

ΠΜΣ «Προηγμένα Συστήματα Πληροφορικής»
Τμήμα Πληροφορικής, Πανεπιστήμιο Πειραιώς

Σημασιολογικά Κατευθυνόμενη Εξόρυξη Γνώσης με Χρήση
Οντολογιών

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΑΤΡΙΒΗ

Ανδρέας Δ. Πετράκης ΜΠΣΠ/09060

**Επιβλέπων: Γιάννης Θεοδωρίδης (Αναπλ. Καθηγητής),
Συνεπιβλέπων: Ευάγγελος Κοτσιφάκος (Διδάκτωρ)**

Information Systems Laboratory (InfoLab)

<http://infolab.cs.unipi.gr>



- Διαδικασία Ανακάλυψης Γνώσης από Δεδομένα (KDD)
 - Εξόρυξη ενδιαφέρουσας γνώσης (π.χ. αξιόπιστη, μη αναμενόμενη, χρήσιμη)
- Οντολογίες
 - Ο καλύτερος τρόπος αναπαράστασης της Γνώσης ενός Πεδίου
- Καινοτομία
 - Ενσωμάτωση της Γνώσης ενός Πεδίου στην KDD
 - Αντί της στατιστικής προσέγγισης
- Νέος τρόπος προσέγγισης;
 - Ανάπτυξη Σημασιολογικού Ιστού
 - Καθιέρωση Οντολογιών, ωρίμανση υποστηρικτικών εργαλείων

- Θεωρητικό Υπόβαθρο
 - Τρόπος Αναπαράστασης Γνώσης → Οντολογίες
 - Αξιοποίηση Οντολογιών → Μηχανές Συλλογιστικής Ανάλυσης
 - Αξιοποίηση Σημασιολογίας → Μετρικές Σημασιολ. Συνάφειας
- Ανάλυση, Σχεδίαση & Υλοποίηση
 - Οντολογία → OWL
 - Τεχνική → Κανόνες Συσχετίσεων
 - Αλγόριθμος → Weka Apriori
- Περίπτωση Χρήσης
 - Πεδίο Γνώσης → Τρομοκρατία
 - Ύπαρξη Κανόνων → Ειδικοί
 - Σύγκριση προσεγγίσεων
- Συμπεράσματα

- Τυπική, ρητή προδιαγραφή μίας σύλληψης
 - Τυπική: αναγνώσιμη από υπολογιστικές μηχανές
 - Ρητή: σαφώς καθορισμένες έννοιες, περιορισμοί
 - Σύλληψη: αφηρημένη/ απλοποιημένη άποψη μίας περιοχής ενδιαφέροντος

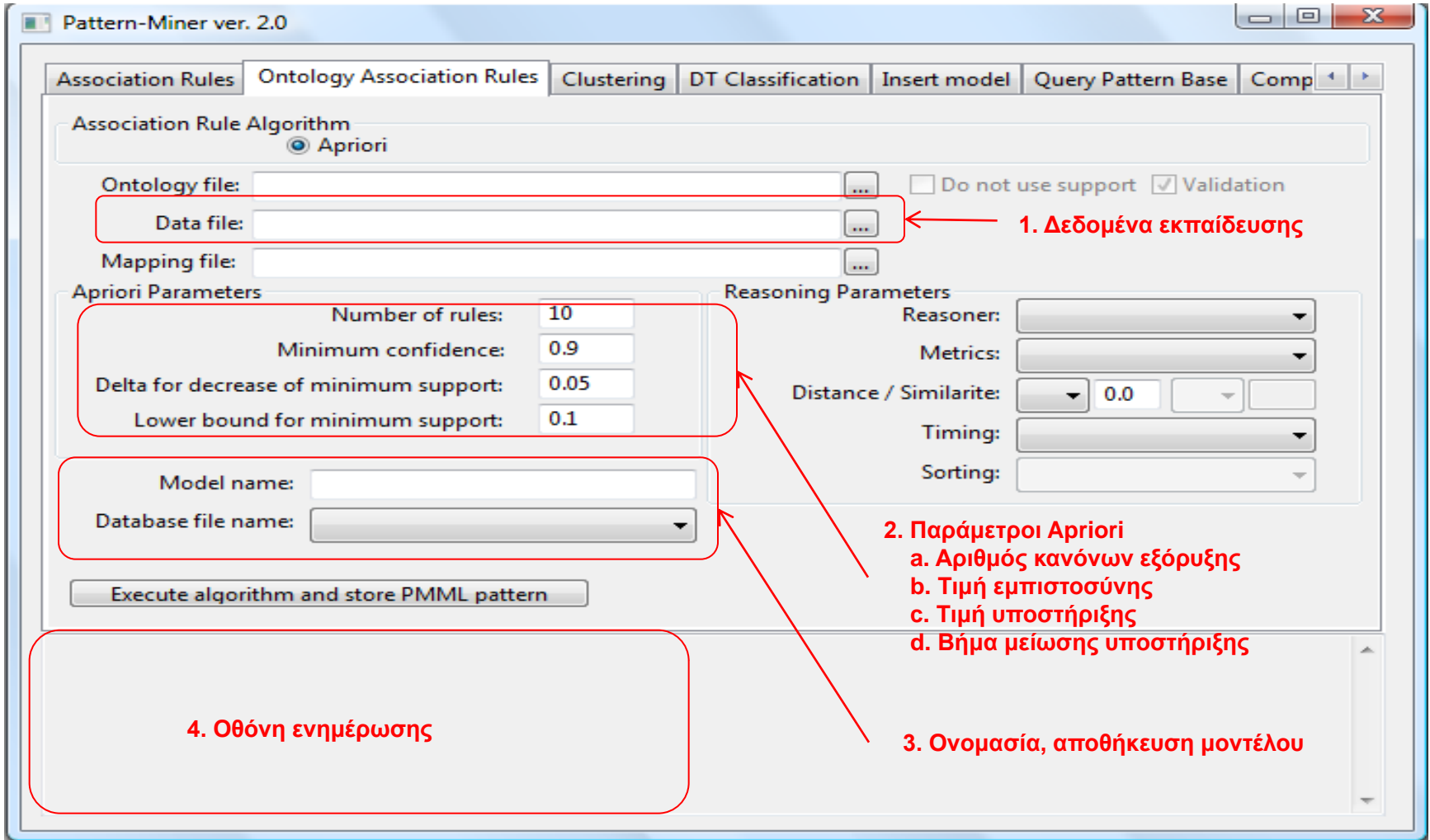
- **Έννοιες:** αντικείμενα μίας σύλληψης/ λεξιλόγιο
 - Ιεραρχική δομή με πολλαπλή κληρονομικότητα
- **Σχέσεις:** σύνδεση μεταξύ των εννοιών
 - Πάνω τους ορίζονται περιορισμοί (καθολικός, υπαρξιακός, πεδίο τιμών, πεδίο ορισμού)
 - Ιεραρχική δομή
- **Συναρτήσεις:** ειδικές περιπτώσεις σχέσεων
 - Ιδιότητες: αντιστροφής, συναρτησιακές, μεταβατικότητας, συμμετρικότητα
- **Αξιώματα:**
 - Προτάσεις που αληθεύουν πάντα
 - Συνδυασμός εννοιών και σχέσεων
- **Στιγμιότυπα:**
 - Τιμές εννοιών, ακολουθούν τους περιορισμούς- αξιώματα

- Εργαλείο παραγωγής λογικών συνεπειών
- Έλεγχος πάνω στην Οντολογία
 - Συνέπεια, ικανοποιησιμότητα, στιγμιότυπων, ιεραρχίας, ερωτήσεων
- Ταξινόμηση, πραγμάτωση
 - Αναδιοργάνωση του δέντρου ιεραρχιών
- Ανάγκη τυπικότητας & εκμετάλλευση εκφραστικότητας
- ↑ Εκφραστικότητας τότε ↑ Πολυπλοκότητας
 - Μη αποκρισιμότητα

- Ομοιότητα, ανομοιότητα, αποστάσεις εννοιών
- Εκμετάλλευση βάθους, πλάτους, πυκνότητα
- Διαφορετικές ανάγκες διαφορετικές μετρικές
 - Κατανοητά αποτελέσματα
 - Συνδυασμός μετρικών
 - Διακριτές τιμές, ασαφή σύνολα
 - Χρήση βαρών
 - Πολλαπλά μονοπάτια, ασύμμετρα μονοπάτια

- Βασική λειτουργία Weka Apriori
 - Δεδομένα εκπαίδευσης
 - Παράμετροι Apriori
- Ενσωμάτωση Οντολογιών
 - Δεδομένα εκπαίδευσης
 - Παράμετροι Apriori
 - Παράμετροι Συλλογιστικής
 - Μηχανή συλλογιστικής ανάλυσης, μετρική, σημασιολογικό πλαίσιο, τότε γίνεται η ενσωμάτωση γνώσης
- Σημασιολογικά προσανατολισμένοι κανόνες
 - Δεδομένα εκπαίδευσης
 - Παράμετροι Apriori *χωρίς μετρική Υποστήριξης*
 - Παράμετροι Συλλογιστικής
 - εκτέλεση κατά την διάρκεια, επιλογή ταξινόμησης αποτελεσμάτων

Βασική λειτουργία Weka Apriori



Pattern-Miner ver. 2.0

Association Rules | Ontology Association Rules | Clustering | DT Classification | Insert model | Query Pattern Base | Comp

Association Rule Algorithm
● Apriori

Ontology file: ... Do not use support Validation

Data file: ... ← 1. Δεδομένα εκπαίδευσης

Mapping file: ...

Apriori Parameters

Number of rules:	10
Minimum confidence:	0.9
Delta for decrease of minimum support:	0.05
Lower bound for minimum support:	0.1

Reasoning Parameters

Reasoner:

Metrics:

Distance / Similarite: 0.0

Timing:

Sorting:

Model name:

Database file name:

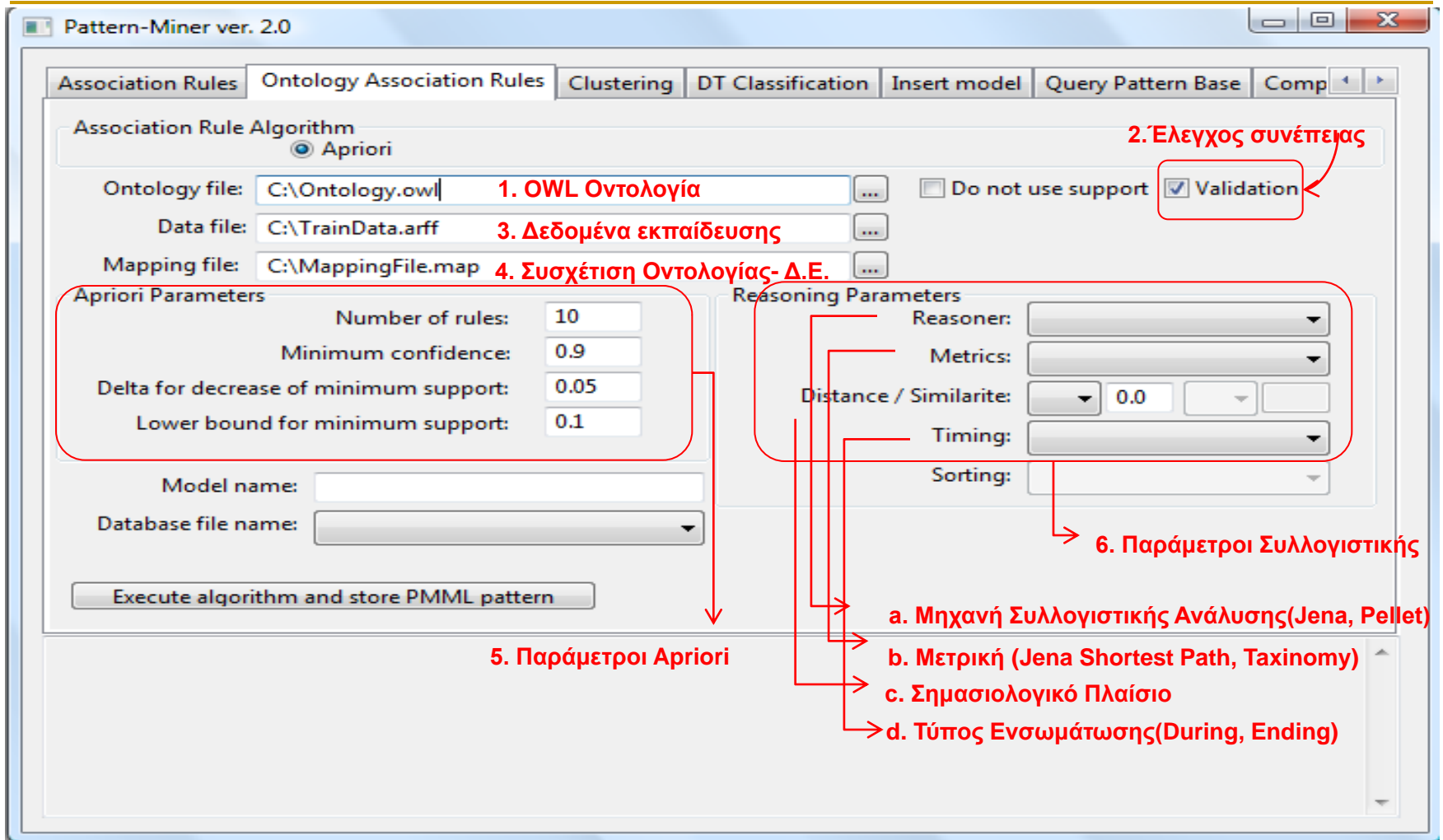
Execute algorithm and store PMML pattern

4. Οθόνη ενημέρωσης

2. Παράμετροι Apriori
a. Αριθμός κανόνων εξόρυξης
b. Τιμή εμπιστοσύνης
c. Τιμή υποστήριξης
d. Βήμα μείωσης υποστήριξης

3. Ονομασία, αποθήκευση μοντέλου

Ενσωμάτωση Οντολογιών



The screenshot shows the 'Ontology Association Rules' tab in the Pattern-Miner software. The interface is annotated with red text and arrows pointing to specific fields:

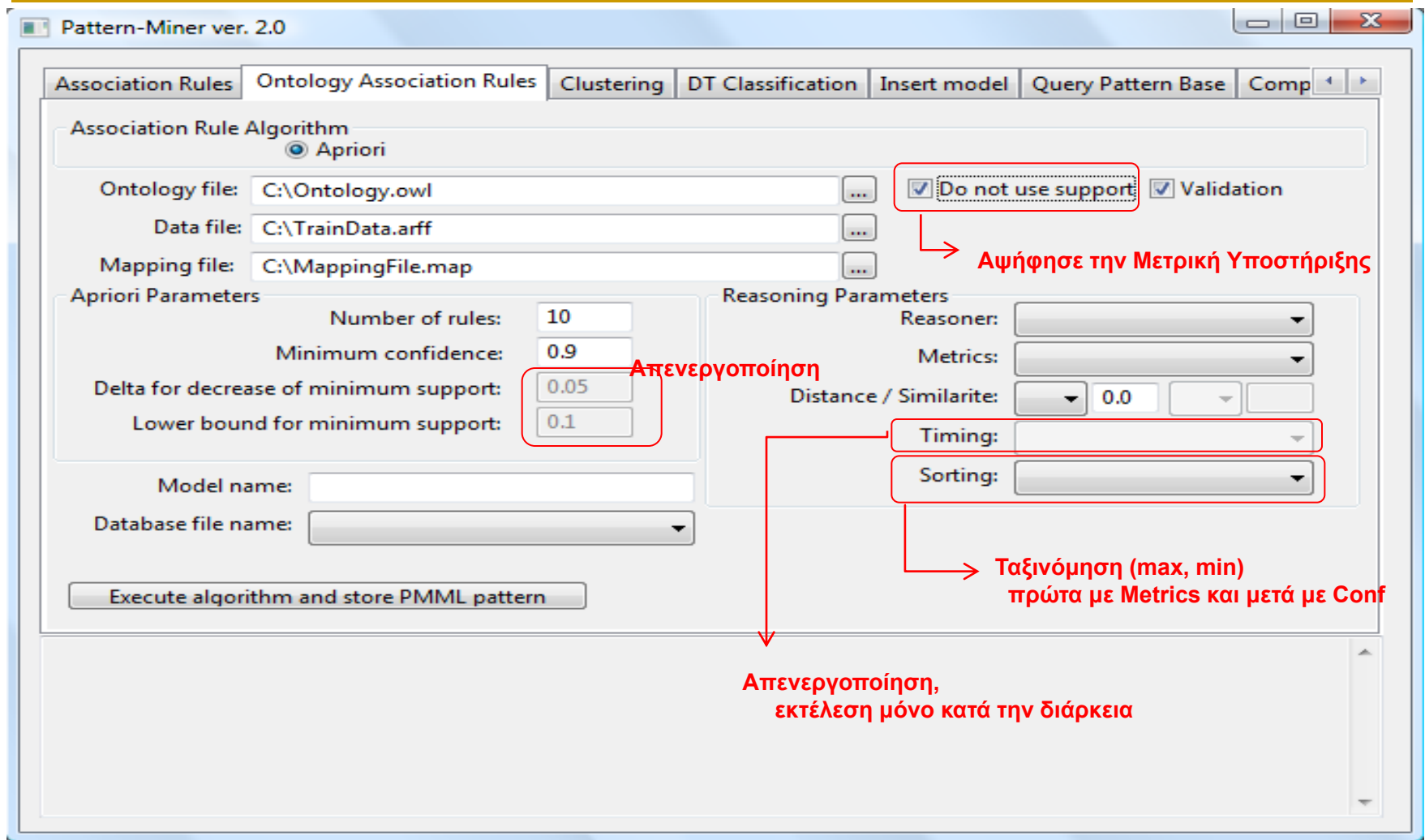
- 1. OWL Οντολογία**: Points to the 'Ontology file' field containing 'C:\Ontology.owl'.
- 2. Έλεγχος συνέπειας**: Points to the 'Validation' checkbox, which is checked.
- 3. Δεδομένα εκπαίδευσης**: Points to the 'Data file' field containing 'C:\TrainData.arff'.
- 4. Συσχέτιση Οντολογίας- Δ.Ε.**: Points to the 'Mapping file' field containing 'C:\MappingFile.map'.
- 5. Παράμετροι Apriori**: Points to the 'Apriori Parameters' section, which includes:
 - Number of rules: 10
 - Minimum confidence: 0.9
 - Delta for decrease of minimum support: 0.05
 - Lower bound for minimum support: 0.1
- 6. Παράμετροι Συλλογιστικής**: Points to the 'Reasoning Parameters' section, which includes:
 - Reasoner: (dropdown menu)
 - Metrics: (dropdown menu)
 - Distance / Similarite: (dropdown menu with value 0.0)
 - Timing: (dropdown menu)
 - Sorting: (dropdown menu)

Below the Reasoning Parameters section, a list of options is provided:

- a. Μηχανή Συλλογιστικής Ανάλυσης (Jena, Pellet)
- b. Μετρική (Jena Shortest Path, Taxinomy)
- c. Σημασιολογικό Πλαίσιο
- d. Τύπος Ενσωμάτωσης (During, Ending)

At the bottom of the interface, there is a button labeled 'Execute algorithm and store PMML pattern'.

Σημασιολογικά Προσανατολισμένοι Κανόνες



The screenshot shows the 'Ontology Association Rules' tab in the Pattern-Miner software. The 'Association Rule Algorithm' is set to 'Apriori'. The 'Do not use support' checkbox is checked, with a red box around it and an arrow pointing to the text 'Αφήφισε την Μετρική Υποστήριξης'. The 'Apriori Parameters' section has 'Number of rules' set to 10, 'Minimum confidence' set to 0.9, 'Delta for decrease of minimum support' set to 0.05, and 'Lower bound for minimum support' set to 0.1. A red box around the 0.05 and 0.1 values has an arrow pointing to the text 'Απενεργοποίηση'. The 'Reasoning Parameters' section has 'Reasoner' and 'Metrics' dropdowns, and 'Distance / Similarite' set to 0.0. The 'Timing' and 'Sorting' dropdowns are also highlighted with red boxes and an arrow pointing to the text 'Ταξινόμηση (max, min) πρώτα με Metrics και μετά με Conf'. A large red arrow points from the 'Reasoning Parameters' section to the text 'Απενεργοποίηση, εκτέλεση μόνο κατά την διάρκεια'. The 'Execute algorithm and store PMML pattern' button is visible at the bottom.

- Απομόνωση Οντολογίας Umbel περί τρομοκρατίας
 - Υποσύνολο της OpenCyc (112 έννοιες)
- Δεδομένα Εκπαίδευσης Global Terrorism Database
 - ΒΔ τρομοκρατικών ενεργειών από το 1970 έως το 2008
 - Απομόνωση 21 τρομοκρατικών οργανώσεων
 - Χρήση 12 μεταβλητών από 120

- Υποθέσεις σε μορφή Κανόνων, (Chasdi, 1997)
 - Πολιτική ιδεολογία => Είδος του στόχου
 - Ειδικοί: επιβεβαιώνεται / Ontology-AR : επιβεβαιώνεται
 - Πολιτική ιδεολογία => Αριθμό των νεκρών
 - Ειδικοί: επιβεβαιώνεται / Ontology-AR : επιβεβαιώνεται
 - Πολιτική ιδεολογία => Αριθμό των τραυματιών
 - Ειδικοί: επιβεβαιώνεται / Ontology-AR : επιβεβαιώνεται
 - Πολιτική ιδεολογία => Μέγεθος υλικών ζημιών
 - Ειδικοί: επιβεβαιώνεται / Ontology-AR : επιβεβαιώνεται
 - Περιοχή τρομοκρατικής ενέργειας => Είδος στόχου
 - Ειδικοί: επιβεβαιώνεται / Ontology-AR : επιβεβαιώνεται (no sup.)
 - Πολιτικά γεγονότα => Είδος στόχου
 - Ειδικοί: Δεν επιβεβαιώνεται / Ontology-AR : Ελλιπή δεδομένα

Σύγκριση Προσεγγίσεων

Κατηγ.	Χαρακτηριστικά	Apriori	οApriori	οApriori Without Support
Απόδοση	Χρήση μνήμης		+πίνακα σημασιολογικών αποστάσεων	+πίνακα σημασιολογικών αποστάσεων
	Κόστος κατασκευής πίνακα αποστάσεων	Όχι	1 φορά	1 φορά
	Κόστος ελέγχου	Support Confidence	Support Confidence Σημασιολογικός έλεγχος στοιχειοσυνόλων Σημασιολογικός έλεγχος κανόνων	Confidence Σημασιολογικός έλεγχος στοιχειοσυνόλων
	Στοιχειοσύνολο		Μείωση στοιχειοσυνόλων	Μεγάλο σύνολο στοιχειοσυνόλων
	Κύκλοι εκτέλεσης		Μείωση	1
	Κανόνες	Πολύ	Λίγοι	Μέτρια

Σύγκριση Προσεγγίσεων

Κατηγορία	Χαρακτηριστικά	Apriori	οApriori	οApriori Without Support
Σημασιολογία	Πλούσια σημ. ερμηνεία	Χαμηλή	Υψηλή	Υψηλή
	Κατανόηση αποτελ.	Χαμηλή	Υψηλή	Υψηλή
	Στοχευόμενοι κανόνες	Όχι	Ναι	Ναι
	Ανάδειξη κανόνων με μικρό support/ confidence	Χαμηλή	Μέτρια	Υψηλή
	Τετριμένη γνώση	Υψηλή (ισχυροί κανόνες μικρή σημασιολ.)	Μέτρια	Χαμηλή
	Λανθασμένη	Υψηλή	Χαμηλή	Χαμηλή
	Επανάληψη κανόνων	Υψηλή	Χαμηλή	Μέτρια
	Ενδιαφέρον κανόνες	Χαμηλή	Υψηλή/ ευελιξία μετρικών	Υψηλή/ ευελιξία μετρικών

- Ενσωμάτωσης Οντολογιών στην KDD
 - Αποδοτική και αναγκαία
- Χρήση κατάλληλων Οντολογιών
 - Εκφραστικές, πολυπλοκότητα
- Χρήση κατάλληλων Μηχανών Συλλογιστικής Ανάλυσης
 - Έλεγχο και εκμετάλλευση εκφραστικότητας Οντολογιών
- Χρήση κατάλληλων Μετρικών
 - Εξόρυξη ενδιαφέροντων και κατανοητών προτύπων
- Χρήση κατάλληλων τεχνικών DM
 - Επιλογή σημείου ενσωμάτωσης γνώσης και συνδυασμό υποκειμενικών, αντικειμενικών, σημασιολογικών μέτρων

