Isard κ.ά. 2003, Androutsopoulos κ.ά. 2007). Η συλλογή αυτή περιλαμβάνει τα ακόλουθα είδη εκθεμάτων:

Ομάδα	Πλήθος εκθεμάτων	Ομάδα	Πλήθος εκθεμάτων
Αγάλματα	8	**Νομίσματα**	6
Αγγεία	20	**Όστρακα**	1
Ανάγλυφα	8	**Πανοπλίες**	2
Διακοσμητικά	1	**Πίνακες**	1
Ειδώλια	2		

Χωρίσαμε τα εκθέματα της συλλογής σε τρεις κατηγορίες-αίθουσες, ώστε κάθε αίθουσα να περιέχει εκθέματα του ιδίου ή κατά το δυνατόν συναφών τύπων. Στην πρώτη αίθουσα υπάρχουν μόνο τα 20 αγγεία της οντολογίας (υψηλός βαθμός συνάφειας). Στη δεύτερη υπάρχουν τα αγάλματα, τα ανάγλυφα, τα ειδώλια, ο μοναδικός πίνακας και το μοναδικό όστρακο της συλλογής (συνολικά 20 εκθέματα, με μικρότερο βαθμό συνάφειας από την προηγούμενη αίθουσα). Τέλος, στην τρίτη αίθουσα βρίσκονται τα νομίσματα, οι πανοπλίες, το μοναδικό διακοσμητικό και (πάλι) ο μοναδικός πίνακας (συνολικά 10 εκθέματα, με ακόμα μικρότερο βαθμό ομοιότητας). Χωρίσαμε τα εκθέματα σε αίθουσες και θεωρήσαμε ότι οι επισκέπτες εξετάζουν τα εκθέματα ανά αίθουσα, ώστε να υπάρχουν περισσότερες ευκαιρίες για ενδιαφέρουσες συγκρίσεις. (Όταν τα εκθέματα που εξετάζει ο επισκέπτης είναι πολύ διαφορετικά, οι ευκαιρίες για ενδιαφέρουσες συγκρίσεις μειώνονται.) Ο διαφορετικός βαθμός συνάφειας των τριών αιθουσών μάς έδωσε την ευκαιρία να μελετήσουμε πώς επηρεάζεται η ποιότητα των παραγόμενων συγκρίσεων από τη συνάφεια των εκθεμάτων που εξετάζει ο χρήστης.

Για κάθε μία αίθουσα φτιάξαμε μία γραφική διεπαφή που επιτρέπει στο χρήστη να επιλέξει τα εκθέματα για τα οποία θέλει να διαβάσει κάποια περιγραφή. Στην κορυφή της εφαρμογής υπάρχουν εικονίδια των εκθεμάτων, τα οποία μπορεί ο χρήστης να επιλέξει με το ποντίκι. Η σειρά με την οποία εμφανίζονται τα εικονίδια είναι τυχαία και αλλάζει σε κάθε επισκέπτη. Κάθε έκθεμα (εικονίδιο) μπορεί να επιλεγεί μόνο μία φορά. Όταν επιλέξει ο χρήστης ένα εικονίδιο, ανοίγει στο κάτω μέρος της εφαρμογής μία νέα καρτέλα (tab) με μία μεγαλύτερη φωτογραφία του εκθέματος και δύο περιγραφές του: μία περιγραφή με συγκρίσεις και μία χωρίς. Σε κάθε έκθεμα, παράγεται το πολύ μία σύγκριση. Σε περίπτωση που δεν είναι δυνατή η παραγωγή καμίας σύγκρισης, εμφανίζεται μόνο μία περιγραφή, χωρίς συγκρίσεις. Ανά πάσα στιγμή ο χρήστης μπορεί να γυρίσει σε προηγούμενες καρτέλες και να ξαναδιαβάσει τις περιγραφές που έχει ήδη δει. Αυτό γίνεται για να μπορεί ο χρήστης να επιβεβαιώνει ότι οι πληροφορίες που του παρουσιάζει το σύστημα είναι σωστές (π.χ. σε περίπτωση που το σύστημα κάνει μία σύγκριση με κάποιο προηγούμενο έκθεμα,

να μπορεί να ελέγξει ότι όντως αυτή η σύγκριση ισχύει). Η σειρά με την οποία εμφανίζονται οι περιγραφές κάθε εκθέματος (αν θα εμφανιστεί, δηλαδή, πρώτα η περιγραφή με τις συγκρίσεις ή η περιγραφή χωρίς τις συγκρίσεις) είναι επίσης τυχαία. Όταν το σύστημα παράγει δύο περιγραφές για ένα έκθεμα (με και χωρίς συγκρίσεις), οι διαφορές των δύο κειμένων φαίνονται με έντονα γράμματα, ώστε να τις εντοπίζει ευκολότερα ο αναγνώστης. Ο αναγνώστης καλείται να επιλέξει εκείνη που του άρεσε περισσότερο. Επίσης, στο κάτω μέρος του παραθύρου υπάρχει χώρος για να τεκμηριώσει την επιλογή του.



Ο κάθε συμμετέχοντας στο πείραμα έπρεπε να επιλέξει ακριβώς 10 εκθέματα της αίθουσας (με όποια σειρά ήθελε) και να κρίνει τις περιγραφές. Στο τέλος, αφού είχε επιλέξει και τα 10, μπορούσε να γράψει συνολικά την άποψή του σχετικά με τις παραχθείσες συγκρίσεις.

Στο πείραμα πήραν μέρος προπτυχιακοί φοιτητές του τμήματος Πληροφορικής του ΟΠΑ. Κάθε φοιτητής εκτελούσε το πείραμα για μία αίθουσα (ο πρώτος φοιτητής έβλεπε τα εκθέματα της πρώτης αίθουσας, ο δεύτερος της δεύτερης, ο τρίτος της τρίτης, ο τέταρτος πάλι της πρώτης κ.ο.κ.)

δύο φορές. Την πρώτη φορά υπήρχαν κείμενα χωρίς συγκρίσεις και κείμενα που χρησιμοποιούσαν μόνο τον αλγόριθμο Μελέγκογλου (χωρίς τη γενίκευσή του, που συγκρίνει με όλα τα εκθέματα της οντολογίας). Τη δεύτερη φορά, χρησιμοποιούσαμε μόνο το συμπληρωματικό αλγόριθμο, που συγκρίνει με όλα τα εκθέματα της οντολογίας (ανεξαρτήτως του αν τα έχει δει ή όχι ο επισκέπτης) και παράγει συγκρίσεις μοναδικότητας και καθολικής ομοιότητας, αλλά λαμβάναμε υπόψη μας μόνο τα εκθέματα της συγκεκριμένης αίθουσας και δεν επιτρέπαμε, χάριν απλότητας, συγκρίσεις πλειονότητας. Σε εκθέματα όπου ήταν δυνατόν να παραχθεί τόσο σύγκριση μοναδικότητας όσο και καθολικής ομοιότητας, προτιμούσαμε τη δεύτερη, πάλι για λόγους ποικιλίας (οι ευκαιρίες παραγωγής συγκρίσεων καθολικής ομοιότητας είναι λιγότερες).

Η σειρά με την οποία χρησιμοποιούσε ο κάθε φοιτητής τους δύο αλγορίθμους (Μελέγκογλου και συγκρίσεων με όλα τα εκθέματα της αίθουσας) ήταν τυχαία. Ο τύπος του χρήστη που χρησιμοποιήσαμε (μεταξύ αυτών που ορίζει η οντολογία του M-PIRO) κατά τη διάρκεια των πειραμάτων ήταν ο ίδιος: ενήλικας Έλληνας. Σε κάθε περίοδο λόγου επιτρέπαμε να υπάρχουν μέχρι και δύο προτάσεις.

4.3 Πειραματικά αποτελέσματα

4.3.1 Συγκρίσεις με όλα τα εκθέματα της αίθουσας

Στο συγκεκριμένο πείραμα πήραν μέρος ανά αίθουσα:

Αίθουσα 1	34
Αίθουσα 2	29
Αίθουσα 3	24
Συνολικά	**87**

Στον παρακάτω πίνακα και το αντίστοιχο διάγραμμα υπάρχουν τα αποτελέσματα για τις συγκρίσεις:

- Καθολικότητας

  Όπως όλοι οι κύλικες της αίθουσας, αυτός ο κύλικας ακολουθεί τον ερυθρόμορφο ρυθμό.

- Μοναδικής οντότητας τύπου

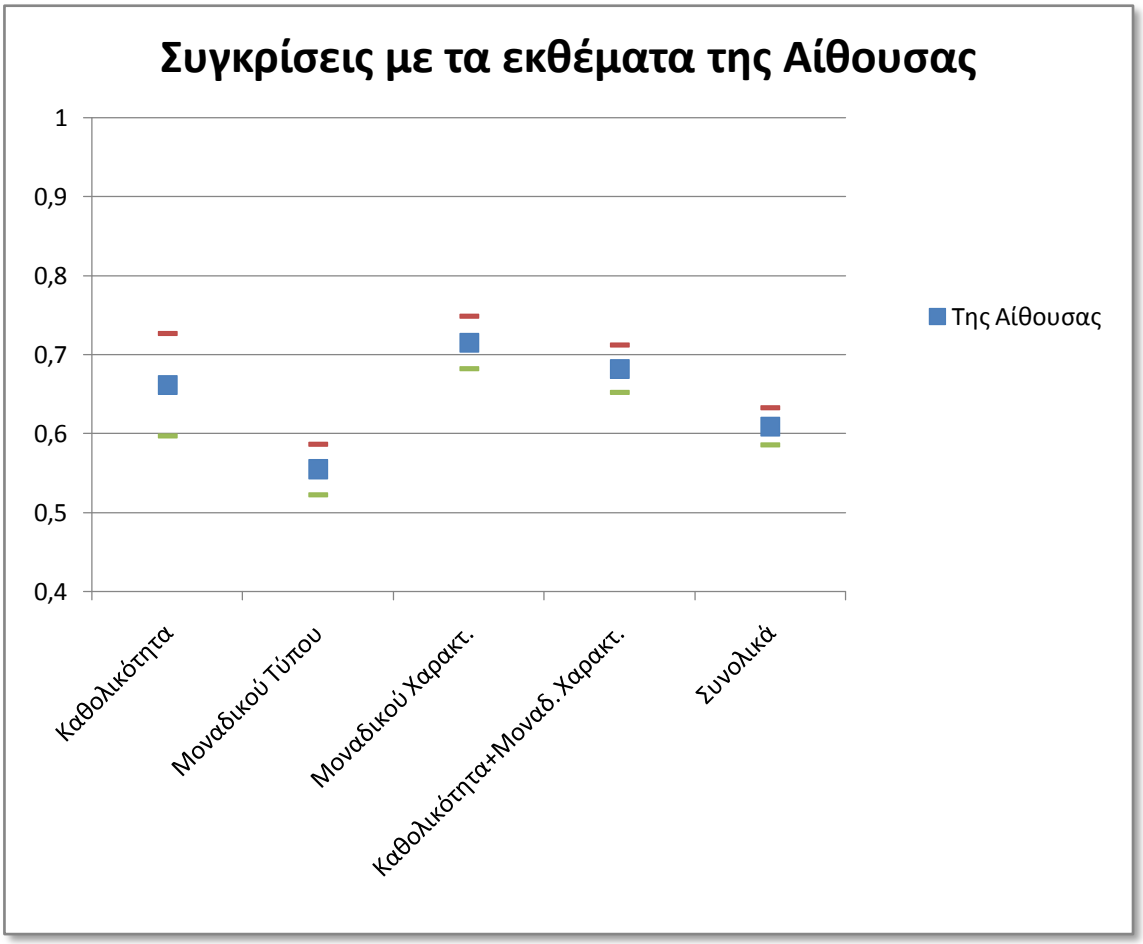
  Αυτή είναι η μόνη λευκή λήκυθος της αίθουσας.

- Μοναδικού χαρακτηριστικού

  Αυτή η πρόχους είναι το μόνο έκθεμα της αίθουσας που έχει φτιαχτεί από ορείχαλκο.

Τα διαστήματα εμπιστοσύνης είναι για βαθμό βεβαιότητας 95%.

Τύπος Συγκρίσεων	Καθολικότητα	Μοναδικού τύπου	Μοναδικού χαρακτ.	Μοναδικού χαρακτ. και καθολικότητας	Συνολικά
Ποσοστό Αποδοχής	0.662 (±0.065)	0.554 (±0.032)	0.715 (±0.033)	0.682 (±0.03)	0.609 (±0.023)
Πλήθος συγκρίσεων	64	216	313	280	593

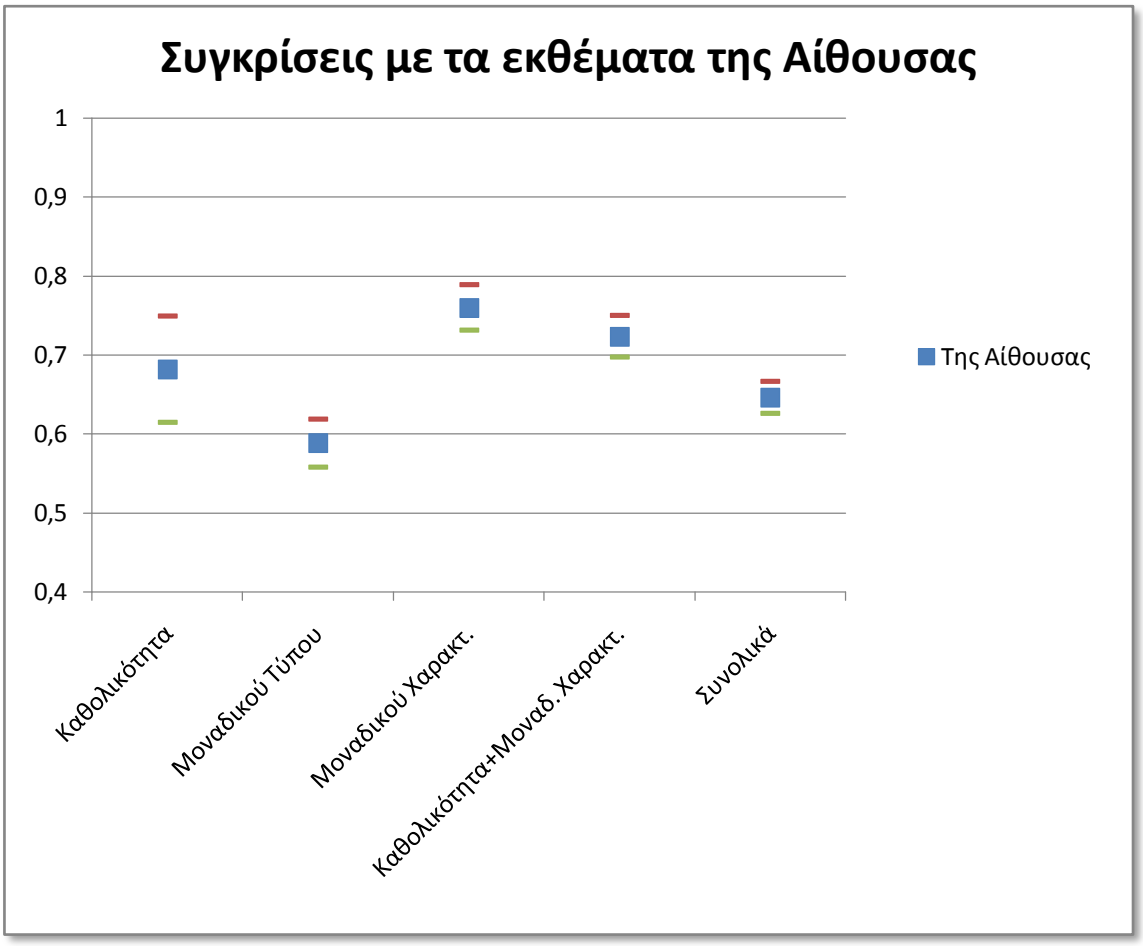


Παρατηρούμε ότι το μικρότερο ποσοστό αποδοχής έχουν οι συγκρίσεις μοναδικής οντότητας τύπου. Βάσει των σχολίων που έγραψαν στα πεδία τεκμηρίωσης, προκύπτει ότι οι χρήστες δεν θεωρούν πως η πληροφορία *δεν υπάρχει στην αίθουσα άλλο έκθεμα του ίδιου τύπου με αυτό που περιγράφεται* είναι άξια αναφοράς. (Εκτελέσαμε και μερικά πειράματα όπου συγκρίναμε με όλα τα εκθέματα της συλλογής, όχι μόνο της αίθουσας, αλλά τα αποτελέσματα ήταν παρόμοια.) Επίσης, σχετικά χαμηλό ποσοστό λαμβάνουν και οι συγκρίσεις καθολικότητας. Αντίθετα, οι συγκρίσεις που αναφέρονται στα μοναδικά χαρακτηριστικά του εκθέματος έχουν ποσοστό αποδοχής 71,5%.

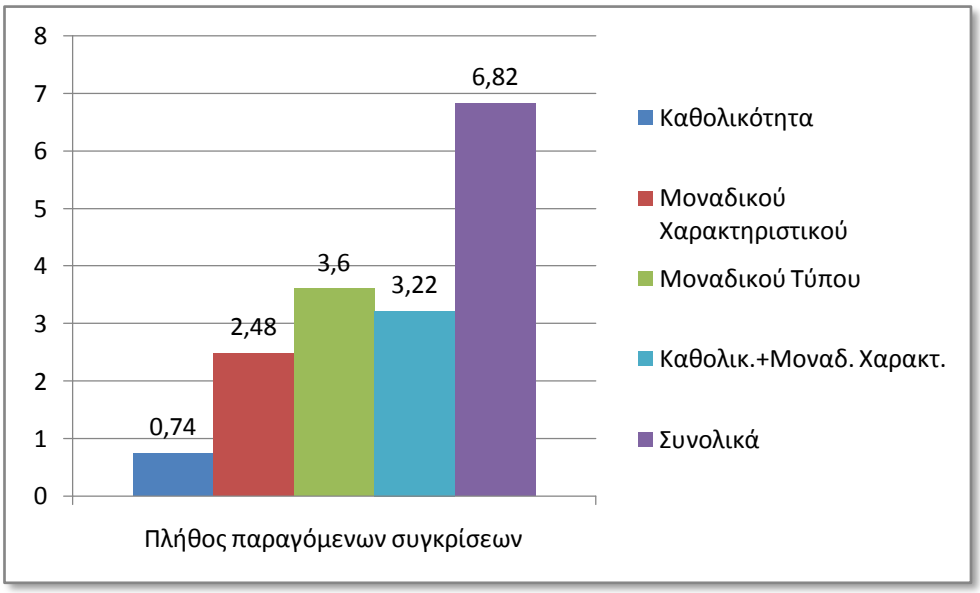
Από το σύνολο των 87 ατόμων που πήραν μέρος στο πείραμα, 5 ήταν εντελώς αρνητικοί στην ύπαρξη συγκρίσεων (σύμφωνα με τα σχόλια τους, αλλά και γιατί δεν επέλεξαν ποτέ περιγραφή με συγκρίσεις[7]). Αφαιρώντας αυτούς τους 5 χρήστες (θα ήταν εύκολο να τους επιτρέπεται να απενεργοποιήσουν την παραγωγή συγκρίσεων) τα ποσοστά αποδοχής γίνονται:

Τύπος Συγκρίσεων	Καθολικότητα	Μοναδικού τύπου	Μοναδικού χαρακτ.	Μοναδικού χαρακτ. και καθολικότητας	Συνολικά
Ποσοστό Αποδοχής	0.682 (±0.067)	0.589 (±0.03)	0.76 (±0.029)	0.724 (±0.026)	0.646 (±0.02)

[7] Για την ακρίβεια, υπήρξαν 7 χρήστες που δεν επέλεξαν καμία περιγραφή με συγκρίσεις, αλλά οι δύο από αυτούς δεν ανέφεραν ρητώς στα σχόλιά τους ότι δεν επιθυμούσαν καμία σύγκριση και/ή παραπονέθηκαν για τις συγκεκριμένες φράσεις των συγκρίσεων και όχι για την ίδια την ύπαρξη των συγκρίσεων.



Για κάθε τύπο σύγκρισης, ο μέσος αριθμός συγκρίσεων που παρήχθη ανά χρήστη ήταν:

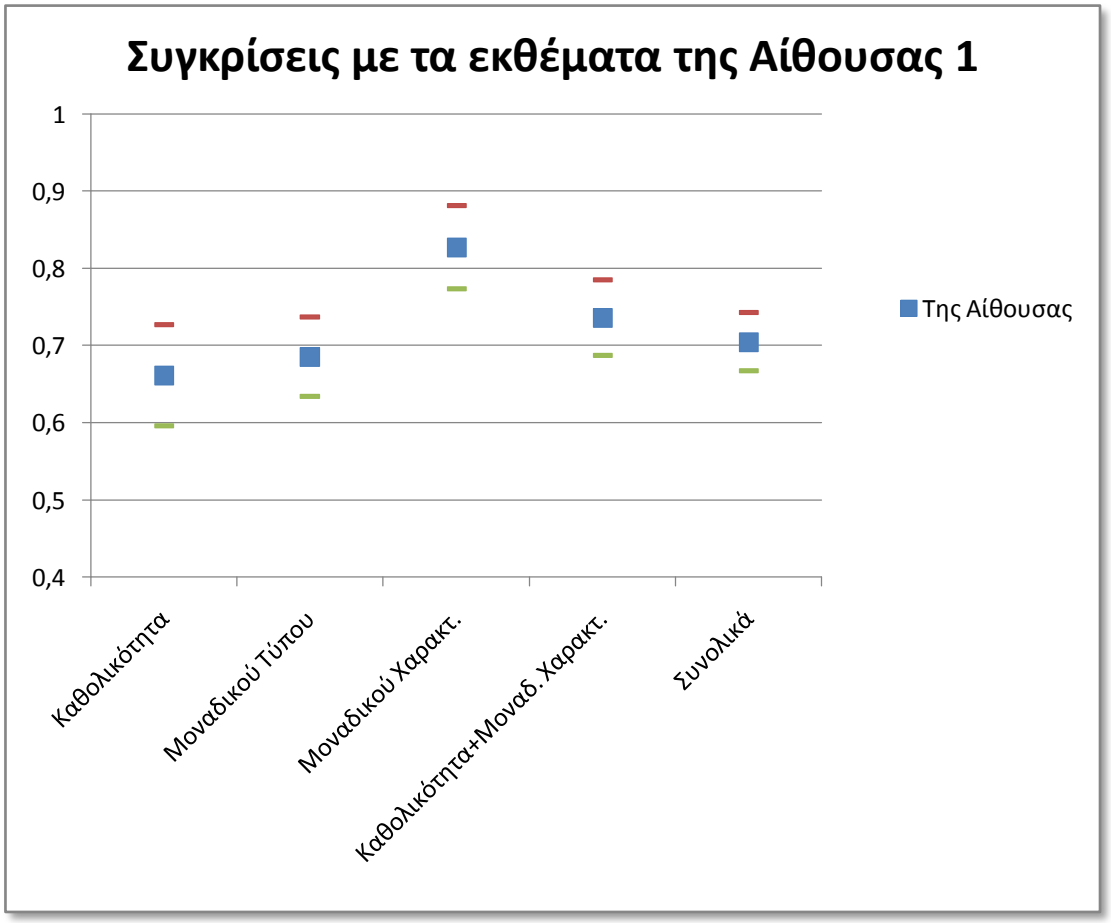


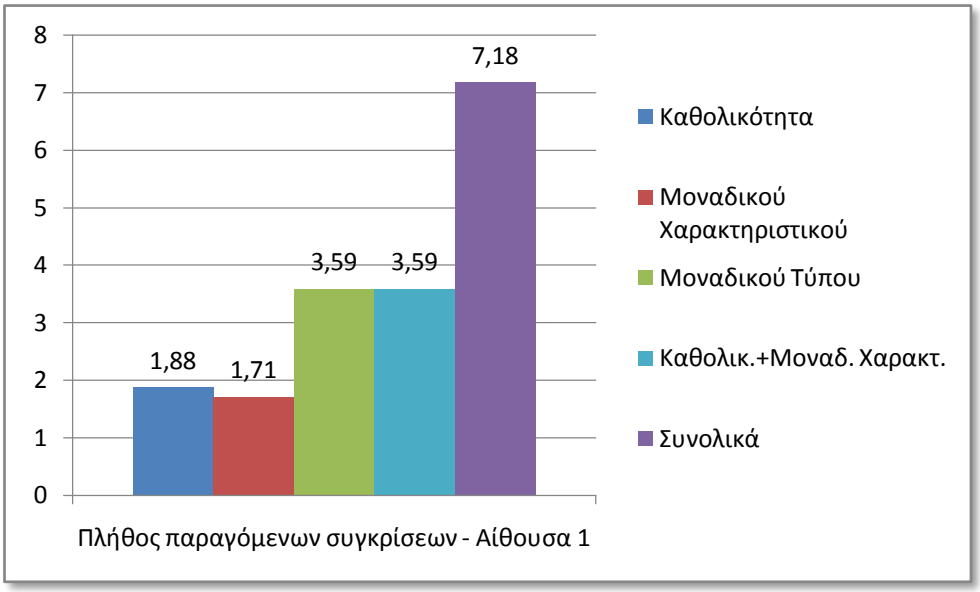
Με το συνδυασμό των συγκρίσεων Καθολικής Ομοιότητας και Μοναδικού Χαρακτηριστικού, επιτυγχάνουμε ένα σχετικά καλό ποσοστό αποδοχής (68,2% ±3%) με έναν αρκετά μεγάλο μέσο αριθμό συγκρίσεων ανά χρήστη (3,22). Αν εξαιρέσουμε τους χρήστες που δεν φαίνεται να επιθυμούν καθόλου συγκρίσεις, το ποσοστό αποδοχής γίνεται 72,4% (±2,6%).

Για κάθε αίθουσα ξεχωριστά, τα αποτελέσματα ήταν:

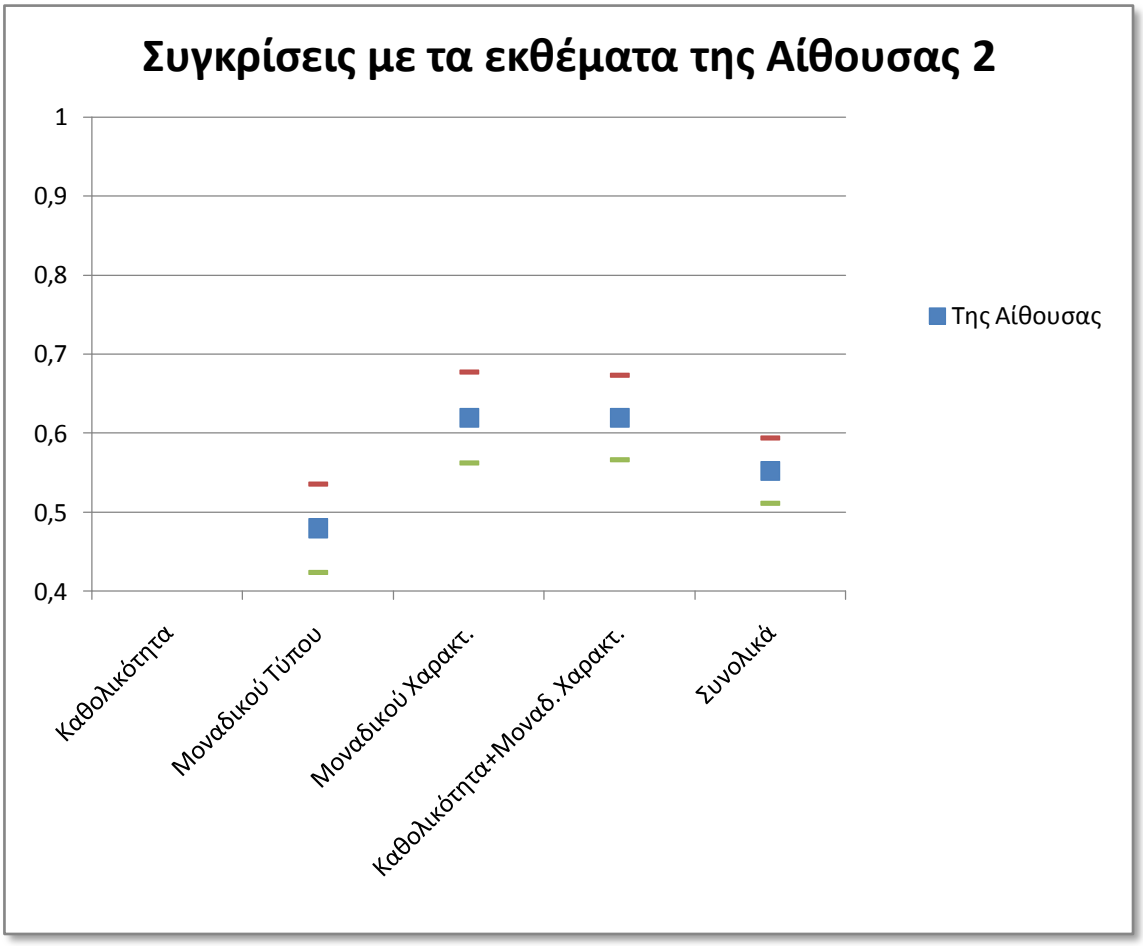
Τύπος Συγκρίσεων		Καθολικότητα	Μοναδικού τύπου	Μοναδικού χαρακτ.	Μοναδικού χαρακτ. και καθολικότητας	Συνολικά
Αίθουσα 1	Ποσοστό Αποδοχής	0.662 (±0.066)	0.686 (±0.051)	0.828 (±0.054)	0.737 (±0.049)	0.705 (±0.038)
	Πλήθος Συγκρίσεων	64	122	58	122	244
Αίθουσα 2	Ποσοστό Αποδοχής	Δεν παράχθηκε σύγκριση	0.479 (±0.056)	0.619 (±0.058)	0.619(±0.053)	0.552 (±0.041)
	Πλήθος Συγκρίσεων	0	71	86	86	157
Αίθουσα 3	Ποσοστό Αποδοχής	Δεν παράχθηκε σύγκριση	0.4562 (±0.065)	0.68 (±0.067)	0.68 (±0.062)	0.542 (±0.048)
	Πλήθος Συγκρίσεων	0	120	72	72	192

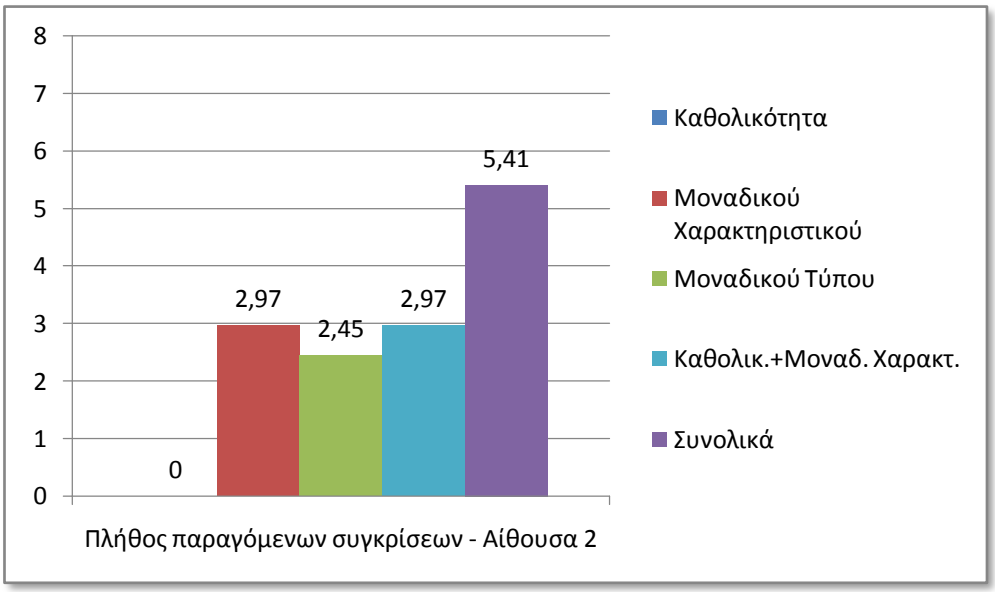
Τα διαγράμματα για την αίθουσα 1 (μεγάλος βαθμός συνάφειας):



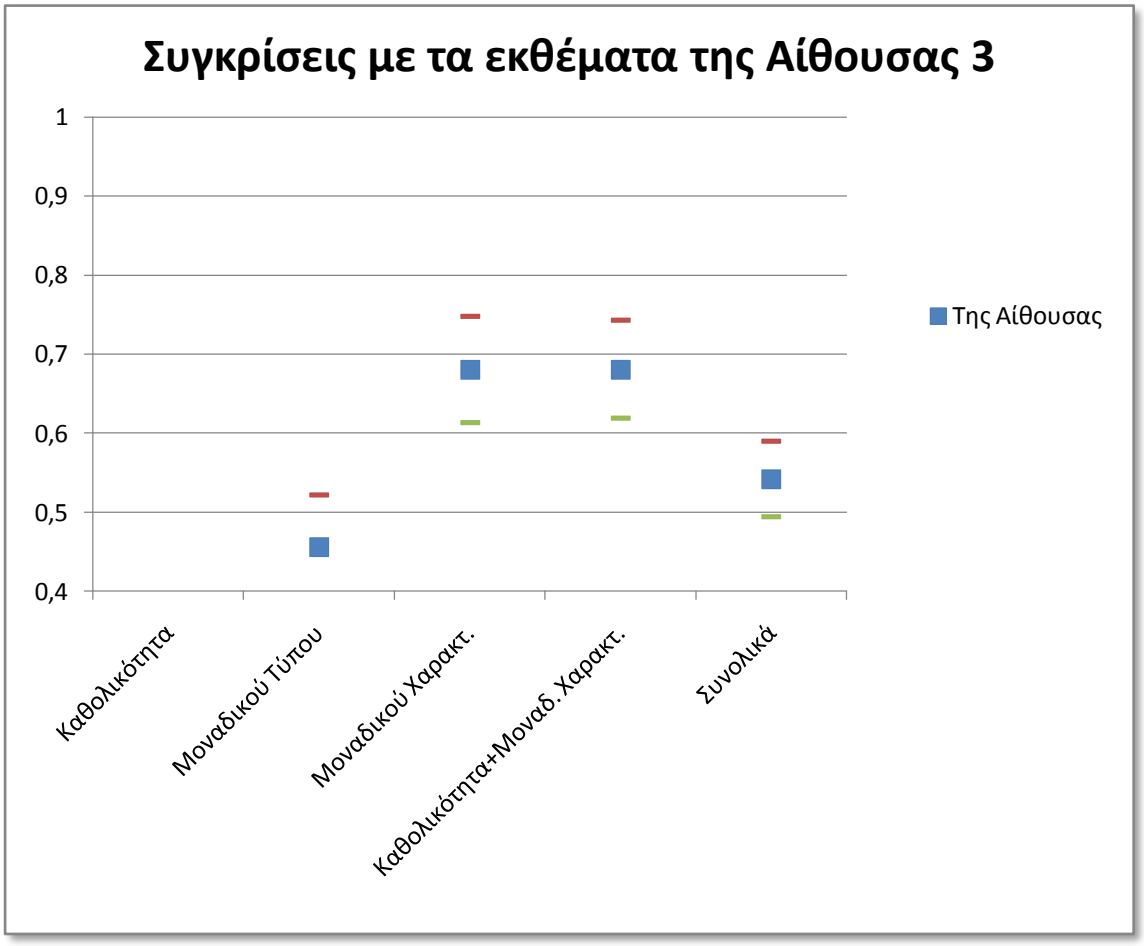


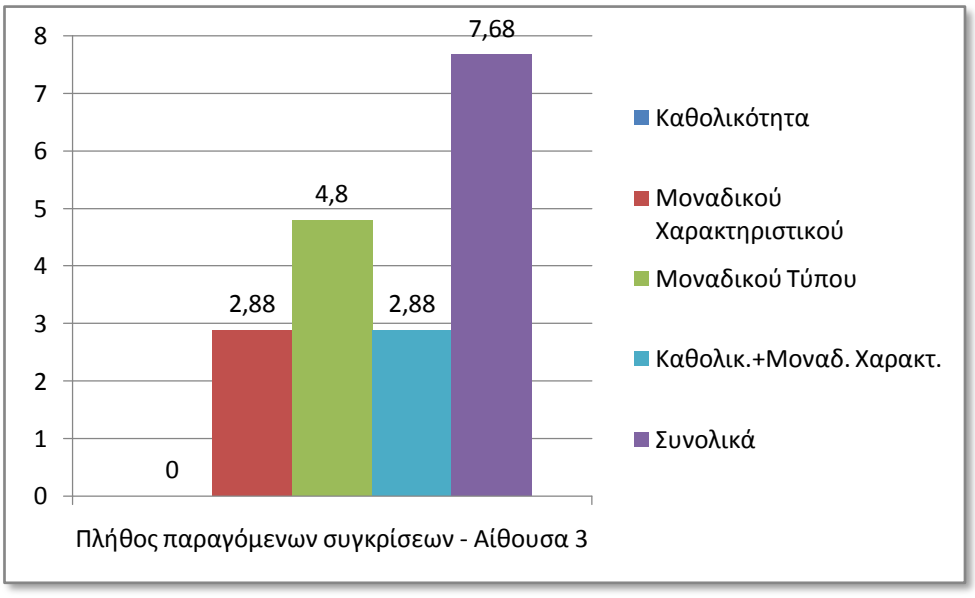
Τα διαγράμματα για την αίθουσα 2 (μέτριος βαθμός συνάφειας):





Τα διαγράμματα για την αίθουσα 3 (μικρός βαθμός συνάφειας):





Παρατηρούμε ότι η πρώτη αίθουσα με το μεγαλύτερο βαθμό συνάφειας έχει ιδιαίτερα υψηλά ποσοστά αποδοχής, ακόμα και στην περίπτωση των συγκρίσεων μοναδικού τύπου. Φαίνεται δηλαδή ότι σε μία αίθουσα στην οποία υπάρχουν αντικείμενα παρόμοιου τύπου, οι χρήστες αποδέχονται συγκρίσεις με τα υπόλοιπα αντικείμενα της αίθουσας. Αντίθετα στις αίθουσες στις

οποίες δεν υπήρχε μεγάλη συνάφεια των εκθεμάτων, αυτού του είδους οι συγκρίσεις απορρίφθηκαν. Ωστόσο πρέπει να σημειωθεί ότι το πλήθος των ερωτηθέντων για τις αίθουσες 2 και 3 δεν είναι ιδιαίτερα μεγάλο και άρα στατιστικά σημαντικό. Επίσης ένα ενδεχόμενο είναι να επηρεαστήκαν οι χρήστες και από το είδος των εκθεμάτων. Στην πρώτη αίθουσα υπήρχαν μόνο αγγεία, τα οποία ενδέχεται να είχαν μεγαλύτερο ενδιαφέρον από ό,τι τα αγάλματα. Το ίδιο ενδέχεται να συμβαίνει και στην αίθουσα 3 με τα νομίσματα (να ήταν πιο ενδιαφέροντα ως εκθέματα σε σχέση με τα αγάλματα).

Συγκρίσεις καθολικής ομοιότητας παράχθηκαν μόνο στην πρώτη αίθουσα. Αυτό είχε ως αποτέλεσμα την αύξηση του αριθμού των συγκρίσεων για τα μοναδικά χαρακτηριστικά (αφού οι συγκρίσεις καθολικότητας είχαν προτεραιότητα στην επιλογή) στις άλλες δύο αίθουσες. Όπως ήταν φυσιολογικό, όσο χαλαρώνει η συνάφεια των εκθεμάτων, αυξάνεται το πλήθος των συγκρίσεων μοναδικής οντότητας τύπου (αφού υπάρχουν περισσότερα διαφορετικά εκθέματα στην κάθε αίθουσα). Συνολικά μπορούμε να παρατηρήσουμε ότι σε μία συλλογή με υψηλό βαθμό συνάφειας των οντοτήτων της, οι συγκρίσεις προτιμώνται από τους χρηστές.

4.3.2 Συγκρίσεις με τον αλγόριθμο Μελέγκογλου

Σε αυτήν την μορφή του πειράματος πήραν μέρος:

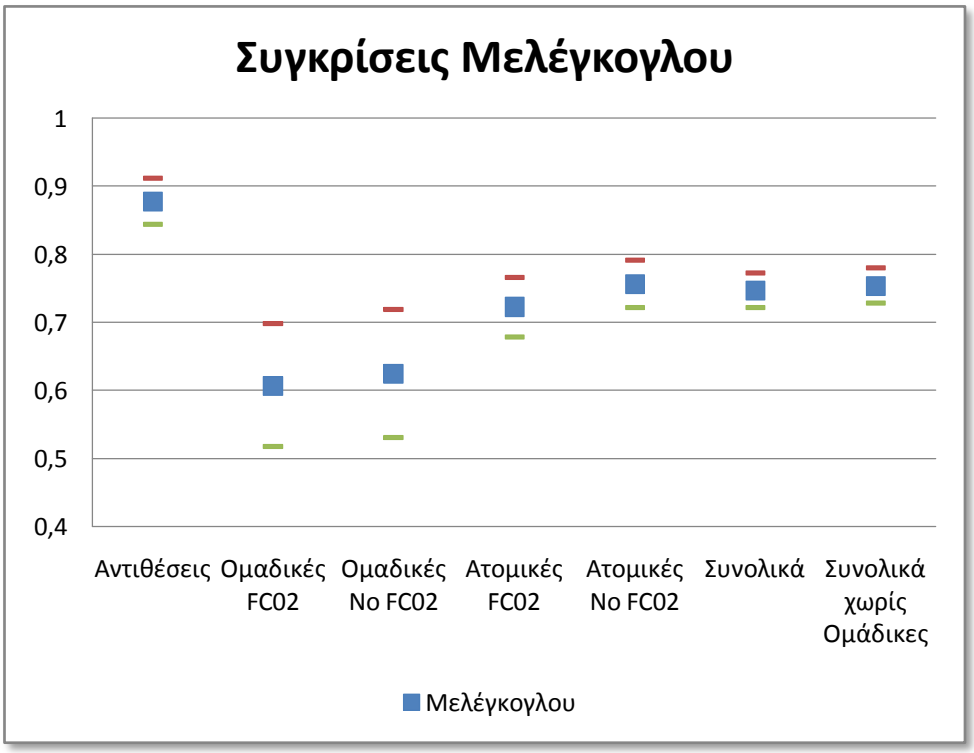
Αίθουσα 1	34
Αίθουσα 2	29
Αίθουσα 3	25
Συνολικά	**88**

Τα αποτελέσματα για κάθε είδος σύγκρισης ξεχωριστά, καθώς και διαφόρων συνδυασμών τους ήταν:

Τύπος Συγκρίσεων	**Συγκρίσεις που ικανοποιούν το 5ο κριτήριο του αλγορίθμου (FC02)**			
	Ομαδικές	Ατομικές	Μαζί	Αντιθέσεις
Ποσοστό Αποδοχής	0.607 (±0.091)	0.722 (±0.044)	0.714 (±0.039)	0.878 (±0.034)
Πλήθος συγκρίσεων	47	92	139	51

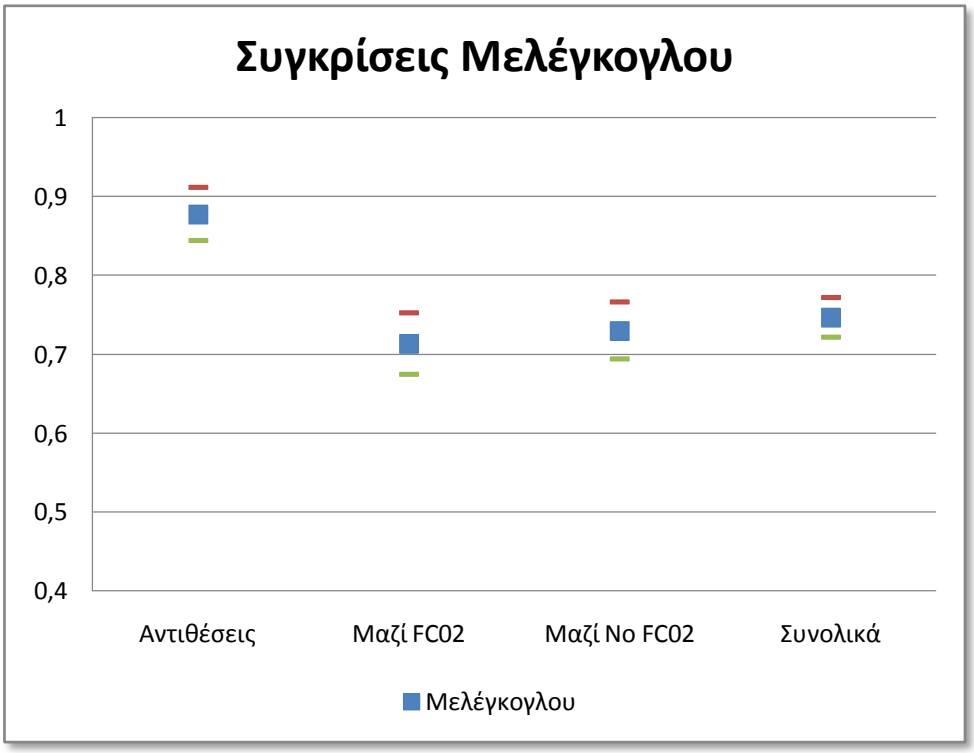
Τύπος Συγκρίσεων	Συγκρίσεις που δεν ικανοποιούν το 5ο κριτήριο του αλγορίθμου (No FC02)				
	Ομαδικές	Ατομικές	Μαζί	Συνολικά	Συνολικά Χωρίς Ομαδικές
Ποσοστό Αποδοχής	0.625 (±0.094)	0.757 (±0.035)	0.73 (±0.036)	0.747 (±0.025)	0.754 (±0.026)
Πλήθος Συγκρίσεων	28	132	160	350	275

Πρέπει να παρατηρήσουμε ότι οι ομαδικές συγκρίσεις ομοιότητας δεν εμφανίζονται συχνά (από τις 350 συγκρίσεις που παρήγαγε το σύστημα, μόνο 75 ήταν συγκρίσεις με ολόκληρη ομάδα εκθεμάτων) και για αυτό τα αντίστοιχα διαστήματα εμπιστοσύνης είναι ιδιαίτερα μεγάλα.



Είναι χαρακτηριστικό ότι οι χρήστες προτιμούν τις συγκρίσεις αντίθεσης. Όπως συνέβη και στο πείραμα με όλα τα εκθέματα της αίθουσας, το ποσοστό αποδοχής των συγκρίσεων που τονίζουν τη διαφορετικότητα της περιγραφόμενης οντότητας ήταν μεγαλύτερο από τα υπόλοιπα είδη συγκρίσεων. Ένα δεύτερο συμπέρασμα είναι ότι οι χρήστες δεν φαίνεται να εκτιμούν τις συγκρίσεις με ολόκληρες ομάδες εκθεμάτων. Αυτό ίσως να οφείλεται στο γεγονός ότι συχνά έχουν προηγηθεί (σε προηγούμενες περιγραφές) συγκρίσεις με μεμονωμένα εκθέματα (που αργότερα περιλαμβάνονται στις ομάδες με τις οποίες συγκρίνουμε) οι οποίες βασίζονται στην ίδια ιδιότητα.

Αυτό έχει ως αποτέλεσμα να διαβάζει ο επισκέπτης ομαδικές συγκρίσεις που βασίζονται σε ιδιότητες οι οποίες έχουν ήδη χρησιμοποιηθεί σε προηγούμενες (ατομικές) συγκρίσεις και τελικά να κουράζεται. Αυτό φάνηκε και από τα σχόλια των χρηστών, που κάποιες φορές παραπονέθηκαν για την επανάληψη συγκρίσεων.

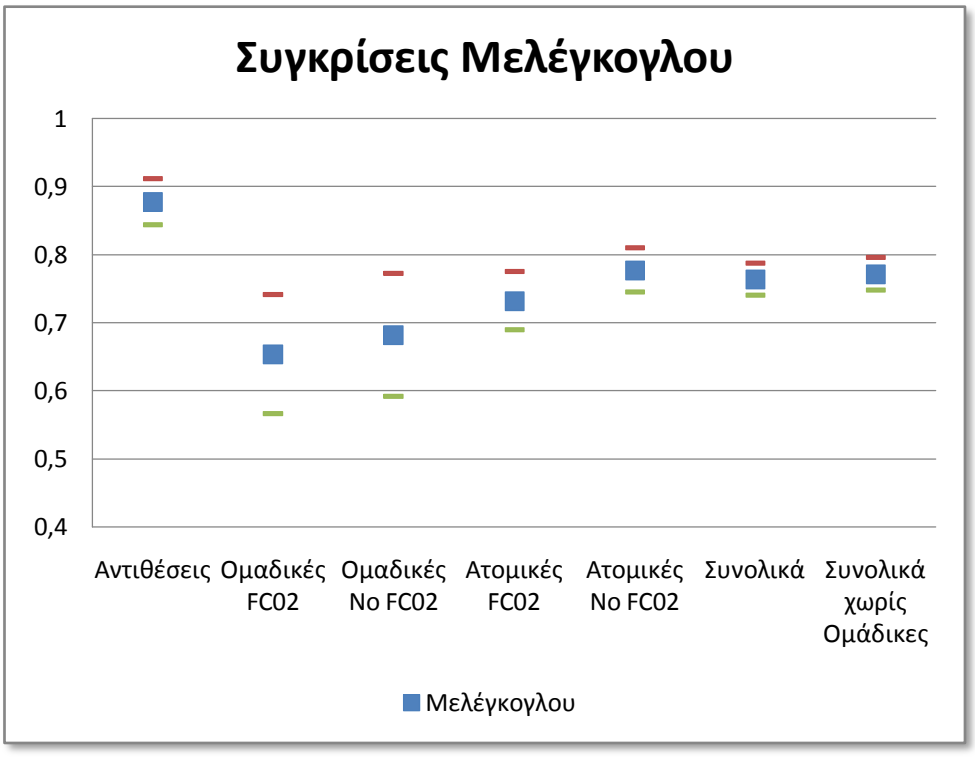


Τα πειραματικά αποτελέσματα δείχνουν, επίσης, ότι η χρήση του πέμπτου κριτηρίου (Filtering Criterion 2 – FC02) του αλγορίθμου Μελέγκογλου δεν βελτιώνει το ποσοστό αποδοχής των ομαδικών και ατομικών συγκρίσεων ομοιότητας (αντιθέτως, φαίνεται να οδηγεί σε ελαφρά χειρότερα αποτελέσματα, αν και η διαφορά δεν είναι στατιστικά σημαντική). Αυτό μάλλον συμβαίνει λόγω του τρόπου με τον οποίο επιλέγουμε την οντότητα με την οποία θα συγκρίνουμε, όταν δεν ικανοποιείται το κριτήριο. Αντί να διαλέξουμε κάποια τυχαία, επιλέγουμε εκείνη την οντότητα (ή την ομάδα της) που είδε ο επισκέπτης πιο πρόσφατα (αρκεί να είναι στις ανάμεσα στις τέσσερις τελευταίες) και άρα τη θυμάται πιο εύκολα. Θα είχε ενδιαφέρον να διερευνηθεί στο μέλλον κατά πόσον η παράλληλη χρήση του κριτηρίου της απόστασης και στις περιπτώσεις όπου ικανοποιείται το πέμπτο κριτήριο θα επιβεβαίωνε την αξία του κριτηρίου αυτού.

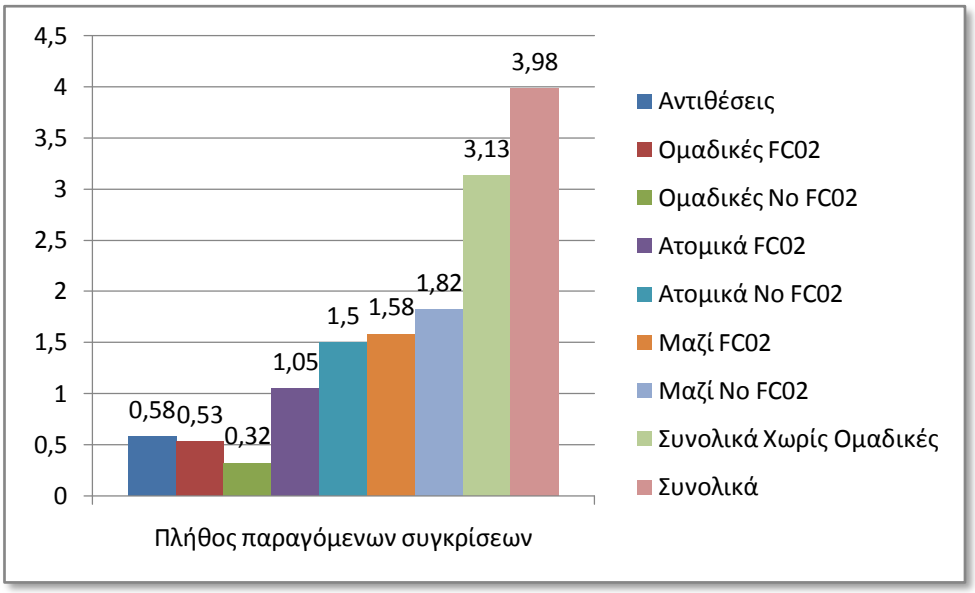
Στην περίπτωση του αλγορίθμου Μελέγκογλου, μόνο 2 άτομα φαίνεται να ήταν εντελώς αρνητικά στην ύπαρξη των συγκρίσεων (έγραψαν στα σχόλιά τους ότι οι συγκρίσεις τούς αποσπούσαν την προσοχή από το τρέχον έκθεμα και δεν επέλεξαν καμία περιγραφή με συγκρίσεις). Αφαιρώντας αυτούς τους χρήστες από τα στοιχεία μας, τα αποτελέσματα γίνονται:

Τύπος Συγκρίσεων	Συγκρίσεις που ικανοποιούν το 5ο κριτήριο του αλγορίθμου (FC02)			
	Ομαδικές	Ατομικές	Μαζί	Αντιθέσεις
Ποσοστό Αποδοχής	0.653 (±0.088)	0.732 (±0.043)	0.732 (±0.038)	0.878 (±0.034)

Τύπος Συγκρίσεων	Συγκρίσεις που δεν ικανοποιούν το 5ο κριτήριο του αλγορίθμου (No FC02)				
	Ομαδικές	Ατομικές	Μαζί	Συνολικά	Συνολικά Χωρίς Ομαδικές
Ποσοστό Αποδοχής	0.682 (±0.09)	0.777 (±0.033)	0.75 (±0.034)	0.764 (±0.023)	0.771 (±0.029)



Για κάθε είδος σύγκρισης, το πλήθος των συγκρίσεων που παρήγαγε το σύστημά μας ήταν:



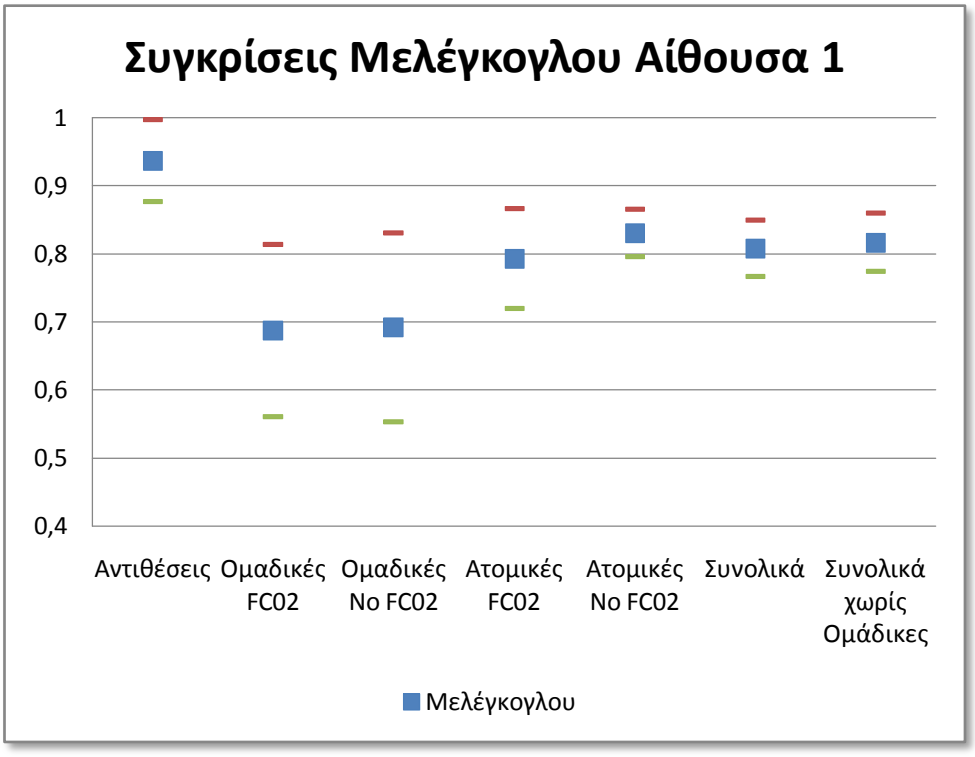
Κρατώντας μόνο τις συγκρίσεις αντίθεσης και ατομικής ομοιότητας, επιτυγχάνουμε ένα πολύ καλό ποσοστό αποδοχής (75,4% ±02,6%) με έναν αρκετά μεγάλο μέσο αριθμό συγκρίσεων ανά χρήστη (3,123). Αν εξαιρέσουμε τους χρήστες που δεν φαίνεται να επιθυμούν καθόλου συγκρίσεις, το ποσοστό αποδοχής γίνεται 77,2% (±02.4%).

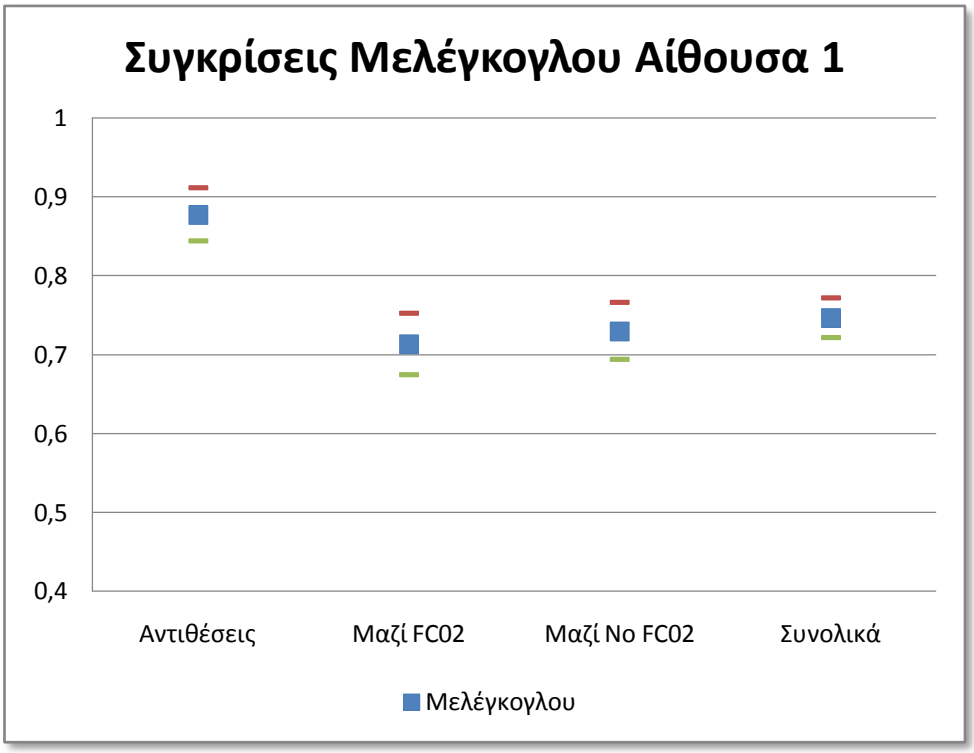
Για κάθε αίθουσα ξεχωριστά, τα αποτελέσματα ήταν:

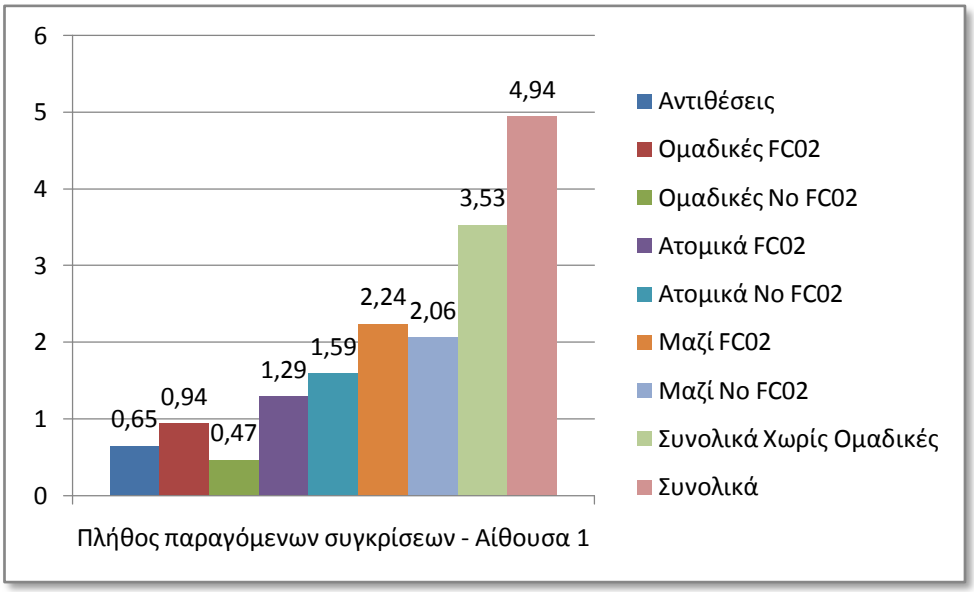
Τύπος Συγκρίσεων		**Συγκρίσεις που ικανοποιούν το 5ο κριτήριο του αλγορίθμου (FC02)**			
		Ομαδικές	Ατομικές	Μαζί	Αντιθέσεις
Αίθουσα 1	**Ποσοστό Αποδοχής**	0.687 (±0.126)	0.793 (±0.073)	0.78 (±0.063)	0.938 (±0.06)
	Πλήθος συγκρίσεων	32	44	76	22
Αίθουσα 2	**Ποσοστό Αποδοχής**	0.2 (±0.297)	0.591 (±0.088)	0.563 (±0.078)	1.0 (-0.070)
	Πλήθος συγκρίσεων	6	25	31	16
Αίθουσα 3	**Ποσοστό Αποδοχής**	0.714 (±0.16)	0.762 (±0.088)	0.783 (±0.078)	0.727 (±0.063)
	Πλήθος συγκρίσεων	9	23	32	13

Τύπος Συγκρίσεων		Συγκρίσεις που δεν ικανοποιούν το 5ο κριτήριο του αλγορίθμου (No FC02)				
		Ομαδικές	Ατομικές	Μαζί	Συνολικά	Συνολικά Χωρίς Ομαδικές
Αίθουσα 1	Ποσοστό Αποδοχής	0.692 (±0.139)	0.83 (±0.035)	0.78 (±0.06)	0.809 (±0.041)	0.817 (±0.043)
	Πλήθος συγκρίσεων	16	54	70	168	120
Αίθουσα 2	Ποσοστό Αποδοχής	0.571 (±0.216)	0.568 (±0.072)	0.56 (±0.072)	0.632 (±0.048)	0.649 (±0.049)
	Πλήθος συγκρίσεων	8	32	40	87	73
Αίθουσα 3	Ποσοστό Αποδοχής	0.5 (±0.357)	0.84 (±0.068)	0.833 (±0.072)	0.796 (±0.051)	0.79 (±0.052)
	Πλήθος συγκρίσεων	4	46	50	95	82

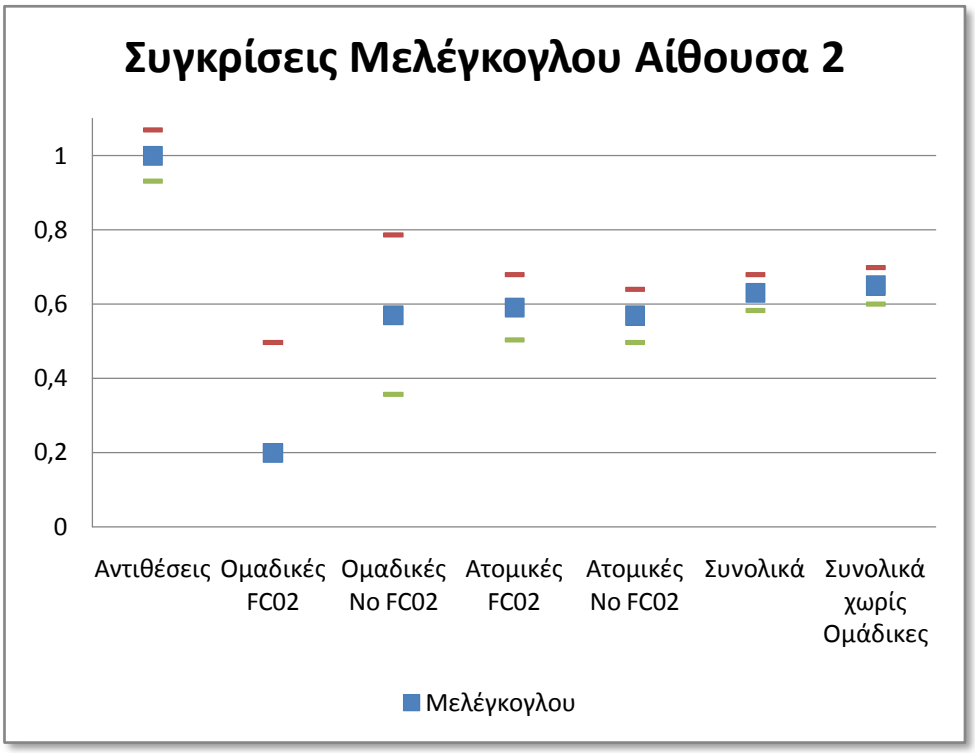
Τα διαγράμματα για την αίθουσα 1 (μεγάλος βαθμός συνάφειας):

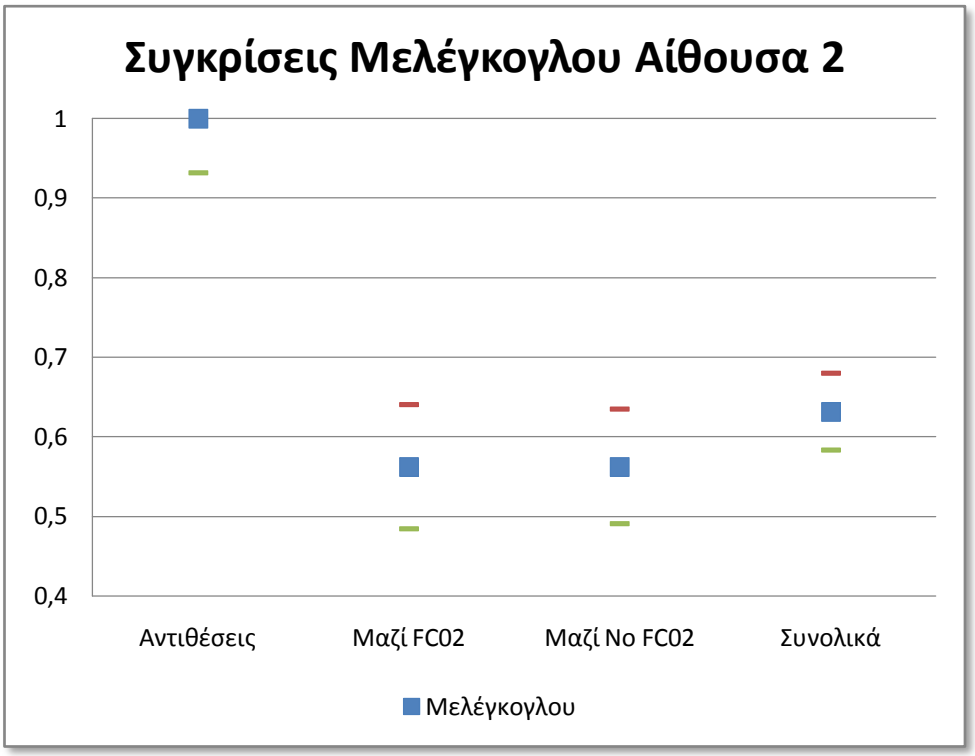


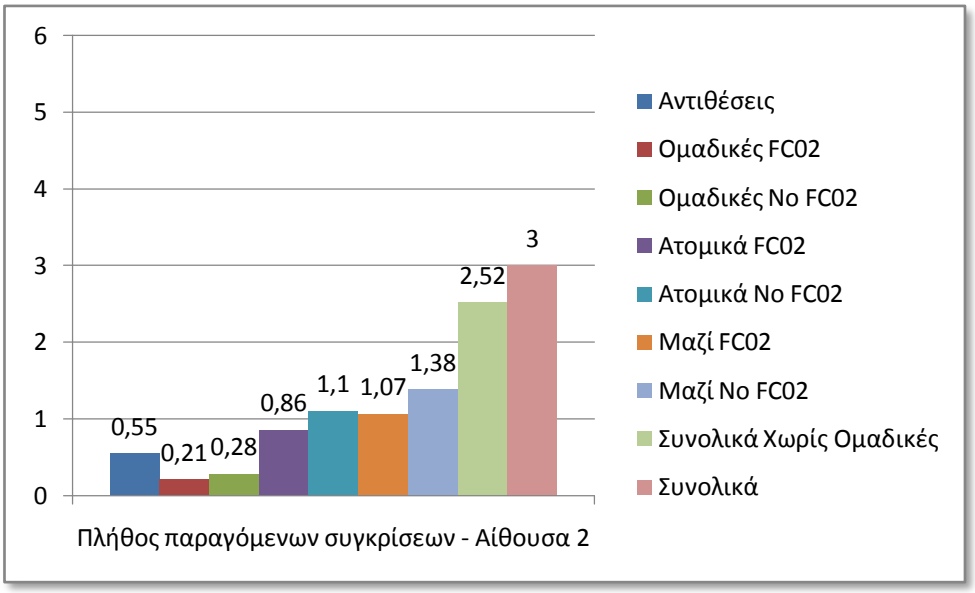

Συγκρίσεις Μελέγκογλου Αίθουσα 1
1
0,9
0,8
0,7
0,6
0,5
0,4
Αντιθέσεις
Μαζί FC02
Μαζί No FC02
Συνολικά
Μελέγκογλου


6
5
4
3
2
1
0
0,65
0,94
0,47
1,29
1,59
2,24
2,06
3,53
4,94
Πλήθος παραγόμενων συγκρίσεων - Αίθουσα 1
Αντιθέσεις
Ομαδικές FC02
Ομαδικές No FC02
Ατομικά FC02
Ατομικά No FC02
Μαζί FC02
Μαζί No FC02
Συνολικά Χωρίς Ομαδικές
Συνολικά

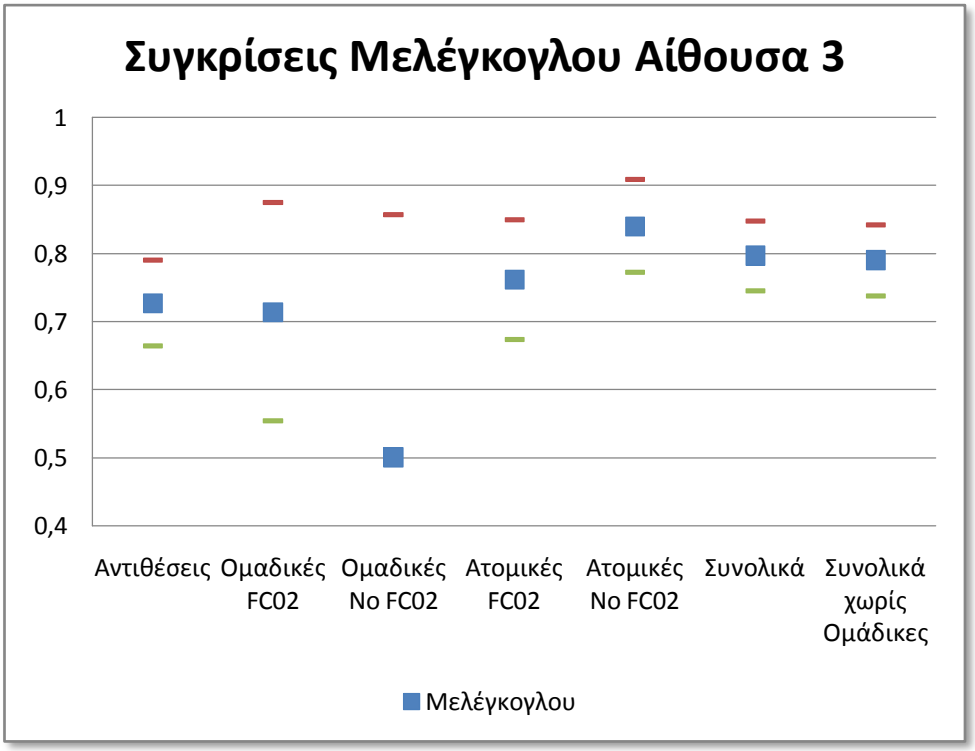
Τα διαγράμματα για την αίθουσα 2 (μέτριος βαθμός συνάφειας):

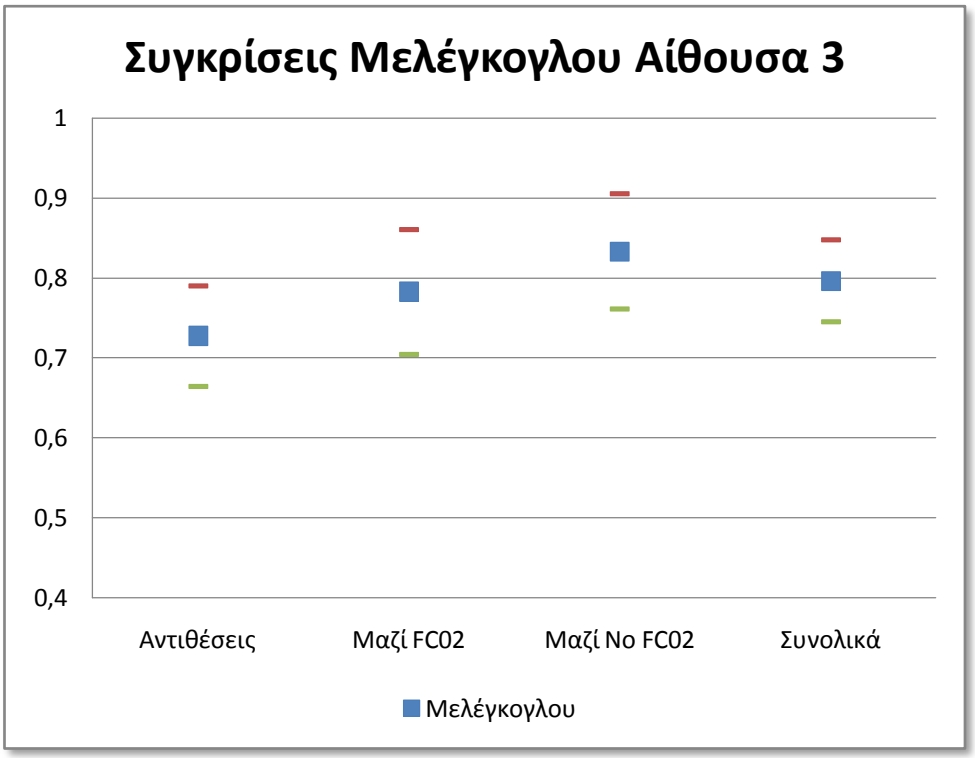


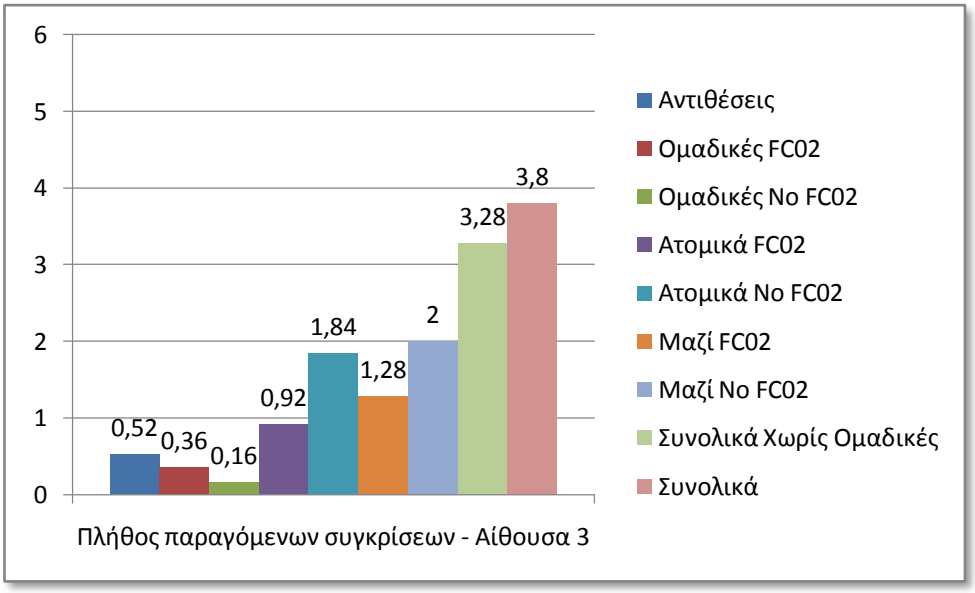




Τα διαγράμματα για την αίθουσα 3 (μικρός βαθμός συνάφειας):







Λόγω των μικρού δείγματος (και της σπανιότητας εμφάνισης κάποιων ειδών συγκρίσεων) τα διαστήματα εμπιστοσύνης είναι ιδιαίτερα μεγάλα. Εδώ παρατηρούμε κάτι σχετικά περίεργο: τα ποσοστά αποδοχής των συγκρίσεων είναι ιδιαίτερα υψηλά στην πρώτη αίθουσα (με το μεγάλο βαθμό συνάφειας των εκθεμάτων) και στην τρίτη αίθουσα (με το μικρότερο βαθμό συνάφειας). Ίσως αυτό να οφείλεται στο ότι τα εκθέματα των δύο αυτών αιθουσών ήταν πιο ενδιαφέροντα από της άλλης. Στην τρίτη αίθουσα βλέπουμε ιδιαίτερα χαμηλά ποσοστά στις συγκρίσεις αντίθεσης. Αυτό μάλλον οφείλεται στο γεγονός ότι δεν εμφανίστηκαν πολλές συγκρίσεις, αυτού του τύπου (μόνο 10).Επίσης πρέπει να σημειώσουμε ότι το πλήθος των συγκρίσεων αντίθεσης που παράγεται παραμένει σταθερό, ανεξάρτητα από το βαθμό συνάφειας της αίθουσας. Ωστόσο το πλήθος των

συγκρίσεων που ικανοποιούν το FC02 είναι αρκετά υψηλότερο στην αίθουσα με την μεγάλη συνάφεια. Αυτό είναι λογικό, αφού σε μία τέτοια αίθουσα υπάρχουν περισσότερες οντότητες που πληρούν το συγκεκριμένο κριτήριο. Το πλήθος των ομαδικών συγκρίσεων μειώνεται με τη μείωση του βαθμού συνάφειας, καθώς όλο και περισσότερες οντότητες είναι διαφορετικού τύπου (ενώ παράλληλα αυξάνεται το πλήθος των ατομικών συγκρίσεων). Γενικά πάντως παρατηρούμε (όπως και στο προηγούμενο πείραμα) ότι οι συγκρίσεις ενδείκνυνται σε συλλογές με αυξημένο βαθμό συνάφειας.

"I never think of the future, it comes soon enough"
Albert Einstein

5 Συμπεράσματα και μελλοντικές κατευθύνσεις

5.1 Συμπεράσματα

Τα αποτελέσματα των πειραμάτων έδειξαν ότι εν γένει οι αναγνώστες προτιμούν τις περιγραφές εκθεμάτων που περιέχουν τις συγκρίσεις που παράγει το σύστημά μας, έναντι των περιγραφών χωρίς συγκρίσεις. Έχοντας εξετάσει τρεις διαφορετικές ομάδες («αίθουσες») εκθεμάτων, με διαφορετικούς βαθμούς συνάφειας η κάθε μία, είδαμε ότι χρησιμοποιώντας έναν απλό αλγόριθμο αναζήτησης μοναδικών χαρακτηριστικών και χαρακτηριστικών που είναι κοινά για ολόκληρους τύπους οντοτήτων, το ποσοστό αποδοχής των συγκρίσεων ήταν 68,2% (±3%). Όταν αφαιρέσαμε από τα δεδομένα μας εκείνους τους χρήστες που δεν ήθελαν καθόλου συγκρίσεις, το ποσοστό αυξήθηκε στο 72,4% (±2,6%). Τα δε ποσοστά αποδοχής για τον αλγόριθμο Μελέγκογλου ήταν ακόμα υψηλότερα. Συγκεκριμένα έφτασαν το 75,4% (±2,6), ενώ αφαιρώντας από την ανάλυση όσους δεν ήθελαν καθόλου συγκρίσεις το ποσοστό γίνεται 77,1% (±2,9%). Σημειώνουμε ότι μόνο 5 και 2 χρήστες από τους 87 και 88, στα αντίστοιχα πειράματα με τους δύο αλγορίθμους, έγραψαν στα σχόλιά τους ότι θα προτιμούσαν να μην υπάρχουν καθόλου συγκρίσεις.

Επίσης τα πειράματα έδειξαν ότι οι συγκρίσεις που τονίζουν τη διαφορετικότητα ενός εκθέματος σε σχέση με τα υπόλοιπα, έχουν μεγαλύτερο ποσοστό αποδοχής από ό,τι οι συγκρίσεις ομοιότητας. Στην περίπτωση του πρώτου αλγορίθμου, το ποσοστό αποδοχής των συγκρίσεων διαφορετικότητας ήταν 76% (±2,9%), ενώ στον αλγόριθμο Μελέγκογλου ήταν 87,8% (±3,4%).

Τέλος, αναλύοντας τα επιμέρους αποτελέσματα για κάθε αίθουσα, παρατηρήσαμε ότι στην πρώτη αίθουσα, στην οποία όλα τα εκθέματα ήταν του ίδιου τύπου (αγγεία), τα ποσοστά αποδοχής των συγκρίσεων ήταν υψηλότερα από το μέσο όρο. Για τον αλγόριθμο καθολικής ομοιότητας και μοναδικότητας το ποσοστό είναι 73,7 (+5,5% από το μέσο όρο όλων των αιθουσών) και για τον αλγόριθμο Μελέγκογλου είναι 81,7% (+6,3% σε σχέση με το μέσο όρο).

Ο αλγόριθμος καθολικής ομοιότητας και μοναδικότητας είχε καλά αποτελέσματα και στην τρίτη αίθουσα, η οποία είχε τα εκθέματα με τη μεγαλύτερη ποικιλομορφία (τα ποσοστά αποδοχής

ταυτίζονταν με το μέσο όρο). Ωστόσο στη δεύτερη αίθουσα με μέτρια ποικιλία εκθεμάτων, ο αλγόριθμος απέδωσε πολύ χαμηλά, περίπου 61%. Το ίδιο φαινόμενο παρατηρήσαμε και κατά την αξιολόγηση του αλγορίθμου Μελέγκογλου: στην τρίτη αίθουσα αξιολογήθηκε περίπου με 79%, ενώ στην δεύτερη με μόλις 65%. Ενδέχεται η συγκεκριμένη αίθουσα να μην είχε τόσο ενδιαφέροντα εκθέματα σε σχέση με τις άλλες δύο, με αποτέλεσμα οι χρήστες να απορρίπτουν τις συγκρίσεις. Γενικά, όμως, και οι δύο αλγόριθμοι απέδωσαν καλύτερα σε αίθουσες με μεγάλο βαθμό συνάφειας των εκθεμάτων τους και σε αίθουσες με σχετικά ανόμοια εκθέματα.

5.2 Μελλοντικές επεκτάσεις

Ο αλγόριθμος Μελέγκογλου παράγει συγκρίσεις με προηγούμενα εκθέματα που έχει δει ο χρήστης, χωρίς να λαμβάνει υπόψη του ότι ο χρήστης ενδεχομένως δεν θυμάται κάποια από τα προηγούμενα εκθέματα ή χαρακτηριστικά τους. Σε μία πιθανή βελτίωση του αλγορίθμου, οι συγκρίσεις δεν θα γίνονταν με τις πραγματικές οντότητες του παρελθόντος, αλλά με γενικευμένες οντότητες, που θα είχαν τα κυρίαρχα χαρακτηριστικά ομάδων οντοτήτων που έχει δει ο επισκέπτης. Μία άλλη προσέγγιση θα ήταν με την πάροδο του χρόνου το σύστημα να «ξεχνά» ότι έχει περιγράψει στον επισκέπτη τις παλαιότερες οντότητες. Αντί να κρατάμε στο ιστορικό μας όλες τις οντότητες για τις οποίες παραγάγαμε κείμενο, καθώς περιγράφουμε νέες, να αφαιρούμε κάποιες από τις προηγούμενες. Κάτι τέτοιο ωστόσο θα πρέπει να γίνεται με αρκετή προσοχή, γιατί ενδέχεται να προκαλέσει σύγχυση σε κάποιον χρήστη με καλή μνήμη. Επίσης, οντότητες για τις οποίες έχει γίνει κάποια σύγκριση αντίθεσης ή μοναδικότητας θα πρέπει να παραμένουν περισσότερο στη «μνήμη», αφού συνήθως ο άνθρωπος εντοπίζει και θυμάται πιο εύκολα τις διαφορές, παρά τις ομοιότητες (κάτι που φάνηκε και από τα αποτελέσματα των πειραμάτων μας).

Ένα άλλο σημείο με το οποίο δεν ασχοληθήκαμε καθόλου είναι οι συγκρίσεις μέσα στο χώρο. Όταν ένας επισκέπτης βρίσκεται σε μία αίθουσα ενός πραγματικού μουσείου, μπορεί να δει πολλά από τα εκθέματα της αίθουσας γύρω του. Έχοντας αντίληψη του πού βρίσκεται θα μπορούσαμε να προκρίνουμε συγκρίσεις με γειτονικά εκθέματα. Για κάτι τέτοιο θα έπρεπε η μηχανή παραγωγής φυσικής γλώσσας να είναι σε θέση να παράγει χωρικές εκφράσεις (π.χ. «*σε αντίθεση με το έκθεμα που βρίσκεται πίσω σας*», «*στα αριστερά του εκθέματος που βλέπετε*»).

Τέλος, αναφερθήκαμε ήδη στην αξία που θα είχε η παράλληλη χρήση του κριτηρίου της απόστασης (πόσο πριν είχε περιγραφεί κάθε έκθεμα) και στην περίπτωση των συγκρίσεων ομοιότητας που ικανοποιούν το πέμπτο κριτήριο του αλγορίθμου Μελέγκογλου, προκειμένου να διερευνηθεί περαιτέρω η αξία αυτού του κριτηρίου. Επίσης, θα ήταν καλό να αποφεύγονται οι επανειλημμένες συγκρίσεις που βασίζονται στην ίδια ιδιότητα, ώστε να μη κουράζεται ο

επισκέπτης, κάτι που ενδέχεται να βελτιώσει το ποσοστό αποδοχής των ομαδικών συγκρίσεων. Για να επιτευχθεί αυτό θα πρέπει το σύστημα να θυμάται, εκτός από τις οντότητες που έχει περιγράψει, και τις συγκρίσεις που έχει χρησιμοποιήσει.

Αναφορές

Galanis, D., & Androutsopoulos, I. (2007). Generating Multilingual Descriptions from Linguistically Annotated OWL Ontologies: the NaturalOWL System. 11th European Workshop on Natural Language Generation (ENLG 2007). Schloss Dagstuhl, Germany.

Isard, A., Oberlander, J., Androutsopoulos, I. & Matheson, C. (2003). Speaking the User's Languages. IEEE Intelligent Systems . Volume: 18, Issue: 1, p. 40-45.

Milosavljevic, M. (1997a) Augmenting the user's knowledge via comparison. User Modeling, Proc of Sixth International Conference UM97, ed. by A Jameson, C Paris, and C Tasso, p. 119-130, Springer.

Milosavljevic, M. (1997b). Content Selection in Comparison Generation. 6th European Workshop on Natural Language Generation (EWNLG-97), Duisburg, Germany p. 72-81.

Milosavljevic, M. (1999). The Automatic Generation of Comparison in Description of Entities. Sydney NSW, Australia: PhD Thesis, Division of Information and Communication Sciences Macquarie University.

Milosavljevic, M., & Dale, R. (1996). Strategies for Comparison in Encyclopedia Descriptions. 8th International Workshop on Natural Language Generation, Herstmonceux, Sussex, UK, 12-15 June 1996. p. 161-170.

Reiter, E., & Dale, R. (2006). *Building Natural Language Generation Systems.* Cambridge University Pr*ess*.

Γαλάνης, Δ. (2006). *Ανάπτυξη μηχανής παραγωγής φυσικής γλώσσας για οντολογίες OWL.* Διπλωματική Εργασία Μεταπτυχιακού Διπλώματος Ειδίκευσης, Τμήμα Πληροφορικής, Οικονομικό Πανεπιστήμιο Αθηνών.