

Σύστημα Συστάσεων eNOVATION CBRN

1. Εισαγωγή

Το παρόν έγγραφο αναλύει την αρχιτεκτονική και τη λογική του **eNOVATION CBRN Recommender**, ενός συστήματος λήψης αποφάσεων βασισμένου σε σημασιολογικούς κανόνες (semantic rule-based system). Σκοπός του συστήματος είναι η υπόδειξη του βέλτιστου Κέντρου Εκπαίδευσης (Training Centre - TC) βάσει δύο παραμέτρων εισόδου:

1. **Technology:** Η απαιτούμενη τεχνολογία.
2. **Scenario:** Το επιχειρησιακό σενάριο.

Το σύστημα χρησιμοποιεί έναν Γράφο Γνώσης (Knowledge Graph) για την εξαγωγή συμπερασμάτων και έναν Αλγόριθμο Βαθμολόγησης Πολλαπλών Κριτηρίων για την κατάταξη των αποτελεσμάτων.

2. Rule Based Engine

Ο πυρήνας του συστήματος είναι ένα δυναμικό SPARQL query που εκτελείται στον Apache Jena Fuseki. Το σύστημα εντοπίζει και μετρά σημασιολογικές σχέσεις διασχίζοντας την οντολογία μέσω συγκεκριμένων σημασιολογικών διαδρομών (semantic paths). Το SPARQL query υπολογίζει πόσες φορές ικανοποιείται ένα κριτήριο, δηλαδή πόσες σχετικές σημασιολογικές διαδρομές εντοπίζονται για κάθε Training Centre. Τα αποτελέσματα αυτής της διαδικασίας αποτελούν τη βάση για τον υπολογισμό των επιμέρους Scores.

Τα κριτήρια μπορούν να χωριστούν σε δύο κατηγορίες, τα δυναμικά και τα στατικά κριτήρια.

2.1. Dynamic Criteria (Εξαρτώνται από το Input)

Αυτά τα κριτήρια καθορίζουν τη σχετικότητα και υπολογίζονται δυναμικά με βάση τις επιλογές του χρήστη.

1. **Tech Use:** Διαθέτει το Κέντρο την επιλεγμένη τεχνολογία;
 - *Semantic Path:* TC --[usesTechnology]--> Selected_Technology
2. **Tech Training:** Παρέχει το Κέντρο μάθημα που εκπαιδεύει πάνω στη συγκεκριμένη τεχνολογία;
 - *Semantic Path:* TC --[providesCourse]--> Course --[trainsOn]--> Selected_Technology
3. **Incident Coverage:** Έχει το Κέντρο εμπειρία σε περιστατικά που αποτελούν τη βάση του επιλεγμένου σεναρίου;

- *Semantic Path:* TC --[tacklesIncident]--> Incident <--[isBasedOnIncident]-- Selected_Scenario
4. **Threat Capability:** Διαθέτει το Κέντρο πόρους για την απειλή που εμπλέκεται στο σενάριο;
- *Semantic Path:* TC --[hasEquipment/usesTech]--> Resource --[addressesThreat]--> Threat <--[involvesThreat]-- Selected_Scenario

2.2. Static Criteria (Γενικές Δυνατότητες)

Αυτά τα κριτήρια είναι ανεξάρτητα από το input και μετρούν τη συνολική δυναμικότητα και τις υποδομές του κέντρου.

5. **Facilities:** Πλήθος εγκαταστάσεων.

- *Semantic Path:* TC --[hasFacility]--> Facility

6. **Disciplines:** Πλήθος επιστημονικών κλάδων (Disciplines).

- *Semantic Path:* TC --[hasTCDiscipline]--> Discipline

7. **Courses:** Συνολικό πλήθος μαθημάτων που προσφέρει το κέντρο.

- *Semantic Path:* TC --[providesTrainingCourse]--> Course

8. **Networks:** Πλήθος δικτύων στα οποία συμμετέχει.

- *Semantic Path:* TC --[connectsWithNetwork]--> Network

Κάθε κριτήριο μπορεί να λάβει τιμές σε διαφορετικό εύρος, ανάλογα με τη φύση του. Ορισμένα κριτήρια είναι δυαδικά, λαμβάνοντας τιμές από **0 έως 1**, ενώ άλλα βασίζονται σε πλήθος εμφανίσεων semantic paths και μπορούν να λάβουν τιμές έως και **0–33**, ανάλογα με το πόσες σχετικές τριπλέτες εντοπίζονται στο γραφικό μοντέλο.

Για παράδειγμα, το κριτήριο **Tech Use** είναι δυαδικό. Λαμβάνει τιμή **1** όταν το SPARQL query εντοπίζει semantic path της μορφής:

TC —[usesTechnology]→ Selected_Technology

δηλαδή όταν το Κέντρο πράγματι χρησιμοποιεί την επιλεγμένη τεχνολογία. Αν δεν εντοπιστεί αντίστοιχη διαδρομή, το κριτήριο λαμβάνει τιμή **0**.

Αντίθετα, κριτήρια όπως το **Networks** βασίζονται σε πολλαπλές εμφανίσεις semantic paths. Το σύστημα εντοπίζει όλες τις διαδρομές της μορφής:

TC —[connectsWithNetwork]→ Network

και ο αριθμός των δικτύων που βρίσκει μετατρέπεται στο αντίστοιχο αριθμητικό αποτέλεσμα του κριτηρίου. Έτσι, το τελικό score του κριτηρίου αντιπροσωπεύει το πλήθος των δικτύων στα οποία συμμετέχει το Κέντρο.

3. Ο Αλγόριθμος Βαθμολόγησης (Scoring Algorithm)

Η κατάταξη των κέντρων ακολουθεί τις αρχές της Πολυκριτηριακής Λήψης Αποφάσεων (MCDM), χρησιμοποιώντας ιεραρχική δομή (Clustering).

3.1. Κανονικοποίηση (Normalization)

Για να συγκριθούν ανόμοια μεγέθη, εφαρμόζεται κανονικοποίηση βάσει της μέγιστης τιμής που παρατηρείται στα τρέχοντα αποτελέσματα.

- *Tύπος:* $V_{norm-i} = \frac{V_i}{V_{max}}$

Όπου V_{norm-i} : η κανονικοποιημένη τιμή, V_i : η πραγματική τιμή που εμφανίζει το κριτήριο, V_{max} : η μέγιστη τιμή που εμφανίζει το κριτήριο σε κάποιο κέντρο με δεδομένη τεχνολογία και σενάριο.

3.2. Clustering & Weights

Τα κριτήρια ομαδοποιούνται σε Clusters. Τα βάρη έχουν επιλεγεί με βάση την κρισιμότητα, όπως ορίζεται σε πρότυπα NATO και ISO και τον βαθμό συσχέτισης με το ζητούμενο.

A. Operational Fit (Cluster Σχετικότητας)

Εξαρτάται από τα Dynamic Criteria.

Tύπος: Score_{op} = 0.35 · TechUse + 0.20 · TechTraining + 0.25 · IncidentCoverage + 0.20 · ThreatCapability

Τεκμηρίωση Βαρών:

1. Tech Use (0.35) & Training (0.20): Η διαθεσιμότητα φυσικών πόρων και εξοπλισμού θεωρείται κρίσιμη παράμετρος για την ανθεκτικότητα (Resilience) κατά το πρότυπο **ISO 22300**. Η έλλειψη ειδικού εξοπλισμού (π.χ. Mass Spectrometer) καθιστά την εκπαίδευση αδύνατη (hard constraint), γι' αυτό λαμβάνει το υψηλότερο βάρος.

2. Threat Capability & Incident: Σύμφωνα με τα πρότυπα **NATO ATP-3.8.1 (CBRN Defence Standards)**, η ικανότητα εξουδετέρωσης συγκεκριμένων παραγόντων (Threat Agents) είναι ο πρωταρχικός δείκτης επιτυχίας μιας μονάδας. Τα κριτήρια αυτά λειτουργούν συμπληρωματικά για να πλαισιώσουν την τεχνολογία στο επιχειρησιακό περιβάλλον.

B. Training Capacity (Cluster Δυναμικότητας)

- Τύπος: $Score_{cap} = 0.60 \cdot Courses + 0.40 \cdot Disciplines$

C. Infrastructure & Networks (Cluster Υποστήριξης)

- Τύπος: $Score_{infra} = 0.60 \cdot Facilities + 0.40 \cdot Networks$

3.3. Υπολογισμός Τελικού Σκορ & Core Score Logic

1. Base Score:

$$BaseScore = 0.65 \cdot Score_{op} + 0.20 \cdot Score_{infra} + 0.15 \cdot Score_{cap}$$

Το Operational Fit κυριαρχεί (65%), διασφαλίζοντας ότι προτείνονται κέντρα που είναι σχετικά και όχι απλώς "μεγάλα".

Σημείωση: Σε μελλοντική έκδοση της εφαρμογής, το σύστημα δύναται να ενσωματώσει λειτουργικότητα εξατομίκευσης βαρών. Μέσω αυτής, ο χρήστης θα έχει τη δυνατότητα να ρυθμίζει δυναμικά τις προτεραιότητες του αλγορίθμου, επιλέγοντας αν η σύσταση θα εστιάζει πρωτίστως στην Καταλληλότητα (Operational Fit) ή στη Συνολική Δυναμικότητα του κέντρου. Επιπλέον, θα μπορούσε να παρέχεται η ευελιξία καθορισμού της σχετικής βαρύτητας μεταξύ Τεχνολογίας και Σεναρίου, επιτρέποντας την προσαρμογή του μοντέλου σε ειδικές απαιτήσεις (Technology-centric vs Scenario-centric approach).

4. User Interface, Explainability & Feedback

4.1. Δυναμική Φόρτωση & Φιλτράρισμα

Το Frontend αντλεί τις λίστες επιλογών δυναμικά (/api/options). Χρησιμοποιείται φίλτρο owl:NamedIndividual στο SPARQL ώστε τα dropdowns να περιέχουν μόνο συγκεκριμένα instances (π.χ. "Foam disinfectants") και όχι αφηρημένες κλάσεις, διασφαλίζοντας την ακρίβεια των queries.

4.2. Explainability (XAI)

Το σύστημα ενισχύει την εμπιστοσύνη μέσω δύο επιπέδων επεξήγησης:

- **Simple Explanations:** Κείμενο φυσικής γλώσσας που εξηγεί γιατί επιλέχθηκε ένα κέντρο.
- **Justification Graph:** Οπτικοποίηση των ακριβών τριπλετών RDF που συνδέουν το Κέντρο με το Αίτημα (π.χ. Center A → providesCourse → Course B). Αυτό προσφέρει διαφάνεια.

4.3. Μηχανισμός Feedback

Το σύστημα καταγράφει τις αντιδράσεις των χρηστών (Bad/Neutral/Good) σε αρχείο JSONL. Σε περιπτώσεις πολλαπλών αλληλεπιδράσεων από τον ίδιο χρήστη για το ίδιο αποτέλεσμα ("αναποφάσιστος χρήστης"), για την ανάλυση πρέπει να λαμβάνεται υπόψη η εγγραφή με το μεταγενέστερο timestamp, καθώς αντικατοπτρίζει την τελική κρίση του χρήστη.

5. Διάγραμμα Ροής Συστήματος

Σύνοψη Διαδικασίας:

1. Το **Frontend** ζητά Σύσταση από το Backend.
2. Το **Backend** κατασκευάζει το ερώτημα SPARQL.
3. Ο **Fuseki Server** εκτελεί το ερώτημα στην Οντολογία.
4. Το **Backend** επεξεργάζεται τα αποτελέσματα, υπολογίζει μαθηματικά σκορ (χρησιμοποιώντας βάρη) και παράγει επεξηγήσεις.
5. Το **Frontend** αποδίδει την καταταγμένη λίστα με τα μονοπάτια αιτιολόγησης.