



ΠΑΝΕΠΙΣΤΗΜΙΟ
ΘΕΣΣΑΛΙΑΣ

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

ΔΙΑΧΕΙΡΙΣΗ ΕΤΕΡΟΓΕΝΕΙΑΣ ΣΕ ΡΟΕΣ ΔΕΔΟΜΕΝΩΝ ΓΙΑ ΤΗΝ ΕΞΑΓΩΓΗ ΓΝΩΣΗΣ

ΧΡΙΣΤΟΦΟΡΙΔΗΣ ΙΩΑΝΝΗΣ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

ΥΠΕΥΘΥΝΟΣ

Κων/νος Κολομβάτσος
Αναπληρωτής Καθηγητής Πανεπιστημίου Θεσσαλίας

Λαμία 3/7 έτος 2025



ΠΑΝΕΠΙΣΤΗΜΙΟ
ΘΕΣΣΑΛΙΑΣ

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

ΔΙΑΧΕΙΡΙΣΗ ΕΤΕΡΟΓΕΝΕΙΑΣ ΣΕ ΡΟΕΣ ΔΕΔΟΜΕΝΩΝ ΓΙΑ ΤΗΝ ΕΞΑΓΩΓΗ
ΓΝΩΣΗΣ

ΧΡΙΣΤΟΦΟΡΙΔΗΣ ΙΩΑΝΝΗΣ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

ΥΠΕΥΘΥΝΟΣ

Κων/νος Κολομβάτσος
Αναπληρωτής Καθηγητής Πανεπιστημίου Θεσσαλίας

Λαμία 3/7 έτος 2025



UNIVERSITY OF
THESSALY

SCHOOL OF SCIENCE

DEPARTMENT OF COMPUTER SCIENCE & TELECOMMUNICATIONS

MANAGING HETEROGENEITY IN DATA STREAMS FOR KNOWLEDGE
EXTRACTION

CHRISTOFORIDIS IOANNIS

FINAL THESIS

ADVISOR

Konstantinos Kolomvatsos
Associate Professor of the University of Thessaly

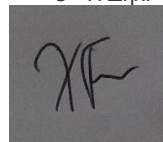
Lamia 3/7 year 2025

«Με ατομική μου ευθύνη και γνωρίζοντας τις κυρώσεις ⁽¹⁾, που προβλέπονται από της διατάξεις της παρ. 6 του άρθρου 22 του Ν. 1599/1986, δηλώνω ότι:

1. Δεν παραθέτω κομμάτια βιβλίων ή άρθρων ή εργασιών άλλων αυτολεξεί **χωρίς να τα περικλείω σε εισαγωγικά** και χωρίς να αναφέρω το συγγραφέα, τη χρονολογία, τη σελίδα. Η αυτολεξεί παράθεση χωρίς εισαγωγικά χωρίς αναφορά στην πηγή, είναι λογοκλοπή. Πέραν της αυτολεξεί παράθεσης, λογοκλοπή θεωρείται και η παράφραση εδαφίων από έργα άλλων, συμπεριλαμβανομένων και έργων συμφοιτητών μου, καθώς και η παράθεση στοιχείων που άλλοι συνέλεξαν ή επεξεργάστηκαν, χωρίς αναφορά στην πηγή. Αναφέρω πάντοτε με πληρότητα την πηγή κάτω από τον πίνακα ή σχέδιο, όπως στα παραθέματα.
2. Δέχομαι ότι η αυτολεξεί **παράθεση χωρίς εισαγωγικά**, ακόμα κι αν συνοδεύεται από αναφορά στην πηγή σε κάποιο άλλο σημείο του κειμένου ή στο τέλος του, είναι αντιγραφή. Η αναφορά στην πηγή στο τέλος π.χ. μιας παραγράφου ή μιας σελίδας, δεν δικαιολογεί συρραφή εδαφίων έργου άλλου συγγραφέα, έστω και παραφρασμένων, και παρουσίασή τους ως δική μου εργασία.
3. Δέχομαι ότι υπάρχει επίσης περιορισμός στο μέγεθος και στη συχνότητα των παραθεμάτων που μπορώ να εντάξω στην εργασία μου εντός εισαγωγικών. Κάθε μεγάλο παράθεμα (π.χ. σε πίνακα ή πλαίσιο, κλπ), προϋποθέτει ειδικές ρυθμίσεις, και όταν δημοσιεύεται προϋποθέτει την άδεια του συγγραφέα ή του εκδότη. Το ίδιο και οι πίνακες και τα σχέδια
4. Δέχομαι όλες τις συνέπειες σε περίπτωση λογοκλοπής ή αντιγραφής.

Ημερομηνία: 10/7/2025

Ο - Η Δηλ.



(1) «Όποιος εν γνώσει του δηλώνει ψευδή γεγονότα ή αρνείται ή αποκρύπτει τα αληθινά με έγγραφη υπεύθυνη δήλωση του άρθρου 8 παρ. 4 Ν. 1599/1986 τιμωρείται με φυλάκιση τουλάχιστον τριών μηνών. Εάν ο υπαίτιος αυτών των πράξεων σκόπευε να προσπορίσει στον εαυτόν του ή σε άλλον περιουσιακό όφελος βλάπτοντας τρίτον ή σκόπευε να βλάψει άλλον, τιμωρείται με κάθειρξη μέχρι 10 ετών.»

ΠΕΡΙΛΗΨΗ

Με τα συστήματα αστικών μεταφορών να γίνονται πιο περίπλοκα, χρειαζόμαστε νέους τρόπους παρακολούθησης της κυκλοφορίας που να μπορούν να χειριστούν τις μεγάλες ποσότητες διαφορετικών τύπων δεδομένων που παράγουν οι σύγχρονες πόλεις. Αυτή η μελέτη αναφέρεται σε ένα σημασιολογικά καθοδηγούμενο πλαίσιο για την έξυπνη παρακολούθηση της κυκλοφορίας στις πόλεις. Εστιάζει στον τρόπο συνδυασμού, ενσωμάτωσης και εναρμόνισης διαφόρων τύπων ροών δεδομένων από πηγές του Διαδικτύου των Πραγμάτων (IoT) όπως GPS, κάμερες κυκλοφορίας και βρόχους επαγωγής σε πραγματικό χρόνο.

Η προτεινόμενη αρχιτεκτονική έχει σχεδιασμό συστήματος πολλαπλών επιπέδων με ενότητες για επεξεργασία ροής σε πραγματικό χρόνο, χαρτογράφηση σχήματος και σημασιολογική σχολίαση. Η σημασιολογική διαλειτουργικότητα είναι σημαντική, πράγμα που σημαίνει ότι τα δεδομένα από διαφορετικές πηγές αντιστοιχίζονται χρησιμοποιώντας οντολογίες όπως το Δίκτυο Σημασιολογικών Αισθητήρων (SSN). Υπάρχουν δύο κύριοι αλγόριθμοι στο σύστημα: μια διαδικασία σημασιολογικής εναρμόνισης βασισμένη σε κανόνες και ένα μοντέλο σύντηξης βασισμένο σε δέντρο που συνδυάζει ετικέτες σημασιολογικής κυκλοφορίας σε μεμονωμένους περιγραφείς. Δημιουργήσαμε ένα πρωτότυπο χρησιμοποιώντας Python, RDF και Apache Kafka και το δοκιμάσαμε σε ένα προσομοιωμένο περιβάλλον κυκλοφορίας στην πόλη. Τα αποτελέσματα δείχνουν ότι η σημασιολογική ευθυγράμμιση είναι αρκετά ακριβής, η καθυστέρηση είναι πολύ χαμηλή και η επεκτασιμότητα είναι πολύ ισχυρή όταν υπάρχουν πολλά δεδομένα. Σε σύγκριση με τις μεθόδους συγχώνευσης επίπεδων δεδομένων, οι συγκριτικές αξιολογήσεις δείχνουν ότι αυτές οι μέθοδοι λειτουργούν καλύτερα, ειδικά όσον αφορά την επίγνωση του πλαισίου και την ικανότητα συλλογισμού.

Η εργασία ολοκληρώνεται με την απαρίθμηση των αδυναμιών του συστήματος, όπως η αδυναμία του να χειριστεί ασυνήθιστες ετικέτες και θορυβώδεις εισόδους, και στη συνέχεια αναφέρεται στο τι πρέπει να γίνει στη συνέχεια, όπως η προσθήκη στην οντολογία, η βελτίωση της υπολογιστικής αιχμής και η ανάλυση των δεδομένων κυκλοφορίας με τρόπο που προστατεύει την ιδιωτικότητα. Αυτό το πλαίσιο θέτει τις βάσεις για έξυπνες λύσεις μεταφορών στην πόλη που είναι γεμάτες δεδομένα και έξυπνες στη χρήση τους.

Λέξεις-κλειδιά: Έξυπνη Παρακολούθηση Κυκλοφορίας, Διαδίκτυο των Πραγμάτων (IoT), Σύντηξη Ροής Δεδομένων, Σημασιολογική Εναρμόνιση

ABSTRACT

As urban transportation systems get more complicated, we need new ways to monitor traffic that can handle the large amounts of different types of data that modern cities produce. This study talks about a semantic-driven framework for smart traffic monitoring in cities. It focuses on how to combine, integrate, and harmonize several types of data streams from Internet of Things (IoT) sources like GPS, traffic cameras, and induction loops in real time.

The suggested architecture has a multi-level system design with modules for real-time stream processing, schema mapping, and semantic annotation. Semantic interoperability is important, which means that data from diverse sources is matched using ontologies like the Semantic Sensor Network (SSN). There are two main algorithms in the system: a rule-based semantic harmonization process and a tree-based fusion model that combines semantic traffic tags into single descriptors. We made a prototype using Python, RDF, and Apache Kafka and tested it in a simulated city traffic setting. The results show that the semantic alignment is quite accurate, the latency is very low, and the scalability is very strong when there is a lot of data. When compared to flat data merging methods, comparative evaluations show that these methods work better, especially when it comes to being aware of context and being able to reason.

The paper ends by listing the system's weaknesses, such as its inability to handle unusual labels and noisy inputs, and then it talks about what needs to be done next, such adding to the ontology, improving edge computing, and analyzing traffic data in a way that protects privacy. This framework sets the stage for smart city transportation solutions that are full of data and smart in their use of it.

Keywords: Smart Traffic Monitoring, Internet of Things (IoT), Data Stream Fusion, Semantic Harmonization

Table of Contents

ΠΕΡΙΛΗΨΗ	I
ABSTRACT	IV
ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ	3
1.1 ΟΡΙΣΜΟΣ ΤΟΥ ΙΟΤ ΚΑΙ ΣΗΜΑΣΙΑ ΣΤΗ ΣΥΓΧΡΟΝΗ ΕΠΟΧΗ	3
1.2 ΠΡΟΒΛΗΜΑΤΑ ΚΑΙ ΠΡΟΚΛΗΣΕΙΣ ΣΤΗΝ ΕΤΕΡΟΓΕΝΕΙΑ ΤΩΝ ΡΟΩΝ ΔΕΔΟΜΕΝΩΝ	4
1.3 ΕΞΑΓΩΓΗ ΓΝΩΣΗΣ ΑΠΟ ΚΑΤΑΝΕΜΗΜΕΝΕΣ ΡΟΕΣ ΔΕΔΟΜΕΝΩΝ	6
1.4 ΣΚΟΠΟΣ ΚΑΙ ΣΤΟΧΟΙ ΤΗΣ ΠΑΡΟΥΣΑΣ ΜΕΛΕΤΗΣ	7
1.5 ΜΕΘΟΔΟΛΟΓΙΚΗ ΠΡΟΣΕΓΓΙΣΗ ΚΑΙ ΔΟΜΗ ΕΡΓΑΣΙΑΣ.....	8
ΚΕΦΑΛΑΙΟ 2: ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ ΚΑΙ ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΕΠΙΣΚΟΠΗΣΗ	9
2.1 ΜΟΡΦΕΣ ΚΑΙ ΠΗΓΕΣ ΕΤΕΡΟΓΕΝΕΙΑΣ ΣΤΟ ΙΟΤ.....	9
2.2 ΠΡΟΣΕΓΓΙΣΕΙΣ ΕΝΟΠΟΙΗΣΗΣ ΔΕΔΟΜΕΝΩΝ (ΤΕΧΝΙΚΕΣ, ΜΟΝΤΕΛΑ, ΠΡΟΤΥΠΑ)	11
2.3 ΣΗΜΑΣΙΟΛΟΓΙΚΗ ΔΙΑΛΕΙΤΟΥΡΓΙΚΟΤΗΤΑ ΚΑΙ ΟΝΤΟΛΟΓΙΕΣ	14
2.4 ΣΥΓΧΩΝΕΥΣΗ ΔΕΔΟΜΕΝΩΝ: ΕΠΙΠΕΔΗ, ΙΕΡΑΡΧΙΚΗ, ΔΕΝΤΡΑ, ΠΛΕΓΜΑ, ΧΩΡΙΣ ΔΟΜΗ	17
2.5 ΥΦΙΣΤΑΜΕΝΕΣ ΜΕΛΕΤΕΣ ΓΙΑ ΕΞΑΓΩΓΗ ΓΝΩΣΗΣ ΣΕ ΕΤΕΡΟΓΕΝΗ ΙΟΤ ΠΕΡΙΒΑΛΛΟΝΤΑ	19
2.6 ΚΡΙΤΙΚΗ ΑΠΟΤΙΜΗΣΗ ΚΑΙ ΕΡΕΥΝΗΤΙΚΟ ΚΕΝΟ	21
ΚΕΦΑΛΑΙΟ 3: ΠΛΑΙΣΙΟ ΑΡΧΙΤΕΚΤΟΝΙΚΗΣ ΚΑΙ ΡΟΗΣ ΠΛΗΡΟΦΟΡΙΑΣ ...	25
3.1 ΑΠΑΙΤΗΣΕΙΣ ΤΟΥ ΣΥΣΤΗΜΑΤΟΣ ΚΑΙ ΑΝΑΛΥΣΗ ΠΕΔΙΟΥ	25
3.2 ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΠΟΛΛΑΠΛΩΝ ΕΠΙΠΕΔΩΝ	28
3.3 ΜΟΝΤΕΛΟ ΣΗΜΑΣΙΟΛΟΓΙΚΗΣ ΣΥΝΕΝΩΣΗΣ ΚΑΙ ΧΑΡΤΟΓΡΑΦΗΣΗ ΣΧΗΜΑΤΩΝ.....	30
ΚΕΦΑΛΑΙΟ 4: ΥΛΟΠΟΙΗΣΗ ΠΡΩΤΟΤΥΠΟΥ ΚΑΙ ΣΥΓΧΩΝΕΥΣΗ ΡΟΩΝ	37
4.1 ΠΕΡΙΓΡΑΦΗ ΕΦΑΡΜΟΓΗΣ / ΠΕΙΡΑΜΑΤΙΚΗΣ ΔΟΜΗΣ	37
4.2 ΣΥΓΧΩΝΕΥΣΗ ΡΟΩΝ ΔΕΔΟΜΕΝΩΝ ΜΕ ΔΕΝΤΡΑ	38
4.3 ΕΜΠΛΟΥΤΙΣΜΟΣ ΜΕ ΣΥΜΦΡΑΖΟΜΕΝΑ (CONTEXT-AWARE FUSION).....	40
4.4 ΠΑΡΑΓΟΜΕΝΟ RDF ΚΑΙ ΑΠΟΤΕΛΕΣΜΑΤΑ ΕΝΟΠΟΙΗΣΗΣ.....	42
ΚΕΦΑΛΑΙΟ 5: ΑΞΙΟΛΟΓΗΣΗ, ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΜΕΛΛΟΝΤΙΚΕΣ ΚΑΤΕΥΘΥΝΣΕΙΣ.....	45
5.1 ΚΡΙΤΗΡΙΑ ΑΞΙΟΛΟΓΗΣΗΣ ΤΗΣ ΣΥΓΧΩΝΕΥΣΗΣ	45
5.2 ΣΥΖΗΤΗΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ ΚΑΙ ΣΥΓΚΡΙΤΙΚΗ ΑΞΙΟΛΟΓΗΣΗ	48
5.3 ΠΕΡΙΟΡΙΣΜΟΙ ΚΑΙ ΠΡΟΤΑΣΕΙΣ ΓΙΑ ΜΕΛΛΟΝΤΙΚΗ ΈΡΕΥΝΑ.....	50
ΒΙΒΛΙΟΓΡΑΦΙΑ.....	52

ΚΕΦΑΛΑΙΟ 1: Εισαγωγή

1.1 Ορισμός του IoT και σημασία στη σύγχρονη εποχή

Ένα δίκτυο συνδεδεμένων φυσικών πραγμάτων ενσωματωμένων με αισθητήρες, λογισμικό και άλλες τεχνολογίες, το Διαδίκτυο των Πραγμάτων (IoT) τους επιτρέπει να συλλέγουν και να μοιράζονται δεδομένα μέσω του διαδικτύου. Αυτά τα «πράγματα» μπορεί να ποικίλλουν από οικιακούς θερμοστάτες και φορητές συσκευές παρακολούθησης φυσικής κατάστασης έως βιομηχανικά μηχανήματα και αυτόνομα αυτοκίνητα. Το χαρακτηριστικό που διαφοροποιεί τα συστήματα IoT είναι η ικανότητά τους να επικοινωνούν όχι μόνο με τους ανθρώπους αλλά και μεταξύ τους, επιτρέποντας έτσι τον αυτοματισμό, τη λήψη αποφάσεων σε πραγματικό χρόνο και τις έξυπνες υπηρεσίες σε όλους τους τομείς. Το IoT έχει γίνει ένας σημαντικός μοχλός ψηφιακού μετασχηματισμού σε κλάδους όπως η υγειονομική περίθαλψη, η γεωργία, η εφοδιαστική, η μεταποίηση και οι αστικές υποδομές στη σύγχρονη εποχή. Βελτιώνει την επιχειρησιακή αποτελεσματικότητα, μειώνει τα έξοδα και παράγει πληροφορίες που βασίζονται σε δεδομένα, βελτιστοποιώντας την εμπειρία των χρηστών και τη χρήση πόρων (Akanbi & Masinde, 2020). Οι έξυπνες πόλεις, για παράδειγμα, χρησιμοποιούν το IoT για προσαρμοστική διαχείριση της κυκλοφορίας, περιβαλλοντική παρακολούθηση και ενεργειακά αποδοτικά συστήματα φωτισμού. Και στην υγειονομική περίθαλψη, οι φορητές συσκευές IoT παρέχουν συνεχή παρακολούθηση των ασθενών, επιτρέποντας έτσι την έγκαιρη διάγνωση και την εξατομικευμένη θεραπεία (Corral-Plaza et al., 2020). Η ένταξη του IoT στην καθημερινή ζωή σηματοδοτεί μια στροφή προς την πανταχού παρούσα υπολογιστική, όπου οι υπηρεσίες που λαμβάνουν υπόψη το περιβάλλον ανταποκρίνονται προληπτικά στις απαιτήσεις των χρηστών. Ενισχύοντας τους ευρύτερους στόχους της τεχνητής νοημοσύνης, οι ροές δεδομένων που παράγονται από τις συσκευές IoT συμβάλλουν στη δημιουργία έξυπνων συστημάτων ικανών να μαθαίνουν και να εξελίσσονται με την πάροδο του χρόνου (Gama & Aguilar-Ruiz, 2007). Κατά συνέπεια, το IoT δεν αποτελεί απλώς μια τεχνολογική τάση, αλλά θεμελιώδες στοιχείο των τρεχόντων κυβερνοφυσικών συστημάτων και του έξυπνου περιβάλλοντος.

1.2 Προβλήματα και προκλήσεις στην ετερογένεια των ροών δεδομένων

Αν και έχει μεγάλη δύναμη να αλλάξει, το IoT εγείρει επίσης δύσκολα ερωτήματα - η κυριότερη από αυτές είναι η ετερογένεια των πηγών δεδομένων. Αποτελούμενα από συσκευές από πολλούς κατασκευαστές, τα οικοσυστήματα IoT χρησιμοποιούν διάφορα πρωτόκολλα σύνδεσης, μορφές δεδομένων και ταχύτητες μετάδοσης. Αυτή η ποικιλία δημιουργεί συντακτικές, σημασιολογικές και δομικές διαφορές που καθιστούν την ενσωμάτωση πιο δύσκολη και εμποδίζουν την εξαγωγή γνώσης (Kumar et al., 2021; Singh & Siwach, 2022). Οι διαφορές στις μορφές δεδομένων - όπως JSON, XML ή ιδιόκτητες δομές - προκαλούν συντακτική ετερογένεια, γεγονός που καθιστά δύσκολη για τα συστήματα την επεξεργασία και ερμηνεία των εισερχόμενων δεδομένων με συνέπεια. Αυτές οι αποκλίσεις μπορεί να απαιτούν σημαντική εργασία ομαλοποίησης και επομένως να καθυστερούν τους αγωγούς επεξεργασίας δεδομένων (Adnan & Akbar, 2019). Οι μετρήσεις θερμοκρασίας, για παράδειγμα, μπορεί να χρησιμοποιούν διαφορετικά ονόματα πεδίων ή μονάδες - Κελσίου ή Φαρενάιτ - οι οποίες θα χρειάζονταν προεπεξεργασία πριν συγκριθούν ή συνδυαστούν.

Η σημασιολογική ετερογένεια είναι ο όρος που χρησιμοποιείται για να περιγράψει αποκλίσεις στην έννοια ή το πλαίσιο των στοιχείων δεδομένων σε πολλές πηγές. Δύο συσκευές μπορεί να υποδεικνύουν τιμές "κατάστασης", σε μια περίπτωση, σχετίζεται με το επίπεδο της μπαταρίας, στην άλλη, στον τρόπο λειτουργίας. Τα δεδομένα από πολλές συσκευές δεν μπορούν να συγχωνευθούν ή να ερμηνευθούν με συνέπεια χωρίς σημασιολογική ευθυγράμμιση, με αποτέλεσμα να διακυβεύεται η ακρίβεια των συστημάτων λήψης αποφάσεων (Kolozali et al., 2014; Singh & Siwach, 2022). Η δομική ετερογένεια είναι η διάταξη των δεδομένων. Ορισμένες συσκευές, για παράδειγμα, εμφανίζουν επίπεδες δομές δεδομένων, ενώ άλλες χρησιμοποιούν ιεραρχικές ή πολυεπίπεδες διατάξεις. Για να διατηρηθεί η απόδοση και η αξιοπιστία, τα συστήματα επεξεργασίας ροής πρέπει να προσαρμόζονται δυναμικά σε αυτές τις δομικές αλλαγές (Putrama & Martinek, 2024).

Η επεξεργασία σε πραγματικό χρόνο προσθέτει περισσότερες απαιτήσεις. Η υψηλή ταχύτητα, ο μεγάλος όγκος και η μεταβλητότητα χαρακτηρίζουν τις ροές

δεδομένων του IoT. Χωρίς καθυστερήσεις ή προβλήματα συμφόρησης, τα συστήματα πρέπει να καταναλώνουν, να καθαρίζουν, να εναρμονίζουν και να αναλύουν δεδομένα εν κινήσει. Τα παραδοσιακά συστήματα επεξεργασίας δεδομένων συνήθως δεν έχουν ρυθμιστεί για να διαχειρίζονται αυτό, κάτι που μπορεί να προκαλέσει προβλήματα όπως βραδύτητα, απώλεια δεδομένων ή υπερφόρτωση (Akanbi & Masinde, 2020; Gama & Aguilar-Ruiz, 2007). Επιπλέον, ο ασύγχρονος χαρακτήρας των εισόδων IoT καθιστά την ευθυγράμμιση ροής ένα ζωτικό πρόβλημα. Οι συσκευές θα μπορούσαν να μεταδίδουν δεδομένα σε διαφορετικά χρονικά διαστήματα ή με ανομοιόμορφες καθυστερήσεις, γεγονός που δυσχεραίνει τη συσχέτιση συμβάντων και την ενσωμάτωση χρονοσειρών (Ramachandran et al., 2024). Ο χρονικός κατακερματισμός καθιστά δύσκολη την εξαγωγή σωστών, ολοκληρωμένων συμπερασμάτων από δεδομένα χωρίς συγχρονισμό.

Η ασφάλεια και η προστασία της ιδιωτικότητας προσθέτουν περαιτέρω πολυπλοκότητα. Οι ετερογενείς συσκευές ενδέχεται να έχουν ποικίλες δυνατότητες κρυπτογράφησης ή ελέγχου ταυτότητας, οι οποίες θα μπορούσαν να τις εκθέσουν σε επιθέσεις ή διαρροές δεδομένων. Η συνεπής διαχείριση ταυτότητας και τα πρωτόκολλα ασφαλούς μετάδοσης δεν χρησιμοποιούνται συνήθως καθολικά σε όλα τα επίπεδα του IoT, επομένως η εγγύηση ασφαλούς και αξιόπιστης ενσωμάτωσης απαιτεί αυτές τις πρακτικές. Επιπρόσθετα, η ποικιλία στην ποιότητα των δεδομένων σε όλες τις πηγές - όπως οι ελλείπουσες τιμές, η απόκλιση των αισθητήρων ή ο θόρυβος - μπορεί να θέσει σε κίνδυνο την αξιοπιστία των κατάντη αναλύσεων. Πολλές τεχνικές εξαγωγής γνώσης προϋποθέτουν σταθερά, καθαρά δεδομένα εισόδου, τα οποία σπάνια ταιριάζουν με την πραγματικότητα των συσκευών IoT που αναπτύσσονται στο πεδίο (Ghoshal & Roberts, 2016).

Μια σημαντική δυσκολία εξακολουθεί να είναι η απουσία ευρέως αναγνωρισμένων προτύπων για τη σημασιολογική διαλειτουργικότητα. Αν και έννοιες όπως η οντολογία του Semantic Sensor Network (SSN) παρέχουν χρήσιμη κατεύθυνση, η χρήση τους σπάνια είναι συνεπής και οι εξατομικευμένες λύσεις τείνουν να μην έχουν δυνατότητα γενίκευσης (Kolozali et al., 2014; Singh & Siwach, 2022). Έτσι, η αποτελεσματικότητα του IoT περιορίζεται σε μεγάλο βαθμό από την ποικιλομορφία των ροών δεδομένων, ακόμη και αν δημιουργεί νέους

ορίζοντες για αυτοματοποίηση και έξυπνη λήψη αποφάσεων. Η αντιμετώπιση αυτών των ζητημάτων απαιτεί ένα μείγμα αρχιτεκτονικών, σημασιολογικών και αλγοριθμικών τεχνικών ικανών να διαχειριστούν την πολυπλοκότητα των τρεχόντων δικτύων IoT.

1.3 Εξαγωγή γνώσης από κατανεμημένες ροές δεδομένων

Η εξαγωγή γνώσης από κατανεμημένες ροές δεδομένων είναι η συνεχής, σε πραγματικό χρόνο διαδικασία μετατροπής ακατέργαστων, ποικίλων και συχνά θορυβωδών δεδομένων που συλλέγονται από γεωγραφικά διασκορπισμένες συσκευές IoT σε οργανωμένες πληροφορίες που μπορούν να καθοδηγήσουν τη λήψη αποφάσεων. Σε αντίθεση με την συμβατική εξόρυξη δεδομένων, η οποία χειρίζεται στατικά σύνολα δεδομένων, η εξαγωγή γνώσης στο IoT πρέπει να περιλαμβάνει δεδομένα υψηλής ταχύτητας, μεγάλου όγκου και υψηλής μεταβλητότητας που προέρχονται από πολλές πηγές σε ασύγχρονα μοτίβα (Gama & Aguilar-Ruiz, 2007).

Σε αυτήν την περίπτωση, η επεξεργασία σε πραγματικό χρόνο είναι απολύτως απαραίτητη. Χωρίς την πολυτέλεια της αποθήκευσης δεδομένων για επακόλουθη ανάλυση, τα συστήματα πρέπει να είναι σε θέση να εντοπίζουν μοτίβα, ανωμαλίες ή συσχετίσεις καθώς καταναλώνονται δεδομένα. Αυτό θέτει σοβαρές απαιτήσεις στη διαχείριση μνήμης, την ταχύτητα επεξεργασίας και την ανοχή σφαλμάτων (van Rijn et al., 2017). Για την αντιμετώπιση αυτών των αναγκών χρησιμοποιούνται ευρέως οι μηχανές επεξεργασίας κατανεμημένων ροών και τα συστήματα middleware, τα οποία επιτρέπουν το άμεσο φιλτράρισμα, τον μετασχηματισμό και τη συσσωμάτωση δεδομένων από πολλούς αισθητήρες και συσκευές edge (Akanbi & Masinde, 2020). Η διευκόλυνση της εξαγωγής γνώσης εξαρτάται σε μεγάλο βαθμό από τη σημασιολογική σχολίαση και τον εμπλουτισμό με βάση τα συμφραζόμενα. Τα συστήματα μπορούν να εντοπίσουν συσχετίσεις και να καταλήξουν σε συμπεράσματα πέρα από τη βασική στατιστική ανάλυση, προσθέτοντας οντολογικές ετικέτες και μεταδεδομένα ειδικά για κάθε τομέα σε στοιχεία δεδομένων (Kolozalet al., 2014). Για παράδειγμα, η γνώση ότι μια παρόμοια τοποθεσία έχει ταχεία μείωση της θερμοκρασίας και υψηλό επίπεδο

υγρασίας μπορεί να βοηθήσει στην πρόβλεψη ειδοποιήσεων σχετικά με πιθανές καιρικές ανωμαλίες.

Η διαδικασία εξαγωγής ενσωματώνεται επίσης όλο και περισσότερο με τη μηχανική μάθηση. Ακόμα και με δεδομένα με θόρυβο ή εν μέρει απόντα δεδομένα, μοντέλα όπως οι γκαουσιανές διαδικασίες και τα πλαίσια μάθησης συνόλων βοηθούν στην εξαγωγή περίπλοκων μοτίβων και στη δημιουργία προβλέψεων (Ghoshal & Roberts, 2016; van Rijn et al., 2017). Η προσαρμογή αυτών των μοντέλων σε δυναμικά περιβάλλοντα με ποικίλες εισόδους παρουσιάζει δυσκολίες, ωστόσο η πλήρης αξιοποίηση των δυνατοτήτων των συστημάτων IoT εξαρτάται από την εξαγωγή πληροφοριών από διάσπαρτες ροές δεδομένων. Απαιτείται εναρμονισμένη εισαγωγή δεδομένων, έξυπνη επεξεργασία και συστήματα συλλογισμού που να γνωρίζουν το περιβάλλον, ικανά να διαχειρίζονται την πολυπλοκότητα ποικίλων πηγών πληροφοριών σε πραγματικό χρόνο.

1.4 Σκοπός και στόχοι της παρούσας μελέτης

Με τελικό στόχο τη βελτίωση της ακρίβειας και της αποτελεσματικότητας των λειτουργιών εξαγωγής γνώσης, ο κύριος στόχος της παρούσας εργασίας είναι η δημιουργία και η αξιολόγηση ενός ισχυρού πλαισίου για τον έλεγχο της ετερογένειας των δεδομένων σε απομακρυσμένα συστήματα IoT. Η δυνατότητα συνδυασμού δεδομένων από πολλές, ποικίλες πηγές σε μία μόνο μορφή καθίσταται βασική δυνατότητα για έξυπνη λήψη αποφάσεων και αυτοματοποίηση σε πολλούς τομείς καθώς το IoT αναπτύσσεται.

Η έρευνα αποσκοπεί στα εξής:

- Να εξετάσει τα είδη και την προέλευση της μεταβλητότητας στις ροές δεδομένων IoT.
- Να εξετάσει τις τρέχουσες θεωρίες και μεθόδους ενσωμάτωσης και σύντηξης δεδομένων.
- Να προτείνει και να εφαρμόσει μια νέα προσέγγιση για την εναρμόνιση ανόμοιων δεδομένων χρησιμοποιώντας ένα κοινό σχήμα και λογική σύντηξης βασισμένη σε δέντρα.

- Να αξιολογήσει την αποτελεσματικότητα της προτεινόμενης λύσης όσον αφορά τη σημασιολογική συνέπεια, την υπολογιστική αποδοτικότητα και την ποιότητα της παραγωγής γνώσης.
- Να υποδείξει πιθανές μελλοντικές ερευνητικές κατευθύνσεις στον χειρισμό της ετερογένειας των δεδομένων IoT και την ανάπτυξη προσαρμοστικών μεθόδων εξαγωγής.

Μέσω αυτής της μελέτης, θέλουμε να καλύψουμε σημαντικά κενά στην τρέχουσα γνώση σχετικά με τη σημασιολογική ευθυγράμμιση, την αυτοματοποιημένη ενοποίηση σχημάτων και την απόδοση ολοκλήρωσης σε πραγματικό χρόνο σε ποικίλα περιβάλλοντα (Putrama & Martinek, 2024; Kumar et al., 2021).

1.5 Μεθοδολογική προσέγγιση και δομή εργασίας

Η μεθοδολογική προσέγγιση αυτής της εργασίας αντιμετωπίζει τα ζητήματα της ετερογένειας των δεδομένων σε περιβάλλοντα IoT μέσω θεωρητικής έρευνας σε συνδυασμό με πρακτική εφαρμογή. Για να βρεθούν οι τρέχουσες μέθοδοι, μοντέλα και κενά που συνδέονται με την ετερογενή ενσωμάτωση δεδομένων και την εξαγωγή πληροφοριών, πραγματοποιήθηκε αρχικά μια εις βάθος βιβλιογραφική μελέτη. Βασιζόμενη σε αυτά τα ευρήματα, η έρευνα προσφέρει ένα πλαίσιο σύντηξης δεδομένων βασισμένο σε δέντρο, ικανό να εναρμονίσει ανόμοιες ροές IoT σε ένα ομοιόμορφο σχήμα, επιτρέποντας έτσι τη σημασιολογική συνέπεια και την αποτελεσματική επεξεργασία. Για να αξιολογηθεί η απόδοση ενός πρωτοτύπου συστήματος όσον αφορά την ακρίβεια, την καθυστέρηση και τη διαλειτουργικότητα, χρησιμοποιήθηκαν προσομοιωμένα ετερογενή δεδομένα IoT για τον σχεδιασμό και τη δοκιμή του. Το πλαίσιο της μελέτης περιλαμβάνει μια εισαγωγή στο θέμα και τη συνάφειά του (Κεφάλαιο 1), μια διεξοδική βιβλιογραφική ανασκόπηση (Κεφάλαιο 2), μια περιγραφή της προτεινόμενης αρχιτεκτονικής (Κεφάλαιο 3) και της υλοποίησης του συστήματος (Κεφάλαιο 4), μια αξιολόγηση των ευρημάτων και τελικά σχόλια με συστάσεις για μελλοντική εργασία (Κεφάλαιο 5).

ΚΕΦΑΛΑΙΟ 2: Θεωρητικό Υπόβαθρο και Βιβλιογραφική Επισκόπηση

2.1 Μορφές και πηγές ετερογένειας στο IoT

Η ποικιλία συσκευών, πρωτοκόλλων, μορφών δεδομένων και πεδίων εφαρμογής που περιλαμβάνει το οικοσύστημα του Διαδικτύου των Πραγμάτων (IoT) το καθιστά φυσικά ετερογενές. Η ετερογένεια στο IoT μπορεί να χωριστεί σε τέσσερις τύπους: συντακτικές, σημασιολογικές, δομικές και παραλλαγές σε επίπεδο πλατφόρμας. Σε συστήματα ροής δεδομένων πραγματικού χρόνου, η ετερογενής διαχείριση γίνεται πιο σημαντική, καθώς αυτές οι διαφορές προσθέτουν πολυπλοκότητα στην επεξεργασία δεδομένων, την ενσωμάτωση και την εξαγωγή πληροφοριών. Όταν οι συσκευές IoT δημιουργούν δεδομένα σε πολλές μορφές - συμπεριλαμβανομένων JSON, XML, CSV ή δυαδικών - αυτό οδηγεί σε συντακτική ετερογένεια. Η ανάλυση και η ερμηνεία των δεδομένων με συνέπεια σε όλα τα συστήματα καθίσταται πιο δύσκολη από αυτό το είδος ασυμφωνίας (Kumar et al., 2021). Για παράδειγμα, τα δεδομένα θερμοκρασίας δύο διαφορετικών κατασκευαστών αισθητήρων μπορούν να έχουν διαφορετικά ονόματα πεδίων, μονάδες - Κελσίου έναντι Φαρενάιτ - ή μορφές χρονικής σήμανσης. Η αντιμετώπιση της συντακτικής ποικιλομορφίας συχνά απαιτεί ενδιάμεσα περιτυλίγματα δεδομένων ή μεθόδους κανονικοποίησης (Adnan & Akbar, 2019).

Αντίθετα, η σημασιολογική ετερογένεια η διακύμανση στην έννοια και την ερμηνεία των πεδίων δεδομένων. Δύο σύνολα δεδομένων μπορεί να χρησιμοποιούν τη λέξη «temp» για να αναφερθούν σε διακριτές μεταβλητές - θερμοκρασία περιβάλλοντος και θερμοκρασία CPU, αντίστοιχα - κάτι που θα μπορούσε να προκαλέσει παρεξηγήσεις κατά την ενσωμάτωση (Singh & Siwach, 2022). Συχνά προτείνονται ως λύσεις για αυτό το είδος ποικιλομορφίας οι οντολογίες και τα συστήματα σημασιολογικής σχολίασης, τα οποία θα επέτρεπαν πιο ουσιαστική ευθυγράμμιση και συλλογισμό δεδομένων (Kolozali et al., 2014). Η δομική ετερογένεια είναι ο τρόπος με τον οποίο δομούνται τα δεδομένα μέσα σε κάθε πηγή. Για παράδειγμα, μια συσκευή μπορεί να μεταδίδει θερμοκρασία, υγρασία και θέση σε επίπεδη μορφή. Μια άλλη θα μπορούσε να τα συμπεριλάβει μέσα σε ένθετα αντικείμενα ή ομαδοποιημένους πίνακες (Putrama & Martinek, 2024).

Αυτές οι διαφορές απαιτούν δυναμικό και αποτελεσματικό χειρισμό διαφορετικών σχημάτων μέσω προσαρμοστικών μοντέλων δεδομένων.

Η ετερογένεια πλατφόρμας, η οποία περιλαμβάνει παραλλαγές σε λειτουργικά συστήματα, πρωτόκολλα δικτύου (π.χ., MQTT, CoAP, HTTP) και πρότυπα επικοινωνίας (π.χ., ZigBee, LoRa, Wi-Fi), είναι ένα άλλο σημαντικό είδος. Ειδικά όταν συνδυάζονται δεδομένα από επίπεδα άκρων, ομίχλης και νέφους, αυτό το επίπεδο μεταβλητότητας έχει σημαντικό αντίκτυπο στη διαλειτουργικότητα και την επεξεργασία ροής σε πραγματικό χρόνο (Corral-Plaza et al., 2020). Προτείνεται ότι οι λύσεις middleware και οι κατανεμημένες αρχιτεκτονικές εξαλείφουν τέτοια κενά αφαιρώντας τις υποκείμενες διαδικασίες επικοινωνίας (Akanbi & Masinde, 2020). Τα συστήματα IoT επηρεάζονται επίσης από την χρονική ετερογένεια, καθώς ξεχωριστές συσκευές λειτουργούν με ποικίλους ρυθμούς παραγωγής δεδομένων ή έχουν καθυστερήσεις που προκαλούνται από περιορισμένους πόρους ενέργειας ή εύρους ζώνης. Αυτό έχει ως αποτέλεσμα ασύγχρονες αφίξεις δεδομένων και δυσκολίες στην ακριβή αντιστοίχιση εγγραφών χρονοσειρών (Ramachandran et al., 2024). Η υπέρβαση αυτού του προβλήματος εξαρτάται από αποτελεσματικές τεχνικές συγχρονισμού ροής και προσωρινής αποθήκευσης.

Η ποικιλομορφία σε επίπεδο εφαρμογής δείχνει επίσης ποικιλομορφία στις ανάγκες που αφορούν συγκεκριμένα πεδία. Ο έξυπνος εξοπλισμός υγειονομικής περίθαλψης επικεντρώνεται σε ζωτικά σημεία των ασθενών, ενώ τα έξυπνα γεωργικά συστήματα μπορούν να δώσουν προτεραιότητα στα δεδομένα υγρασίας και φωτός του εδάφους. Αυτή η ποικιλομορφία που αφορά συγκεκριμένα πεδία περιπλέκει την ενσωμάτωση και την ενοποιημένη αναπαράσταση γνώσης, καθώς δεν είναι όλα τα χαρακτηριστικά των δεδομένων καθολικά εφαρμόσιμα ή ερμηνεύσιμα σε όλες τις εφαρμογές (Marshoodulla & Saha, 2023). Επιπρόσθετα, η συνέπεια των ανακτημένων πληροφοριών επηρεάζεται σε μεγάλο βαθμό από την ετερογένεια της ποιότητας των δεδομένων. Οι εφαρμογές στον πραγματικό κόσμο συχνά περιλαμβάνουν ασυνεπείς ρυθμούς δειγματοληψίας, ελλείποντα αποτελέσματα, θόρυβο και δυσλειτουργίες αισθητήρων. Για να μειωθεί ο

αντίκτυπος τέτοιων εισροών χαμηλής ποιότητας, απαιτούνται προηγμένες μέθοδοι προεπεξεργασίας και φιλτραρίσματος (Gama & Aguilar-Ruiz, 2007).

Από την άποψη της εξαγωγής γνώσης, η διαχείριση αυτού του ευρέος φάσματος ετερογένειας καθίσταται κρίσιμη για την εγγύηση της ορθότητας και της αξιοπιστίας των πληροφοριών που λαμβάνονται από τα δεδομένα ροής. Οι προτεινόμενες μέθοδοι για την προσαρμοστική μάθηση από τις εξελισσόμενες πηγές και μορφές δεδομένων περιλαμβάνουν τη μάθηση σε σύνολα, τα πιθανοτικά μοντέλα και την δυναμική εξόρυξη ροής (van Rijn et al., 2017). Επιπλέον, οι τεχνικές που χρησιμοποιούν Γκαουσιανές Διαδικασίες δείχνουν πολλά υποσχόμενα αποτελέσματα στον έλεγχο της αβεβαιότητας και της απρόβλεπτης ικανότητας σε ποικίλες ροές δεδομένων (Ghoshal & Roberts, 2016). Η ετερογένεια που σχετίζεται με την επεκτασιμότητα εμφανίζεται καθώς τα συστήματα IoT αναπτύσσονται. Περισσότερα ζητήματα απόδοσης και ολοκλήρωσης προκύπτουν από τον αυξανόμενο αριθμό συσκευών και κόμβων δικτύου, γεγονός που ενισχύει την ποικιλία και την ποσότητα των δεδομένων. Οι αρχιτεκτονικές πρέπει να είναι αποτελεσματικές και ανεκτικές σε σφάλματα, σε κλίμακα, καθώς και να παρέχουν σημασιολογική και δομική ομαλοποίηση (Sarikoz, 2023). Επομένως, η ετερογένεια στο IoT εκδηλώνεται με πολλούς διαφορετικούς τρόπους, από αποκλίσεις στη μορφή και το σχήμα έως πιο έντονες σημασιολογικές και συμφραζόμενες παραλλαγές. Τόσο οι φυσικές ιδιότητες των συσκευών όσο και η λογική ποικιλία των αναγκών των εφαρμογών προκαλούν αυτές τις διαφορές. Η ακριβής και ουσιαστική εξαγωγή πληροφοριών από διάσπαρτες ροές δεδομένων IoT εξαρτάται από τον αποτελεσματικό χειρισμό αυτής της πολυπλοκότητας.

2.2 Προσεγγίσεις ενοποίησης δεδομένων (τεχνικές, μοντέλα, πρότυπα)

Η ουσιαστική ανάλυση και λήψη αποφάσεων εξαρτώνται από την ενσωμάτωση ετερογενών δεδομένων IoT. Οι μέθοδοι ενσωμάτωσης δεδομένων πρέπει να είναι σε θέση να εναρμονίζουν διαφορετικές ροές δεδομένων σε πραγματικό χρόνο, διατηρώντας παράλληλα το πλαίσιο και μειώνοντας την απώλεια πληροφοριών, δεδομένης της ποικιλίας στις μορφές δεδομένων, τα πρωτόκολλα και τη σημασιολογία σε όλες τις πηγές IoT. Αυτές οι μέθοδοι περιλαμβάνουν πολλούς

τομείς, όπως η συντακτική ενοποίηση, η σημασιολογική χαρτογράφηση, τα συστήματα middleware και οι διαδικασίες τυποποίησης.

Η εναρμόνιση δεδομένων - η διαδικασία μετατροπής δεδομένων από πολλές διαφορετικές πηγές σε μια ομοιόμορφη δομή και μορφή - είναι μια βασική μέθοδος για την ενσωμάτωση. Η εναρμόνιση περιλαμβάνει την ευθυγράμμιση σχήματος, τη μετατροπή μονάδων, τη μετονομασία πεδίων και την επίλυση ασυμφωνιών συνόλου δεδομένων (Kumar et al., 2021). Οι αγωγοί ETL (Εξαγωγή, Μετασχηματισμός, Φόρτωση) ή τα αντίστοιχα streaming όπως το Apache Kafka και το Apache NiFi χρησιμοποιούνται συχνά για την ολοκλήρωση αυτής της διαδικασίας, επιτρέποντας τον συνεχή μετασχηματισμό δεδομένων καθώς καταναλώνονται ροές (Adnan & Akbar, 2019). Η κατανόηση της πολυπλοκότητας της υποκείμενης ετερογένειας υποβοηθείται σε μεγάλο βαθμό από τεχνικές που βασίζονται σε middleware. Αυτά τα συστήματα παρέχουν API και τυπικά πρωτόκολλα επικοινωνίας για να επιτρέπουν την ομαλή μεταφορά δεδομένων, λειτουργώντας ως ενδιάμεσοι μεταξύ συσκευών και εφαρμογών. Για παράδειγμα, το middleware επεξεργασίας κατανεμημένων ροών επιτρέπει τη συλλογή, το σχολιασμό και το φιλτράρισμα δεδομένων σε πραγματικό χρόνο σε γεωγραφικά διασκορπισμένες πηγές, μειώνοντας έτσι σημαντικά το φορτίο ενσωμάτωσης στο επίπεδο εφαρμογής (Akanbi & Masinde, 2020).

Από άποψη μοντελοποίησης, έχουν προταθεί πολλά μοντέλα ενσωμάτωσης, συμπεριλαμβανομένων μοντέλων που βασίζονται σε μεσολαβητές, στρατηγικών αποθήκευσης δεδομένων και ενσωμάτωσης που βασίζεται σε οντολογίες. Προσφέροντας ευελιξία σε βάρος της καθυστέρησης και της πολυπλοκότητας, τα συστήματα που βασίζονται σε μεσολαβητές υποβάλλουν δυναμικά ερωτήματα σε πολλές πηγές και συγκεντρώνουν απαντήσεις. Αντίθετα, τα μοντέλα αποθήκης δεδομένων προσφέρουν ταχύτερες αναζητήσεις, αλλά υποφέρουν από χαμηλή ευελιξία στην ανάπτυξη σχημάτων, καθώς ενοποιούν δεδομένα σε ένα ενιαίο αποθετήριο με ένα συνεπές σχήμα (Putrama & Martinek, 2024). Όλο και πιο συνηθισμένες για τη διαχείριση της σημασιολογικής ετερογένειας είναι οι προσεγγίσεις που βασίζονται σε οντολογίες. Χρησιμοποιώντας τυπική λογική και πρότυπα σημασιολογικού ιστού όπως RDF και OWL, αυτά τα μοντέλα παρέχουν κοινά λεξιλόγια και συνδέσεις μεταξύ των στοιχείων δεδομένων. Τα συστήματα

μπορούν να διεξάγουν συλλογισμό, να συνάγουν συνδέσεις και να συνδυάζουν σημασιολογικά παρόμοιο αλλά συντακτικά διαφορετικό υλικό συνδέοντας οντολογίες που αφορούν συγκεκριμένους τομείς με δεδομένα (Singh & Siwach, 2022). Στους τομείς του IoT χρησιμοποιούνται συνήθως εργαλεία όπως το Protégé και πλαίσια όπως η οντολογία SSN (Semantic Sensor Network) (Kolozali et al., 2014).

Μια άλλη κατηγορία μεθόδων είναι η χαρτογράφηση σχημάτων και η σύνδεση εγγραφών, στις οποίες οι αλγόριθμοι βρίσκουν και αντιστοιχίζουν σχετικά χαρακτηριστικά σε πολλά σύνολα δεδομένων. Για να βρουν χαρτογραφήσεις μεταξύ διαφορετικών σχημάτων, αυτές οι προσεγγίσεις χρησιμοποιούν ευρετικές μεθόδους, μέτρα ομοιότητας - όπως Jaccard και Levenshtein - και μοντέλα μηχανικής μάθησης (Ramachandran et al., 2024). Σε συστήματα πραγματικού χρόνου, όπου τα εισερχόμενα δεδομένα πρέπει να κατανοηθούν και να αντιμετωπιστούν άμεσα, η χαρτογράφηση σχημάτων καθίσταται πολύ σημαντική. Η ενσωμάτωση έχει επίσης επηρεαστεί σε μεγάλο βαθμό από πρωτοβουλίες τυποποίησης. Μορφές δεδομένων όπως JSON, XML και δυαδικές παραλλαγές όπως το CBOR παρέχουν μια κοινή βάση για συντακτική συμβατότητα, πρωτόκολλα όπως MQTT, CoAP και HTTP παρέχουν συνεπή επικοινωνία. Σε σημασιολογικό επίπεδο, έργα όπως το Web of Things (WoT) του W3C παρέχουν ομοιόμορφα μεταδεδομένα και περιγραφές στοιχείων για να επιτρέψουν την ανακάλυψη και τη διαλειτουργικότητα συσκευών (Corral-Plaza et al., 2020). Αυτές οι οδηγίες αυξάνουν την επεκτασιμότητα σε ποικίλα περιβάλλοντα και μειώνουν τα προβλήματα διαλειτουργικότητας μεταξύ των συστημάτων.

Οι προηγμένες μέθοδοι ολοκλήρωσης προσαρμόζονται αυτόματα σε νέες πηγές δεδομένων χρησιμοποιώντας εξόρυξη ροής και μηχανική μάθηση. Για παράδειγμα, τα μοντέλα μάθησης συνόλων μπορούν να ενημερώνουν συνεχώς τις προβλέψεις καθώς εισέρχονται νέα δεδομένα και να αναμειγνύουν πληροφορίες από πολλές ετερογενείς πηγές δεδομένων (van Rijn et al., 2017). Σε δυναμικά περιβάλλοντα IoT όπου οι μορφές και οι πηγές δεδομένων αλλάζουν συχνά, αυτές οι στρατηγικές είναι ιδιαίτερα ωφέλιμες. Επιπλέον, εντοπίζοντας μοτίβα, συσχετίσεις και ανωμαλίες σε πολλές πηγές δεδομένων, οι μηχανές επεξεργασίας σύνθετων συμβάντων (CEP) επιτρέπουν την ολοκλήρωση σε υψηλότερο επίπεδο αφαίρεσης.

Συγκεντρώνοντας και επεξεργαζόμενοι δεδομένα από πολλές συσκευές IoT σε σχεδόν πραγματικό χρόνο χρησιμοποιώντας κανόνες που βασίζονται σε συμβάντα, τα συστήματα CEP παρέχουν αντιδράσεις και πληροφορίες που λαμβάνουν υπόψη το περιβάλλον (Gama & Aguilar-Ruiz, 2007).

Επίσης, προτείνονται συστήματα ολοκλήρωσης βασισμένα στη γνώση που ενσωματώνουν γνώση πεδίου στη διαδικασία ολοκλήρωσης δεδομένων. Πριν τα δεδομένα φτάσουν στη φάση εξαγωγής γνώσης, αυτά τα συστήματα τα ευθυγραμμίζουν και τα εμπλουτίζουν χρησιμοποιώντας συμπεράσματα και κανόνες που λαμβάνουν υπόψη το περιβάλλον. Σε τομείς όπως η υγειονομική περίθαλψη ή η περιβαλλοντική παρακολούθηση, όπου η ερμηνεία ανά πεδίο είναι ζωτικής σημασίας, τέτοια συστήματα είναι πολύ χρήσιμα (Sarikoz, 2023). Η ενσωμάτωση εξακολουθεί να αντιμετωπίζει σημαντικές δυσκολίες παρά τις εξελίξεις. Συνήθη προβλήματα περιλαμβάνουν την επεκτασιμότητα, την ταχύτητα δεδομένων, τις ελλείπουσες πληροφορίες και το κόστος διατήρησης συγχρονισμένων μοντέλων ευθυγράμμισης. Πρόσφατες μελέτες υπογραμμίζουν την ανάγκη για υβριδικές στρατηγικές που συνδυάζουν τεχνικές βασισμένες σε κανόνες, στατιστικές και βασισμένες στη μάθηση για την προσαρμοστική διαχείριση των δραστηριοτήτων ενσωμάτωσης (Ghoshal & Roberts, 2016). Συνοψίζοντας, μια πολυεπίπεδη στρατηγική που περιλαμβάνει συντακτική ομαλοποίηση, σημασιολογική ευθυγράμμιση, επεκτάσιμες αρχιτεκτονικές και προσαρμοστικούς μηχανισμούς μάθησης είναι απαραίτητη για την αποτελεσματική ενσωμάτωση δεδομένων σε περιβάλλοντα IoT. Η υιοθέτηση λύσεων middleware, σημασιολογικών μοντέλων και τυποποιημένων πρωτοκόλλων βοηθά στη γεφύρωση του χάσματος ετερογένειας και επιτρέπει την αποτελεσματική εξαγωγή πληροφοριών από απομακρυσμένες, ποικίλες πηγές δεδομένων.

2.3 Σημασιολογική διαλειτουργικότητα και οντολογίες

Στο οικοσύστημα του Διαδικτύου των Πραγμάτων (IoT), όπου διάφορες συσκευές και πλατφόρμες πρέπει να διαβάζουν και να ερμηνεύουν δεδομένα με συνέπεια παρά τις διαφορές στη δομή, τη μορφή και το πλαίσιο, η σημασιολογική διαλειτουργικότητα είναι ζωτικής σημασίας. Σε αντίθεση με τη συντακτική ή δομική διαλειτουργικότητα, η οποία δίνει έμφαση στη μορφοποίηση δεδομένων και

στα πρωτόκολλα πρόσβασης, η σημασιολογική διαλειτουργικότητα εγγυάται ότι η έννοια των δεδομένων είναι σταθερή σε όλα τα συστήματα και τους τομείς (Singh & Siwach, 2022). Στα πλαίσια του Διαδικτύου των Πραγμάτων, οι οντολογίες είναι οι κύριοι παράγοντες που διευκολύνουν τη σημασιολογική διαλειτουργικότητα. Μια οντολογία επιτρέπει σε διάφορα συστήματα να κατανοούν και να συλλογίζονται τα δεδομένα με τυποποιημένο τρόπο χρησιμοποιώντας μια κοινή γλώσσα και ένα σύνολο σχέσεων για ένα συγκεκριμένο θέμα. Με τη χρήση οντολογιών, οι υπολογιστές μπορούν να συνδέουν σημασιολογικά συγκρίσιμες λέξεις σε πολλές διαφορετικές πηγές δεδομένων, καθώς ενθαρρύνουν τη συνέπεια στην ερμηνεία (Kolozali et al., 2014). Μια οντολογία μπορεί να συνδέσει και τα δύο με την ιδέα της θερμοκρασίας περιβάλλοντος εάν, για παράδειγμα, μία συσκευή χρησιμοποιεί τον όρο "θερμοκρασία" και μία άλλη "θερμ.", μπορούν και οι δύο να συνδεθούν με την έννοια της θερμοκρασίας περιβάλλοντος.

Αναπτυγμένη από το W3C, η οντολογία Semantic Sensor Network (SSN) είναι ένα ευρέως χρησιμοποιούμενο πρότυπο που προσφέρει ένα πλαίσιο για τον ορισμό αισθητήρων, των παρατηρήσεών τους και των σχετικών πληροφοριών. Σχολιασμένη με σημασιολογικές ετικέτες, αυτό το παράδειγμα επιτρέπει το σχολιασμό ροής δεδομένων IoT, επιτρέποντας έτσι την αυτόματη ανακάλυψη, φιλτράρισμα και ενσωμάτωση (Kumar et al., 2021). Συνήθως, οι οντολογίες υλοποιούνται χρησιμοποιώντας το Resource Description Framework (RDF) και τη Web Ontology Language (OWL), οι οποίες επιτρέπουν ορισμούς αναγνώσιμους από μηχανές και δυνατότητες συλλογισμού. Η αποτελεσματική σύντηξη δεδομένων και η εξαγωγή πληροφοριών εξαρτώνται από το σημασιολογικό σχολιασμό των ροών δεδομένων IoT σε πραγματικό χρόνο. Μια προσέγγιση βασισμένη στη γνώση για το σχολιασμό ροής δεδομένων επιτρέπει στα συστήματα IoT να επισημαίνουν και να ευθυγραμμίζουν αυτόματα τα δεδομένα καθώς παράγονται ενσωματώνοντας οντολογίες σε αγωγούς επεξεργασίας δεδομένων (Kolozali et al., 2014). Σε δυναμικά περιβάλλοντα όπως οι έξυπνες πόλεις και οι εφαρμογές υγειονομικής περίθαλψης, αυτή η προσέγγιση επιτρέπει τη λήψη αποφάσεων με επίγνωση του περιβάλλοντος (Corral-Plaza et al., 2020).

Από την άλλη πλευρά, η δημιουργία και η συντήρηση οντολογιών που αφορούν συγκεκριμένους τομείς μπορεί να είναι δύσκολη και δαπανηρή. Συχνά, οι ειδικοί

σε τομείς είναι απαραίτητοι για να καθορίσουν ακριβείς ιδέες, συνδέσεις και όρια. Αυτές οι οντολογίες πρέπει να ενημερώνονται τακτικά ώστε να αντικατοπτρίζουν τις τεχνολογικές ή τις συμφραζόμενες αλλαγές (Adnan & Akbar, 2019). Επιπλέον, ιδιαίτερα όταν συνδυάζονται συστήματα IoT από διάφορους κατασκευαστές ή βιομηχανίες, η εναρμόνιση πολλών οντολογιών σε διάφορους τομείς - γνωστή και ως αντιστοίχιση ή χαρτογράφηση οντολογιών - παραμένει μια τεχνολογική δυσκολία (Putrama & Martinek, 2024). Πρόσφατες μελέτες διερευνούν τη σημασιολογική χαρτογράφηση και την αυτοματοποιημένη δημιουργία οντολογιών χρησιμοποιώντας μεθόδους μηχανικής μάθησης. Με τη χρήση στατιστικής ανάλυσης, ομαδοποίησης και επεξεργασίας φυσικής γλώσσας, αυτές οι μέθοδοι δημιουργούν οντολογικές ιδέες και συσχετίσεις από ακατέργαστα δεδομένα, μειώνοντας έτσι το φόρτο στους ανθρώπινους ειδικούς (Ramachandran et al., 2024). Τέτοιες ημιαυτοματοποιημένες τεχνικές θα μπορούσαν να επιταχύνουν τη χρήση σημασιολογικών τεχνολογιών σε εγκαταστάσεις IoT μεγάλης κλίμακας.

Τα γραφήματα γνώσης, τα οποία δείχνουν οντότητες και τις αλληλεπιδράσεις τους με έναν συνδεδεμένο, οργανωμένο τρόπο, αποτελούν μια άλλη σημαντική χρήση των οντολογιών στο IoT. Με τη χρήση οντολογικών πλαισίων, τα γραφήματα γνώσης συνδέουν σημεία δεδομένων σε πολλές πηγές, βελτιώνοντας έτσι την σημασιολογική αναζήτηση, τη συλλογιστική και την ανάλυση (Singh & Siwach, 2022). Χρησιμοποιούνται όλο και περισσότερο σε τομείς όπως ο βιομηχανικός αυτοματισμός και η διαχείριση της εφοδιαστικής αλυσίδας, παρακολουθούν τις εξαρτήσεις σε πραγματικό χρόνο και βοηθούν στη λήψη σοφών αποφάσεων. Η ομαλή ενσωμάτωση και κατανόηση ποικίλων δεδομένων IoT εξαρτάται από τη σημασιολογική διαλειτουργικότητα που καθίσταται δυνατή από τις οντολογίες. Αν και η δημιουργία και η ευθυγράμμιση οντολογιών εξακολουθούν να παρουσιάζουν δυσκολίες, τα πλεονεκτήματα της ομοιόμορφης αναπαράστασης νοήματος και της αυτοματοποιημένης συλλογιστικής είναι απολύτως ζωτικής σημασίας για την απόκτηση αξιοποιήσιμων πληροφοριών από περίπλοκα, διασκορπισμένα συστήματα IoT (Sarikoz, 2023).

2.4 Συγχώνευση δεδομένων: επίπεδη, ιεραρχική, δέντρα, πλέγμα, χωρίς δομή

Ο όρος «σύντηξη δεδομένων» περιγράφει την τεχνική συνδυασμού πολλών πηγών δεδομένων για την απόκτηση πιο συνεπών, σωστών και σχετικών πληροφοριών. Η σύντηξη δεδομένων είναι κρίσιμη στο IoT για τον συνδυασμό διαφορετικών εισροών σε συνεκτικές δομές γνώσης, καθώς τα δεδομένα προέρχονται από ένα ευρύ φάσμα ετερογενών αισθητήρων, συσκευών και συστημάτων. Μεταξύ των διαφόρων μοντέλων σύντηξης που προτείνονται - επίπεδα, ιεραρχικά, δενδροειδή, πλέγματα και μη δομημένα - το καθένα έχει μοναδικά οφέλη και χρήσεις.

Η επίπεδη σύντηξη είναι η διαδικασία συνδυασμού πηγών δεδομένων στο ίδιο επίπεδο αφαίρεσης χωρίς να λαμβάνονται υπόψη τυχόν ιεραρχικές ή ένθετες συνδέσεις. Αν και θα μπορούσε να χάσει σημασιολογικό ή συμπραζόμενο πλούτο, αυτή η προσέγγιση είναι γρήγορη και εύκολη. Η συγχώνευση τιμών θερμοκρασίας και υγρασίας από πολλούς αισθητήρες σε ένα επίπεδο σύνολο δεδομένων, για παράδειγμα, προσφέρει μια ενοποιημένη προοπτική, αλλά καμία κατανόηση των συνδέσμων ή της αιτιότητας (Adnan & Akbar, 2019). Τα ομοιογενή δίκτυα αισθητήρων ή τα πρώτα επίπεδα συσσωμάτωσης δεδομένων είναι τα πιο κατάλληλα για επίπεδη σύντηξη (Gama & Aguilar-Ruiz, 2007).

Αντίθετα, η ιεραρχική σύντηξη ταξινομεί τα δεδομένα σε πολυεπίπεδες δομές σύμφωνα με γεωγραφικούς, χρονικούς ή λειτουργικούς συνδέσμους. Σε εφαρμογές έξυπνων πόλεων και βιομηχανικών εφαρμογών, όπου τα δεδομένα από διάφορα επίπεδα - συμπεριλαμβανομένων αισθητήρων, πυλών και διακομιστών cloud - πρέπει να ενσωματωθούν με συνοχή, αυτή η έννοια είναι πολύ ωφέλιμη (Putrama & Martinek, 2024). Υποστηρίζοντας την επεκτασιμότητα και την αρθρωτότητα, η ιεραρχική σύντηξη επιτρέπει στα συστήματα να χειρίζονται δεδομένα τοπικά στην άκρη και στη συνέχεια να παρέχουν συγκεντρωτικές περιλήψεις σε κεντρικά συστήματα για περαιτέρω ανάλυση (Akanbi & Masinde, 2020). Τα μοντέλα σύντηξης που βασίζονται σε δέντρα είναι ένα ιδιαίτερο είδος ιεραρχικής σύντηξης στην οποία τα δεδομένα οργανώνονται σε ένα δέντρο με κόμβους που υποδηλώνουν ιδιότητες ή συσκευές και κλάδους που υποδηλώνουν συνδέσεις. Ιδιαίτερα όταν χρησιμοποιούνται μετρήσεις ομοιότητας όπως οι αποστάσεις Hamming ή

Levenshtein για την ομαδοποίηση συγκρίσιμων ετικετών, αυτή η τεχνική είναι ωφέλιμη για την οργάνωση σημασιολογικά σχολιασμένων δεδομένων και επιτρέπει την αποτελεσματική υποβολή ερωτημάτων (Ramachandran et al., 2024). Η διαχείριση της ετερογένειας των ετικετών σε διασκορπισμένες συσκευές γίνεται ιδιαίτερα καλά χρησιμοποιώντας αρχιτεκτονικές δέντρων (Kumar et al., 2021). Χρησιμοποιείται συνήθως σε εφαρμογές όπως η παρακολούθηση του περιβάλλοντος, όπου οι μετρήσεις αισθητήρων είναι γεωγραφικά κατανομημένες σε μια γεωγραφική περιοχή, η σύντηξη που βασίζεται σε πλέγμα οργανώνει τα δεδομένα σε αναπαραστάσεις τύπου μήτρας. Η σύντηξη πλέγματος επιτρέπει την χωρική παρεμβολή, την ανίχνευση ανωμαλιών και την οπτικοποίηση γραφημάτων θερμότητας και περιγράμματος (Akanbi & Masinde, 2020). Επίσης, επιτρέπει την κατανομημένη αποθήκευση και την παράλληλη επεξεργασία, δύο από τα οποία είναι ζωτικής σημασίας για τη διαχείριση δεδομένων μεγάλης κλίμακας από έξυπνα γεωργικά ή μετεωρολογικά συστήματα (Corral-Plaza et al., 2020).

Ένα από τα πιο δύσκολα είδη σύντηξης δεδομένων είναι η μη δομημένη, καθώς σημαίνει συγχώνευση δεδομένων που στερούνται προκαθορισμένου προτύπου ή μορφής. Αυτό καλύπτει ημι-δομημένες εξόδους αισθητήρων, δεδομένα κειμένου από μέσα κοινωνικής δικτύωσης ή αρχεία καταγραφής και υλικό πολυμέσων - εικόνες, βίντεο, ήχο. Τα μη δομημένα δεδομένα μετατρέπονται σε δομημένες μορφές κατάλληλες για ενσωμάτωση χρησιμοποιώντας μεθόδους όπως η βαθιά μάθηση, η επεξεργασία φυσικής γλώσσας και η εξαγωγή χαρακτηριστικών (Ghoshal & Roberts, 2016). Η πολυπλοκότητα της μη δομημένης σύντηξης τονίζει την ανάγκη για ισχυρά συστήματα προεπεξεργασίας, εμπλουτισμού πληροφοριών και σχολιασμού (Kumar et al., 2021). Η επιτυχής σύντηξη δεδομένων σε συστήματα IoT απαιτεί επίσης επίγνωση του περιβάλλοντος, που σημαίνει ότι το σύστημα πρέπει να λαμβάνει υπόψη το περιβάλλον, τον χρόνο και την αξιοπιστία της πηγής δεδομένων των δεδομένων που συνδυάζονται. Η αγνόηση στοιχείων περιβάλλοντος, για παράδειγμα, μπορεί να οδηγήσει σε εσφαλμένα αποτελέσματα, ενώ ο συνδυασμός δεδομένων από ελαττωματικούς αισθητήρες θα μπορούσε να προσθέσει θόρυβο ή μεροληψία (Sarikoz, 2023). Για την παροχή προσαρμοστικής λήψης αποφάσεων σε πραγματικό χρόνο, οι πλατφόρμες middleware και οι λύσεις

edge AI περιλαμβάνουν προοδευτικά μεθόδους σύντηξης που βασίζονται στο περιβάλλον (Marshoodulla & Saha, 2023).

Οι μέθοδοι σύντηξης δεδομένων είναι απαραίτητες για την εναρμόνιση των ανόμοιων δεδομένων και τη διευκόλυνση της έξυπνης ανάλυσης στο IoT. Κάθε μοντέλο σύντηξης βοηθά στο κλείσιμο των κενών που δημιουργούνται από την ποικιλία των συσκευών, την πολυπλοκότητα των δεδομένων και τη σημασιολογική διακύμανση από επίπεδες και ιεραρχικές δομές σε μη δομημένους αγωγούς επεξεργασίας. Η επιλογή της κατάλληλης τεχνικής σύντηξης βασίζεται στους στόχους της εφαρμογής, στους περιορισμούς του συστήματος και στην ποσότητα αφαίρεσης που απαιτείται για την εξαγωγή γνώσης κατάντη (Kolozali et al., 2014).

2.5 Υφιστάμενες μελέτες για εξαγωγή γνώσης σε ετερογενή IoT περιβάλλοντα

Λόγω της εκθετικής επέκτασης των διασκορπισμένων πηγών δεδομένων και της ανάγκης για ανάλυση σε πραγματικό χρόνο, η εξαγωγή πληροφοριών από ετερογενή περιβάλλοντα IoT έχει προσελκύσει σημαντικό επιστημονικό ενδιαφέρον. Η τρέχουσα έρευνα δίνει έμφαση σε πολλές πτυχές, όπως η επεξεργασία ροής, η εναρμόνιση δεδομένων, η σημασιολογική σχολίαση και η μηχανική μάθηση για την αυτοματοποιημένη παραγωγή πληροφοριών.

Τα συστήματα επεξεργασίας ροής σε πραγματικό χρόνο που προορίζονται για τη διαχείριση διαφόρων εισροών IoT αποτελούν ένα σημαντικό πεδίο μελέτης. Για παράδειγμα, οι Corral-Plaza et al. (2020) παρέχουν ένα αρθρωτό σύστημα που χρησιμοποιεί μια μηχανή επεξεργασίας ροής για την ομαλοποίηση και τον χειρισμό δεδομένων σε κίνηση, υποστηρίζοντας έτσι την ενσωμάτωση πολλών ετερογενών πηγών. Η προσθήκη ενοτήτων σημασιολογικού εμπλουτισμού που σχολιάζουν εισερχόμενες ροές δεδομένων, η μέθοδός τους επιτρέπει την επεκτασιμότητα και την προσαρμογή, ενισχύοντας έτσι την εξαγωγή γνώσης. Οι Akanbi και Masinde (2020) παρέχουν επίσης ένα σύστημα middleware για την παρακολούθηση του περιβάλλοντος που επιτρέπει την κατανεμημένη επεξεργασία δεδομένων αισθητήρων σε μεγάλες πλατφόρμες δεδομένων. Η εργασία τους υπογραμμίζει την ανάγκη συμπερίληψης χωροχρονικών πληροφοριών για σωστή ανάλυση με επίγνωση του περιβάλλοντος. Ένας πολυεπίπεδος σχεδιασμός

επιτρέπει στο σύστημα να παρέχει προσαρμοστικό φιλτράρισμα και σύντηξη, επιτρέποντας έτσι την ανακάλυψη γνώσης που εκτείνεται σε πολλούς περιβαλλοντικούς τομείς.

Ένα άλλο κοινό θέμα είναι η εξαγωγή γνώσης χρησιμοποιώντας σημασιολογικά μοντέλα και οντολογίες. Χρησιμοποιώντας οντολογίες τομέα, οι Kolozali et al. (2014) δείχνουν πώς μπορεί να επιτευχθεί σημασιολογική σχολίαση σε πραγματικό χρόνο των ροών δεδομένων IoT, παράγοντας έτσι σχετικές πληροφορίες μέσω μηχανών συλλογισμού. Αυτή η εργασία θέτει τις βάσεις δείχνοντας πώς η σημασιολογική διαλειτουργικότητα είναι απαραίτητη για να γεφυρωθεί το χάσμα μεταξύ της οργανωμένης γνώσης και των ακατέργαστων δεδομένων αισθητήρων. Πολλές έρευνες χρησιμοποιούν τεχνικές μηχανικής μάθησης - ειδικά μοντέλα μάθησης συνόλων και πιθανοτικών μοντέλων - για τη διαχείριση της αβεβαιότητας και της διακύμανσης σε ετερογενή δεδομένα. Χρησιμοποιώντας τεχνικές συνόλων που μαθαίνουν από ροές δεδομένων με μεταβαλλόμενα χαρακτηριστικά, οι van Rijn et al. (2017) παρέχουν ένα πλαίσιο για την εκτίμηση της διαδικτυακής απόδοσης. Η μέθοδός τους διατηρεί εξαιρετική ακρίβεια πρόβλεψης ακόμη και υπό παρέκκλιση ιδεών, ενημερώνοντας συνεχώς τον εαυτό της με εισερχόμενα δεδομένα και προσαρμόζοντας τον εαυτό της στις νέες τάσεις.

Οι διαδικασίες αυτές έχουν επίσης χρησιμοποιηθεί σε καταστάσεις εξαγωγής γνώσης για τη διαχείριση του θορύβου και της ασάφειας ανάγνωσης αισθητήρων. Χρησιμοποιώντας γκαουσιανές διαδικασίες, οι Ghoshal και Roberts (2016) παρέχουν ένα μοντέλο για προγνωστική ανάλυση σε ροές οικονομικών δεδομένων που μπορεί να τροποποιηθεί για ετερογενή περιβάλλοντα IoT όπου η αβεβαιότητα και η αραιότητα των δεδομένων είναι συχνά προβλήματα. Οι Ramachandran et al. (2024) διερευνούν υβριδικές αρχιτεκτονικές για την εξαγωγή γνώσης από απλωμένες πηγές δεδομένων. Η προσέγγισή τους ερμηνεύει ροές δεδομένων με σημασιολογικές ετικέτες χρησιμοποιώντας στατιστικά μοντέλα σε συνδυασμό με συλλογισμό βασισμένο σε κανόνες. Σε έξυπνα συστήματα όπως η κατασκευή και η υγειονομική περίθαλψη, αυτή η διπλή μέθοδος επιτρέπει την ανίχνευση συμβάντων και την αναγνώριση μοτίβων υψηλότερου επιπέδου.

Οι μη δομημένες και ημιδομημένες ρυθμίσεις έχουν επίσης καλυφθεί για την εξαγωγή γνώσης. Οι Adnan και Akbar (2019) εφιστούν την προσοχή στις αδυναμίες των παραδοσιακών μεθόδων εξαγωγής στον χειρισμό μη δομημένων δεδομένων IoT, όπως φωτογραφίες, αρχεία καταγραφής ή ομιλία. Προτείνουν την προσθήκη τεχνικών υπολογιστικής όρασης και επεξεργασίας φυσικής γλώσσας για την αύξηση του εύρους των εξαγόμενων πληροφοριών σε περίπλοκα συστήματα. Οι Kumar et al. (2021) παρέχουν επίσης μια μεθοδική βιβλιογραφική μελέτη σχετικά με τις τεχνικές εναρμόνισης δεδομένων που επιτρέπουν την εξαγωγή γνώσης. Η μελέτη τους εντοπίζει τυπικά εμπόδια, συμπεριλαμβανομένων των ασυμβατοτήτων των σχημάτων και των περιορισμών σε πραγματικό χρόνο, ταξινομώντας τις τρέχουσες λύσεις σύμφωνα με μεθοδολογίες μετασχηματισμού, προσεγγίσεις ολοκλήρωσης και επεκτασιμότητα. Ο Sarikoz (2023) εξετάζει τους αγωγούς εξαγωγής γνώσης από άκρο σε άκρο, επισημαίνοντας σημαντικά βήματα όπως η συλλογή δεδομένων, η προεπεξεργασία, η σύντηξη, ο σημασιολογικός εμπλουτισμός και η συλλογιστική. Η μελέτη τονίζει ότι η ισορροπία μεταξύ της ποιότητας των δεδομένων, της υπολογιστικής αποτελεσματικότητας και της σημασιολογικής συνέπειας ορίζει την αποτελεσματική εξαγωγή γνώσης. Η βιβλιογραφία δείχνει μια μεγάλη ποικιλία τεχνικών και εργαλείων για την εξαγωγή πληροφοριών σε ετερογενή περιβάλλοντα IoT. Αυτές οι μελέτες τονίζουν την ανάγκη για μια πολυεπίπεδη στρατηγική που συνδυάζει συντακτικές, σημασιολογικές και στατιστικές μεθόδους για τη μετατροπή ακατέργαστων, διάσπαρτων ροών δεδομένων σε πρακτικές πληροφορίες.

2.6 Κριτική αποτίμηση και ερευνητικό κενό

Παρόλο που έχουν επιτευχθεί αξιοσημείωτες εξελίξεις στον τομέα της εξαγωγής πληροφοριών από ποικίλα συστήματα IoT, εξακολουθούν να υπάρχουν αρκετά κενά και περιορισμοί. Μια αυστηρή ανασκόπηση της τρέχουσας βιβλιογραφίας αποκαλύπτει περιορισμούς σε τομείς όπως η επεκτασιμότητα, η σημασιολογική συνέπεια, ο αυτοματισμός και η επίγνωση του περιβάλλοντος.

Η επεκτασιμότητα της αρχιτεκτονικής ολοκλήρωσης είναι ένας θεμελιώδης περιορισμός. Πολλά προτεινόμενα πλαίσια δείχνουν καλά αποτελέσματα σε ελεγχόμενα περιβάλλοντα, αλλά χάνουν απόδοση όταν επεκτείνονται σε μεγάλες,

πραγματικές εγκαταστάσεις IoT. Για παράδειγμα, τα συστήματα middleware συχνά υποφέρουν από περιορισμούς καθυστέρησης και εύρους ζώνης καθώς αυξάνεται ο αριθμός των συνδεδεμένων συσκευών (Akanbi & Masinde, 2020; Corral-Plaza et al., 2020). Επιπλέον, οι κεντρικές μέθοδοι επεξεργασίας δεν μπορούν πάντα να διαχειριστούν τον εκρηκτικό και υψηλής ταχύτητας χαρακτήρα των ροών δεδομένων IoT και ενδέχεται να δημιουργήσουν μοναδικά σημεία αποτυχίας (Putrama & Martinek, 2024).

Η σημασιολογική ασάφεια και η διαχείριση οντολογιών αποτελούν μια ακόμη δυσκολία. Παρόλο που οι οντολογίες έχουν αποδειχθεί επιτυχημένες στην αντιστοίχιση των νοημάτων των δεδομένων, η δυναμική τους ευελιξία παρεμποδίζεται, καθώς απαιτούν γνώση του τομέα και ανθρώπινη ανάπτυξη. Πολλές τρέχουσες έρευνες δεν διαθέτουν συστήματα για αυτόματη ανάπτυξη οντολογιών ή ευθυγράμμιση οντολογιών σε πολλαπλούς τομείς (Kolozali et al., 2014; Singh & Siwach, 2022). Σε εξαιρετικά δυναμικά ή διατομεακά περιβάλλοντα IoT, συμπεριλαμβανομένων των έξυπνων πόλεων ή των πολυτομεακών συστημάτων υγειονομικής περίθαλψης, αυτό μειώνει τη σημασία τους. Η εξαγωγή γνώσης σε πραγματικό χρόνο στερείται επίσης προσαρμοστικών συστημάτων μάθησης. Αν και οι γκαουσιανές διαδικασίες και η μάθηση σε σύνολο έχουν δείξει πολλά υποσχόμενα αποτελέσματα στη διαχείριση της μεταβλητότητας και της αβεβαιότητας, τα συστήματα IoT πλήρους κλίμακας σπάνια τις χρησιμοποιούν (van Rijn et al., 2017; Ghoshal & Roberts, 2016). Οι περισσότερες από τις υπάρχουσες μέθοδοι εξαρτώνται από προκαθορισμένα μοντέλα και κανονισμούς, τα οποία θα μπορούσαν να καταστούν ξεπερασμένα σε ταχέως μεταβαλλόμενα περιβάλλοντα.

Στους αγωγούς εξαγωγής γνώσης IoT, η διαχείριση μη δομημένων δεδομένων παραμένει ελλιπής. Παρόλο που μελέτες έχουν δείξει τη σημασία της συμπερίληψης οπτικών, ακουστικών και φυσικής γλώσσας στην ανάλυση, πολύ λίγα συστήματα συνδυάζουν με επιτυχία τέτοιες μεθόδους με δομημένα δεδομένα αισθητήρων (Adnan & Akbar, 2019). Η έλλειψη συνεπών συστημάτων που χειρίζονται πολυτροπικά είδη δεδομένων περιορίζει την ολότητα της ανακτημένης γνώσης. Ένα άλλο υποτιμημένο θέμα είναι η επίγνωση του πλαισίου. Οι περισσότερες μέθοδοι ενσωμάτωσης και εξαγωγής αγνοούν τα στοιχεία πλαισίου,

χρόνου και πηγής που μπορεί να επηρεάσουν την ερμηνεία, αντιμετωπίζοντας όλα τα δεδομένα ως εξίσου σημαντικά. Ειδικά σε τομείς κρίσιμους για την ασφάλεια, όπως η διαχείριση καταστροφών ή ο βιομηχανικός έλεγχος, αυτή η απουσία μοντελοποίησης πλαισίου μπορεί να οδηγήσει σε ψευδή ή ασήμαντα ευρήματα (Sarikoz, 2023).

Από άποψη τυποποίησης, οι τρέχουσες πρωτοβουλίες όπως το MQTT και το WoT χειρίζονται μόνο εν μέρει τη σημασιολογική και δομική ποικιλία των δεδομένων IoT. Δεν υπάρχει κανένα συμφωνημένο πρωτόκολλο που να συνδυάζει πλούσια σημασιολογική αναπαράσταση με συντακτική συμβατότητα. Η απουσία κριτηρίων σημασιολογικής διαλειτουργικότητας περιορίζει την αλληλεπίδραση μεταξύ πλατφορμών και ωθεί τους προγραμματιστές να βασίζονται σε εξατομικευμένες λύσεις, αυξάνοντας έτσι το κόστος συντήρησης και μειώνοντας την επαναχρησιμοποίηση (Putrama & Martinek, 2024). Ένας άλλος τομέας με ανεκπλήρωτες προοπτικές είναι ο αυτοματισμός. Οι μέθοδοι χειροκίνητης εισαγωγής ή οι ευρετικές μέθοδοι εξακολουθούν να κυριαρχούν στην κατασκευή οντολογιών, τη χαρτογράφηση σχημάτων και την εξαγωγή χαρακτηριστικών. Παρόλο που υπάρχει μικρή μελέτη σχετικά με την προσθήκη τέτοιων τεχνολογιών σε λειτουργικά συστήματα IoT, ο αυτοματισμός που βασίζεται στην Τεχνητή Νοημοσύνη θα μπορούσε να βοηθήσει στην κάλυψη αυτού του κενού (Ramachandran et al., 2024; Kumar et al., 2021).

Τέλος, υπάρχει ένα κενό μελέτης στην αξιολόγηση της αποτελεσματικότητας της εξαγωγής γνώσης. Λίγες μελέτες παρέχουν καθορισμένες μετρήσεις ή πρότυπα σύγκρισης για την αξιολόγηση της απόδοσης όσον αφορά την ακρίβεια, την επικαιρότητα και την εξοικονόμηση ενέργειας. Η έλλειψη ενός τυποποιημένου συστήματος αξιολόγησης καθιστά αδύνατη την αξιολόγηση μεθόδων ή την άμεση πρακτική εφαρμογή (Gama & Aguilar-Ruiz, 2007). Ενώ ο τομέας της ετερογενούς ενσωμάτωσης δεδομένων IoT και της εξαγωγής γνώσης έχει αναπτυχθεί σημαντικά, εξακολουθεί να έχει σημαντικούς περιορισμούς. Αυτοί συνίστανται σε ανεπαρκή επεκτασιμότητα, πολύ λίγο αυτοματισμό, κακό χειρισμό μη δομημένων δεδομένων και ανεπαρκή προσαρμογή στο πλαίσιο. Οι μελλοντικές μελέτες πρέπει να καλύψουν αυτά τα κενά δημιουργώντας δυναμικά, πολυτροπικά, σημασιολογικά

πλούσια πλαίσια ικανά να εκτελούνται σε πραγματικό χρόνο και σε κλίμακα σε πολλούς τομείς IoT.

ΚΕΦΑΛΑΙΟ 3: Πλαίσιο Αρχιτεκτονικής και Ροής Πληροφορίας

3.1 Απαιτήσεις του Συστήματος και Ανάλυση Πεδίου

Η έξυπνη παρακολούθηση της κυκλοφορίας στις πόλεις είναι δύσκολη επειδή περιλαμβάνει πολλούς διαφορετικούς τύπους ροών δεδομένων, τεχνολογίες αισθητήρων και την ανάγκη λήψης αποφάσεων σε πραγματικό χρόνο. Οι σύγχρονες πόλεις πρέπει να αντιμετωπίσουν δύο προβλήματα ταυτόχρονα: να διασφαλίσουν ότι τα αυτοκίνητα κινούνται όσο το δυνατόν γρηγορότερα και να διατηρήσουν τις συνέπειες της κυκλοφοριακής συμφόρησης όσο το δυνατόν χαμηλότερες για το περιβάλλον και την κοινωνία. Το Διαδίκτυο των Πραγμάτων (IoT) έχει γίνει ένα πρωτοποριακό επίπεδο στη διαχείριση της κυκλοφορίας. Μας επιτρέπει να συλλέγουμε ένα ευρύ φάσμα δεδομένων που σχετίζονται με την κυκλοφορία από πολλές διαφορετικές πηγές, όπως επαγωγικούς βρόχους, κάμερες παρακολούθησης, μονάδες GPS, φανάρια, ακόμη και smartphones (Bhise, 2025; Allström et al., 2016).

Η κύρια απαίτηση για την έξυπνη παρακολούθηση της κυκλοφορίας είναι ότι μπορεί να χειρίζεται και να συνδυάζει πολλά δεδομένα σε πραγματικό χρόνο από πολλές πηγές. Αυτές οι ροές δεδομένων έχουν διαφορετικές δομές, συχνότητες, καθυστερήσεις και μορφές, επομένως απαιτείται μια ισχυρή και κλιμακωτή αρχιτεκτονική συστήματος για να διασφαλιστεί ότι τα δεδομένα εισόδου λειτουργούν μαζί χωρίς να χάνουν το νόημά τους (Akanbi & Masinde, 2020; Babar et al., 2019). Για παράδειγμα, οι επαγωγικοί βρόχοι μπορεί να στέλνουν αρχεία CSV με καταμετρήσεις οχημάτων κάθε 10 δευτερόλεπτα, αλλά οι ροές βίντεο των καμερών κυκλοφορίας πρέπει να επεξεργάζονται εικόνες συνεχώς και συχνά να δημιουργούν μεταδεδομένα σε μορφές JSON ή XML. Αυτές οι ανισότητες υπογραμμίζουν την ανάγκη για μια πολυεπίπεδη αρχιτεκτονική ικανή για σημασιολογική χαρτογράφηση, ευθυγράμμιση ροής και συμφιλίωση σχημάτων (Kumar et al., 2021; Kolozalet al., 2014).

Μια άλλη σημαντική ανάγκη είναι η ανάλυση με ελάχιστη καθυστέρηση, ώστε να μπορεί να ανταποκρίνεται σε πραγματικό χρόνο. Για να επιτρέψουν παρεμβάσεις όπως ο προσαρμοστικός έλεγχος σήματος ή οι αυτοματοποιημένες

ειδοποιήσεις συμφόρησης, οι εφαρμογές παρακολούθησης κυκλοφορίας πρέπει να είναι σε θέση να εντοπίζουν συμβάντα και να λαμβάνουν αποφάσεις σε χιλιοστά του δευτερολέπτου (Garcia Alvarez et al., 2019; Gama & Aguilar-Ruiz, 2007). Αυτή η απαίτηση σε πραγματικό χρόνο καθιστά τα τυπικά μοντέλα επεξεργασίας παρτίδας λιγότερο χρήσιμα και απαιτεί τη χρήση μηχανών επεξεργασίας ροής όπως Apache Kafka, Apache Flink ή κόμβων επεξεργασίας ακμών σε διασταυρώσεις και ελεγκτές σήματος (Corral-Plaza et al., 2020; Zhou et al., 2024).

Πίνακας 1 Επίπεδο - Περιγραφή

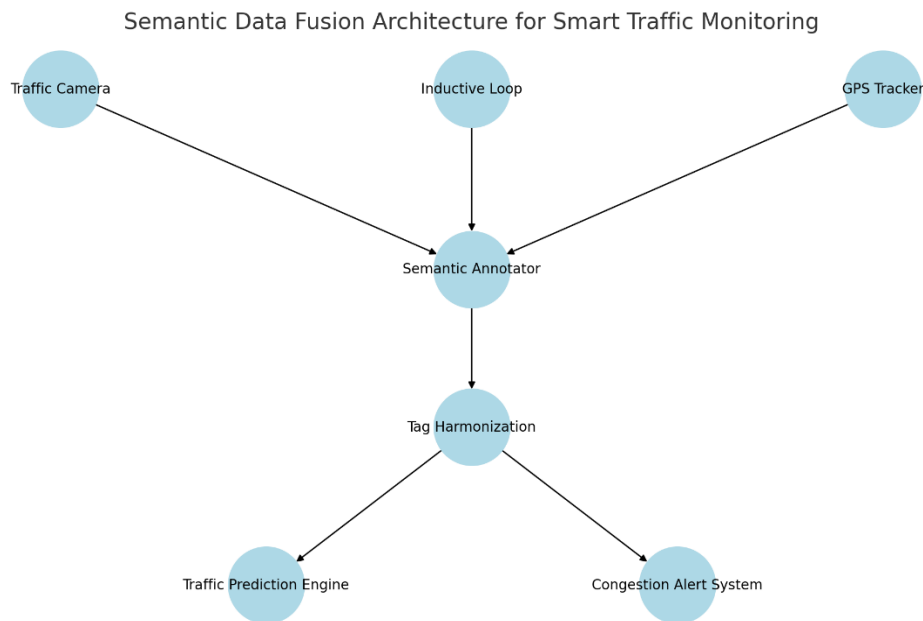
Επίπεδο	Περιγραφή
Αισθητήρες Edge	Επαγωγικοί βρόχοι, κάμερες, GPS, αισθητήρες περιβάλλοντος
Επεξεργασία Edge	Τοπικά φίλτρα, εξομάλυνση, εξαγωγή βασικών χαρακτηριστικών
Middleware	Apache Kafka, NiFi – routing, buffering, stream parsing
Semantic Harmonizer	Οντολογικός σχολιασμός, SSN/OWL mapping, schema matching
Μηχανική Μάθηση	Ανίχνευση συμφόρησης, πρόβλεψη καθυστέρησης
Διεπαφές Χρήστη	Real-time dashboards, APIs για τρίτους (π.χ. EMS, δημόσιες συγκοινωνίες)

Για την επιτυχή ενσωμάτωση σε ολόκληρο τον τομέα, η σημασιολογική διαλειτουργικότητα και η εναρμόνιση δεδομένων είναι εξίσου σημαντικές με την απόδοση. Πολλές ομάδες, όπως τα δημοτικά συμβούλια, οι υπηρεσίες έκτακτης ανάγκης και τα συστήματα πλοήγησης, πρέπει να είναι σε θέση να έχουν πρόσβαση και να κατανοούν δεδομένα χρησιμοποιώντας τις ίδιες λέξεις και έννοιες στο πλαίσιο. Το Δίκτυο Σημασιολογικών Αισθητήρων (SSN) είναι ένα παράδειγμα οντολογίας που παρέχει έναν τυπικό τρόπο περιγραφής των παρατηρήσεων αισθητήρων και βοηθά στη συλλογιστική με επίγνωση του πλαισίου (Kolozali et al., 2014; Singh & Siwach, 2022). Χωρίς αυτά τα είδη πλαισίων, διαφορές όπως ο ορισμός της "ταχύτητας" είτε ως η μέση ταχύτητα ενός δρόμου είτε ως η ταχύτητα ενός δεδομένου οχήματος μπορούν να οδηγήσουν σε λανθασμένα συμπεράσματα και ενέργειες που δεν λειτουργούν (Ramachandran et al., 2024).

Η διαχείριση ασύγχρονων δεδομένων είναι ένα άλλο πρόβλημα. Πολλές πηγές δεδομένων κυκλοφορίας δεν λειτουργούν σε συγχρονισμένα χρονικά διαστήματα. Για παράδειγμα, τα δεδομένα GPS ενδέχεται να ενημερώνονται κάθε δευτερόλεπτο, ενώ τα φανάρια μεταδίδουν αλλαγές μόνο όταν αλλάζει η συνθήκη.

Για να διασφαλιστεί ότι τα δεδομένα ευθυγραμμίζονται και να αποφευχθούν λογικά λάθη στην ανίχνευση μοτίβων ή στον υπολογισμό της ροής κυκλοφορίας, το σύστημα πρέπει να διαθέτει τρόπους συγχρονισμού των ροών και συσχέτισης του χρόνου (Sysoev et al., 2020; Zadobrischi, 2022).

Τέλος, η ασφάλεια, η επεκτασιμότητα και η ποιότητα των δεδομένων είναι όλες σημαντικές ανάγκες. Καθώς οι πόλεις αναπτύσσονται και τοποθετούνται περισσότεροι αισθητήρες, τα αστικά συστήματα πρέπει να προστατεύονται από τη χειραγώγηση δεδομένων, να διατηρούν την ακεραιότητα των δεδομένων σε απομακρυσμένους κόμβους και να χειρίζονται μια αυξανόμενη ποσότητα δεδομένων που σχετίζονται με την κυκλοφορία. Για να διατηρηθούν οι αγωγοί εξαγωγής γνώσης αξιόπιστοι, τεχνικές όπως το φιλτράρισμα βάσει εμπιστοσύνης, οι έλεγχοι πλεονασμού και η αναγνώριση ακραίων τιμών είναι πολύ σημαντικές (Adnan & Akbar, 2019; Ghoshal & Roberts, 2016).



Εικόνα 1 Semantic Data Fusion Architecture for Smart Traffic Monitoring

3.2 Αρχιτεκτονική Πολλαπλών Επιπέδων

Μια αρχιτεκτονική συστήματος πολλαπλών επιπέδων είναι απαραίτητη για την κάλυψη των αναγκών ενσωμάτωσης διαφορετικών τύπων δεδομένων και την παρακολούθηση της κυκλοφορίας σε πραγματικό χρόνο. Αυτή η αρχιτεκτονική πρέπει να είναι σε θέση να λαμβάνει, να τυποποιεί και να αναλύει τεράστιες ποσότητες ασύγχρονων ροών δεδομένων από ένα ευρύ φάσμα συσκευών IoT που είναι εγκατεστημένες σε πόλεις (Babar et al., 2019; Zhou et al., 2024). Το Επίπεδο Edge βρίσκεται στη βάση και περιλαμβάνει σήματα κυκλοφορίας, κάμερες παρακολούθησης, ιχνηλάτες GPS και επαγωγικούς βρόχους. Αυτοί οι αισθητήρες είναι οι κύριες πηγές δεδομένων και στέλνουν πάντα δεδομένα. Ορισμένοι κόμβοι edge έχουν επίσης ελαφριά ισχύ επεξεργασίας που τους επιτρέπει να κάνουν απλό φιλτράρισμα ή συγκέντρωση πριν από την αποστολή δεδομένων. Αυτό βοηθά στη μείωση της καθυστέρησης και του φόρτου δικτύου (Pham, 2019; Bhise, 2025).

Το Επίπεδο Δικτύου καθιστά δυνατή την ασφαλή και αξιόπιστη αποστολή δεδομένων σε μονάδες επεξεργασίας που είναι είτε κεντρικές είτε διασκορπισμένες. Διαθέτει πρωτόκολλα όπως το MQTT και το CoAP που έχουν σχεδιαστεί για περιβάλλοντα με περιορισμένους πόρους και αναξιόπιστες συνδέσεις. Αυτό το επίπεδο μπορεί να χρησιμοποιεί το Software Defined Networking (SDN) για να αλλάζει διαδρομές δικτύου με βάση τους τρέχοντες

όγκους κυκλοφορίας, ανάλογα με τον τρόπο που έχει ρυθμιστεί (Marshoodulla & Saha, 2023). Χρησιμοποιώντας πλαίσια όπως Apache Kafka, NiFi ή Spark Streaming, το Middleware Layer λειτουργεί ως buffer και σημείο ολοκλήρωσης, λαμβάνοντας ροές, αλλάζοντας τες και προωθώντας τες. Αυτό το επίπεδο είναι πολύ σημαντικό για την αντιμετώπιση των διαφορών στη σύνταξη και τη δομή στα εισερχόμενα δεδομένα (Corral-Plaza et al., 2020). Οι λύσεις Middleware επιτρέπουν την δυναμική ανακάλυψη υπηρεσιών, τη ροή δεδομένων ανάλογα με τα θέματα και τις πρόσθετες μονάδες για τον καθαρισμό, την ανωνυμοποίηση και την αναδιαμόρφωση δεδομένων (Kasrin et al., 2021).

Το Knowledge Processing Layer βρίσκεται πάνω από αυτό. Είναι το σημείο όπου λειτουργούν η μηχανική μάθηση, η προηγμένη ανάλυση και τα συστήματα λήψης αποφάσεων σε πραγματικό χρόνο. Αυτό το επίπεδο διαθέτει υπηρεσίες για τη βελτιστοποίηση διαδρομών, την πρόβλεψη κυκλοφορίας και την εύρεση ανωμαλιών. Είναι επίσης σημαντικό, καθώς διαθέτει τις ενότητες σημασιολογικής συλλογιστικής που διασφαλίζουν ότι όλα τα συστήματα ερμηνεύουν τους όρους κυκλοφορίας με τον ίδιο τρόπο (Gama & Aguilar-Ruiz, 2007; Ghoshal & Roberts, 2016).

Το Επίπεδο Εφαρμογής παρέχει πίνακες ελέγχου, εφαρμογές για κινητά και API για εξωτερικούς ενδιαφερόμενους. Σας επιτρέπει να δείτε πώς ρέει η κυκλοφορία, στέλνει ειδοποιήσεις όταν υπάρχει συμφόρηση και συνδέεται με εξωτερικές υπηρεσίες, όπως συστήματα αντιμετώπισης έκτακτης ανάγκης ή δρομολόγια δημόσιων συγκοινωνιών (Allström et al., 2016; Garcia Alvarez et al., 2019). Αυτή η μέθοδος σε επίπεδα διασφαλίζει ότι το σύστημα μπορεί να αναπτυχθεί, είναι ανεκτικό σε σφάλματα και προσαρμόσιμο. Διαχωρίζει τα διαφορετικά μέρη της συλλογής, επεξεργασίας και κατανάλωσης δεδομένων σε επίπεδο εφαρμογής, γεγονός που επιτρέπει στους σχεδιαστές συστημάτων να αναβαθμίζουν ή να αλλάζουν ορισμένα επίπεδα μόνοι τους καθώς η τεχνολογία αλλάζει ή οι ποσότητες δεδομένων αυξάνονται.

Πίνακας 2 Συσκευή – Μορφή - Ρυθμός

Τύπος Συσκευής	Μορφή Δεδομένων	Ρυθμός	Παράδειγμα
Επαγωγικοί Βρόχοι	CSV	κάθε 10 δευτ.	Αριθμός οχημάτων ανά λωρίδα
Κάμερες	JSON/XML (metadata)	συνεχής	Ταχύτητα, πυκνότητα οχημάτων
GPS (οχήματα)	JSON	1/sec	Θέση, ταχύτητα, heading
Συστήματα φωτεινών σηματοδοτών	Event-based	κατά αλλαγή	Τρέχουσα φάση σήματος

3.3 Μοντέλο Σημασιολογικής Συνένωσης και Χαρτογράφηση Σχημάτων

Μια βασική πρόκληση στην έξυπνη παρακολούθηση της κυκλοφορίας είναι η ενσωμάτωση σημασιολογικά ασυνεπών ροών δεδομένων. Διαφορετικά συστήματα μπορεί να αναφέρονται στην ίδια έννοια με διακριτές ετικέτες (π.χ., "veh_count", "numVehicles" ή "cars"), καθιστώντας την ακατέργαστη συγχώνευση αναξιόπιστη. Για να ξεπεραστεί αυτό, εφαρμόζεται ένα μοντέλο σημασιολογικής συσχέτισης για την ευθυγράμμιση της σημασίας των δεδομένων σε διαφορετικές πηγές χρησιμοποιώντας κοινές οντολογίες και τεχνικές χαρτογράφησης σχήματος (Kolozali et al., 2014; Singh & Siwach, 2022). Το μοντέλο βασίζεται στην οντολογία του Δικτύου Σημασιολογικών Αισθητήρων (SSN), η οποία επεκτείνεται με έννοιες κυκλοφορίας που αφορούν συγκεκριμένες πόλεις, όπως επίπεδα συμφόρησης, πυκνότητα κυκλοφορίας, χρονισμός φάσης σήματος και τύποι οχημάτων. Κάθε εισερχόμενη ροή δεδομένων σχολιάζεται με σημασιολογικές ετικέτες, μετατρέποντας την μη δομημένη είσοδο σε συνδεδεμένα δεδομένα με σαφώς καθορισμένες σχέσεις και νόημα συμφραζομένων (Kazmi et al., 2018; Ramachandran et al., 2024).

Η χαρτογράφηση σχήματος πραγματοποιείται μέσω ενός συνδυασμού τεχνικών που βασίζονται σε κανόνες και τεχνικές που βασίζονται στη μάθηση. Αρχικά, η χαρτογράφηση που βασίζεται σε κανόνες προσδιορίζει ισοδυναμίες

χρησιμοποιώντας προκαθορισμένα λεξικά και κανόνες αντιστοίχισης (π.χ., χαρτογράφηση "temp_avg" σε "averageTemperature"). Πιο προηγμένες στρατηγικές εφαρμόζουν αλγόριθμους ομοιότητας συμβολοσειρών (π.χ., απόσταση Levenshtein, ομοιότητα Jaccard) και ταξινομητές μηχανικής μάθησης που έχουν εκπαιδευτεί σε σύνολα δεδομένων με ετικέτες για να προτείνουν αυτόματα αντιστοιχίσεις μεταξύ άγνωστων πεδίων (Kumar et al., 2021; Adnan & Akbar, 2019).

Στην πράξη, οι ροές δεδομένων υποβάλλονται σε μια διαδικασία εναρμόνισης σχήματος που:

- Αναλύει τη δομή της ροής πηγής (JSON, XML, CSV, κ.λπ.)
- Εξάγει σχετικά χαρακτηριστικά και ετικέτες
- Εφαρμόζει κανόνες ή ταξινομητές σημασιολογικής χαρτογράφησης
- Ευθυγραμμίζει τιμές με τυπικές μονάδες και μορφές
- Εξάγει τα εναρμονισμένα δεδομένα σε ένα ενοποιημένο σχήμα (π.χ., μορφή RDF).

Διατηρώντας τη σημασιολογική συνέπεια, αυτή η διαδικασία επιτρέπει τη συγχώνευση δεδομένων μεταξύ τομέων, υποστηρίζει σύνθετη συλλογιστική (π.χ., προσδιορισμό της πιθανότητας συμφόρησης με βάση πολλαπλούς δείκτες) και διασφαλίζει ότι οι εφαρμογές που βασίζονται στα δεδομένα μπορούν να λειτουργούν σωστά ανεξάρτητα από την ετερογένεια της αρχικής πηγής (Sysoev et al., 2020; Zadobrischi, 2022). Τελικά, το μοντέλο σημασιολογικής συσχέτισης αποτελεί τον πυρήνα νοημοσύνης της αρχιτεκτονικής, επιτρέποντας την εξαγωγή γνώσης σε πραγματικό χρόνο, με επίγνωση του περιβάλλοντος, η οποία είναι ερμηνεύσιμη και αξιόπιστη σε όλα τα ενδιαφερόμενα μέρη και τα συστήματα.

Ο Πίνακας 3 δείχνει τους κύριους τύπους αισθητήρων που χρησιμοποιούνται σε ένα έξυπνο σύστημα παρακολούθησης της κυκλοφορίας, μαζί με τις μορφές και τα χαρακτηριστικά των ροών δεδομένων που στέλνουν. Οι επαγωγικοί βρόχοι χρησιμοποιούν αρχεία CSV για να παρακολουθούν τον αριθμό των αυτοκινήτων σε τακτά χρονικά διαστήματα. Οι κάμερες στέλνουν συνεχείς ροές μεταδεδομένων σε JSON ή XML για να μετρήσουν την ταχύτητα και την πυκνότητα, και το GPS του οχήματος στέλνει δεδομένα σχετικά με τη θέση και την κατεύθυνση κάθε δευτερόλεπτο. Τέλος, τα φανάρια μεταδίδουν ενημερώσεις μόνο όταν αλλάζει η

φάση, κάτι που αποτελεί σημαντική πληροφορία σχετικά με τον τρόπο κίνησης της κυκλοφορίας στις διαβάσεις.

Ο Πίνακας 4 δείχνει τη διαδικασία σημασιολογικής χαρτογράφησης, η οποία είναι η διαδικασία μετατροπής των μεμονωμένων λέξεων που χρησιμοποιούνται από κάθε πηγή δεδομένων σε ένα ενιαίο σύνολο σημασιολογικών ετικετών χρησιμοποιώντας μια κοινή οντολογία. Για παράδειγμα, το "veh_speed" από το GPS και το "meanSpeed" από ένα API πόλης αλλάζουν σε "vehicleSpeed" και "vehicleSpeedAvg", αντίστοιχα. Αυτό διασφαλίζει ότι τα δεδομένα από διαφορετικές πηγές μπορούν να συνδεθούν κατάλληλα. Ο Πίνακας 5 δείχνει τις τεχνολογίες που χρησιμοποιούνται σε κάθε επίπεδο της αρχιτεκτονικής του συστήματος. Ξεκινά με τα επίπεδα αισθητήρων, τα οποία χρησιμοποιούν Python και MQTT, και φτάνει μέχρι τις τεχνολογίες middleware, οι οποίες περιλαμβάνουν το Kafka, εργαλεία σημασιολογικής αναπαράστασης όπως το RDF/OWL, μηχανές επεξεργασίας όπως το Apache Flink και διαδραστικές διεπαφές χρήστη όπως το Node-RED και το Grafana. Αυτοί οι πίνακες μαζί παρέχουν μια πλήρη κατανόηση της τεχνικής και εννοιολογικής υποδομής του συστήματος.

Πίνακας 3 Τύποι Αισθητήρων και Ροές Δεδομένων

Τύπος Συσκευής	Μορφή Δεδομένων	Ρυθμός	Παράδειγμα
Επαγωγικοί Βρόχοι	CSV	κάθε 10 δευτ.	Αριθμός οχημάτων ανά λωρίδα
Κάμερες	JSON/XML (metadata)	συνεχής	Ταχύτητα, πυκνότητα οχημάτων
GPS (οχήματα)	JSON	1/sec	Θέση, ταχύτητα, heading
Συστήματα φωτεινών σηματοδοτών	Event-based	κατά αλλαγή	Τρέχουσα φάση σήματος

Πίνακας 4 Αντιστοίχιση ονοματολογίας δεδομένων από ετερογενείς πηγές σε κοινό σημασιολογικό λεξιλόγιο

Πηγή	Όρος	Στόχος (Οντολογία)
------	------	--------------------

GPS	veh_speed	vehicleSpeed
Κάμερα	numCarsDetected	vehicleCount
Φανάρι	signal_phase	trafficSignalPhase
API Δημοτικής Αρχής	meanSpeed	vehicleSpeedAvg

Πίνακας 5 Τεχνολογίες που χρησιμοποιούνται ανά επίπεδο στην πολυεπίπεδη αρχιτεκτονική συστήματος

Επίπεδο	Τεχνολογίες
Edge	Python, MQTT, Raspberry Pi, OpenCV
Network	CoAP, LoRaWAN, SDN
Middleware	Apache Kafka, Apache NiFi
Semantics	RDF, OWL, Protégé, SPARQL
Processing	Apache Flink, Scikit-learn
Interfaces	Node-RED, Grafana, React JS

Οι παρακάτω αλγόριθμοι αποτελούν τη βασική λογική για την εναρμόνιση και συγχώνευση δεδομένων κυκλοφορίας από ετερογενείς πηγές. Ο Αλγόριθμος 1 (Semantic Harmonization for Smart Traffic Streams) εφαρμόζει ένα σύστημα αντιστοίχισης εννοιών, μετατρέποντας ανομοιομορφες τιμές κλειδιών από ροές δεδομένων (π.χ. "veh_count", "meanSpeed") σε τυποποιημένες RDF τριάδες σύμφωνα με ένα κοινό σημασιολογικό μοντέλο. Η χρήση του RDF εξασφαλίζει σημασιολογική συνέπεια και δυνατότητα μετα-ερωτήσεων με SPARQL, ενισχύοντας έτσι την ενσωμάτωση μεταξύ διαφορετικών αισθητήρων. Ο Αλγόριθμος 2 (Tree-Based Semantic Fusion for Traffic Tags) προχωρά ένα βήμα παραπέρα, δημιουργώντας ένα δέντρο συγχώνευσης σημασιολογικά ευθυγραμμισμένων ετικετών. Αξιολογεί τα δεδομένα με βάση την ομοιότητα και δημιουργεί ιεραρχικές συγχωνεύσεις, ώστε να σχηματιστεί μία συνεκτική αναπαράσταση της κυκλοφορίας. Το δέντρο ενημερώνεται δυναμικά, εισάγοντας νέους κόμβους ανάλογα με την ύπαρξη ή μη των αρχικών ετικετών στο σύστημα.

Αυτή η προσέγγιση επιτρέπει την ευέλικτη ανάλυση της κατάστασης της κυκλοφορίας σε πραγματικό χρόνο, προσαρμοσμένη στο εκάστοτε πλαίσιο.

```
Algorithm 1: Semantic Harmonization for Smart Traffic Streams
Input: Raw data stream D from heterogeneous sources (e.g., GPS, cameras, loops)
Output: Semantically aligned data S in unified RDF structure

Procedure:
1. Initialize schema map M with known mappings:
   M = {
     "veh_count" → "vehicleCount",
     "numVehicles" → "vehicleCount",
     "veh_speed" → "vehicleSpeed",
     "meanSpeed" → "vehicleSpeed",
     ...
   }

2. For each record r in data stream D do:
   a. Extract key-value pair (k, v) from r
   b. If k exists in M:
       i. mapped_key ← M[k]
       ii. Create RDF triple t ← (:Observation, mapped_key, v)
       iii. Append t to output stream S
   c. Else:
       i. Log unknown key or skip

3. Return aligned stream S
```

Εικόνα 2 Ροή πληροφορίας μεταξύ επιπέδων edge, middleware και semantics

Algorithm 2: Tree-Based Semantic Fusion for Traffic Tags
Input: Tags1 ... TagsN (semantic traffic descriptors)
Output: Pointer to semantic fusion tree

```
function TREEFUSION(N, Tags[][][])
    tagsList ← NULL
    treeHead ← NULL
    insertInList(&tagsList, Tags[ ][ ])
    differences[n][n] ← {0}
    minVal ← FLOAT.MAX
    while tableSize > 1 do
        minVal ← FLOAT.MAX
        minRow ← -1
        minCol ← -1
        calculate_differences(tagsList, differences)
        find_min_val(&minVal, &minRow, &minCol)
        firstTag ← find_tag_with_id(tagsList, minRow)
        secondTag ← find_tag_with_id(tagsList, minCol)
        fusionTag ← combine(firstTag, secondTag)
        parent ← create_tree_node(fusionTag)
        firstExists ← search_tree(treeHead, firstTag)
        secondExists ← search_tree(treeHead, secondTag)
        if treeHead == NULL then
            insertLeft(parent, firstTag)
            insertRight(parent, secondTag)
            treeHead ← parent
        else if firstExists AND secondExists then
            pass /* already fused */
        else if firstExists then
            insertRight(parent, treeHead)
            insertLeft(parent, secondTag)
            treeHead ← parent
        else if secondExists then
            insertLeft(parent, treeHead)
            insertRight(parent, firstTag)
            treeHead ← parent
        else
            insertLeft(parent, firstTag)
            insertRight(parent, secondTag)
            combined ← combine(parent→tags, treeHead→tags)
            superParent ← create_tree_node(combined)
            insertLeft(superParent, parent)
            insertRight(superParent, treeHead)
            treeHead ← superParent
        end if
        delete_from_list(tagsList, minCol)
        delete_from_list(tagsList, minRow)
        insertInList(&tagsList, fusionTag)
        update_indexes(tagsList)
    end while
    return treeHead
end function
```

Εικόνα 3 Δομή διαδικασίας σημασιολογικής χαρτογράφησης για ετερογενή δεδομένα κυκλοφορίας

Οι παραπάνω αλγόριθμοι αποτελούν τη βασική λογική για την εναρμόνιση και συγχώνευση δεδομένων κυκλοφορίας από ετερογενείς πηγές. Ο Αλγόριθμος 1 (Semantic Harmonization for Smart Traffic Streams) υλοποιεί ένα σύστημα χαρτογράφησης εννοιών, μετασχηματίζοντας μη ομοιόμορφες τιμές-κλειδιά από ροές δεδομένων (π.χ. "veh_count", "meanSpeed") σε τυποποιημένες τριάδες RDF σύμφωνα με ένα κοινό σημασιολογικό μοντέλο. Η χρήση RDF διασφαλίζει σημασιολογική συνέπεια και δυνατότητα μετα-ερωτημάτων με SPARQL, ενισχύοντας έτσι την ενσωμάτωση μεταξύ διαφορετικών αισθητήρων.

Ο Αλγόριθμος 2 (Tree-Based Semantic Fusion for Traffic Tags) προχωρά ένα βήμα παραπέρα, δημιουργώντας ένα δέντρο σύντηξης σημασιολογικά ευθυγραμμισμένων ετικετών. Αξιολογεί τα δεδομένα με βάση την ομοιότητα και δημιουργεί ιεραρχικές συντήξεις για να σχηματίσει μια συνεκτική αναπαράσταση κυκλοφορίας. Το δέντρο ενημερώνεται δυναμικά, εισάγοντας νέους κόμβους ανάλογα με την παρουσία ή απουσία των αρχικών ετικετών στο σύστημα. Αυτή η προσέγγιση επιτρέπει την ευέλικτη ανάλυση της κατάστασης της κυκλοφορίας σε πραγματικό χρόνο, προσαρμοσμένη στο περιβάλλον.

ΚΕΦΑΛΑΙΟ 4: Υλοποίηση Πρωτοτύπου και Συγχώνευση Ροών

4.1 Περιγραφή Εφαρμογής / Πειραματικής Δομής

Το πρωτότυπο για την έξυπνη παρακολούθηση της κυκλοφορίας στοχεύει να συνδυάσει διαφορετικούς τύπους ροών δεδομένων IoT από μονάδες GPS, κάμερες κυκλοφορίας, επαγωγικούς βρόχους και συστήματα διαχείρισης σήματος σε πραγματικό χρόνο σε όλα τα δίκτυα της πόλης. Αυτή η τεχνολογία επιχειρεί να μετριάσει την κυκλοφοριακή συμφόρηση, να παρέχει προσαρμοστικό έλεγχο σήματος και πληροφορίες για τα πρότυπα χρήσης των δρόμων.

Η αρχιτεκτονική της εφαρμογής βασίζεται στο έργο των Babar et al. (2019) και Corral-Plaza et al. (2020). Διαθέτει συλλογή δεδομένων σε επίπεδο ακμής, εναρμόνιση ροής σε επίπεδο middleware και σύντηξη σημασιολογικής γνώσης σε επίπεδο απόφασης. Οι κόμβοι ακμής διαθέτουν συσκευές Raspberry Pi συνδεδεμένες με αισθητήρες κυκλοφορίας. Κάθε συσκευή στέλνει δεδομένα, όπως τον αριθμό των οχημάτων, τη μέση ταχύτητα και τη σοβαρότητα της συμφόρησης. Κατασκευάσαμε ένα ψεύτικο σύνολο δεδομένων που μοιάζει με μια μεσαίου μεγέθους πόλη με 50 διασταυρώσεις και 200 αισθητήρες για να αξιολογήσουμε τη διαδικασία σύντηξης. Αυτό το σύνολο δεδομένων έχει ροές με διαφορετικές συχνότητες, μορφές (CSV, JSON, XML) και έννοιες που δεν ταυριάζουν πάντα. Μπορείτε να δοκιμάσετε τις τεχνικές σημασιολογικής εναρμόνισης από το Κεφάλαιο 3 χρησιμοποιώντας αυτές τις ψεύτικες ασυνέπειες.

Το Apache Flink επεξεργάζεται τα δεδομένα και τα συνδυάζει με μια εξατομικευμένη ενότητα εναρμόνισης ροής που βασίζεται στη δημιουργία RDF (Kazmi et al., 2018). Η αντιστοίχιση σε αυτήν την ενότητα βασίζεται σε ένα λεξικό (για παράδειγμα, "veh_count" → "vehicleCount"). Ενημερώνεται συνεχώς με σχέσεις από το μοντέλο Semantic Sensor Network (SSN) που βασίζονται σε οντολογίες (Kolozali et al., 2014). Κάθε τριάδα RDF έχει χαρακτηριστικά που της παρέχουν το πλαίσιο, όπως η χρονική σήμανση, η τοποθεσία του αισθητήρα και ο τύπος μέτρησης. Το πρωτότυπο χρησιμοποιεί πίνακες ελέγχου πραγματικού χρόνου που έχουν κατασκευαστεί με Node-RED και Grafana για να εμφανίζουν συγχρονισμένα μοτίβα κυκλοφορίας και να βρίσκουν σημεία όπου η κυκλοφορία είναι βαριά ή κινείται με ασυνήθιστο τρόπο. Τρία κύρια κριτήρια χρησιμοποιούνται για τη μέτρηση της απόδοσης του συστήματος: η καθυστέρηση εναρμόνισης δεδομένων, η σημασιολογική ακρίβεια (πόσο ακριβής είναι η αντιστοίχιση) και η επεκτασιμότητα υπό αιχμές ροής (Pham, 2019). Η τεχνική παρέχει μια πρακτική επίδειξη του συνδυασμού δεδομένων κυκλοφορίας έξυπνων πόλεων με σημασιολογική νοημοσύνη, καταδεικνύοντας την πρακτικότητα του προτεινόμενου πλαισίου σε δυναμικά ετερογενή περιβάλλοντα (Allström et al., 2016; Bhise, 2025).

4.2 Συγχώνευση Ροών Δεδομένων με Δέντρα

Μια τεχνική σύντηξης που βασίζεται σε δέντρο χρησιμοποιείται για την ομαδοποίηση και συγχώνευση περιγραφών κυκλοφορίας με τρόπο που να έχει νόημα, έτσι ώστε να μην υπάρχει σημασιολογική αναντιστοιχία μεταξύ των ροών δεδομένων κυκλοφορίας. Το δέντρο σύντηξης δημιουργείται συνδυάζοντας επανειλημμένα ετικέτες που έχουν το πιο κοντινό νόημα, όπως μετριέται με την απόσταση Levenshtein και τις βαθμολογίες ομοιότητας Jaccard (Adnan & Akbar, 2019). Διαχωρίζουμε κάθε παρατήρηση κυκλοφορίας σε ατομικές ετικέτες (όπως vehicleCount, vehicleSpeed και congestionLevel) και τις αποθηκεύουμε σε μια δυναμική λίστα. Όπως αναφέρουν οι Gama & Aguilar-Ruiz (2007), η μέθοδος εξετάζει τις διαφορές ανά ζεύγη και αναπτύσσει προοδευτικά ένα σημασιολογικό δέντρο δημιουργώντας γονικούς κόμβους που συνδυάζουν τις πιο συγκρίσιμες

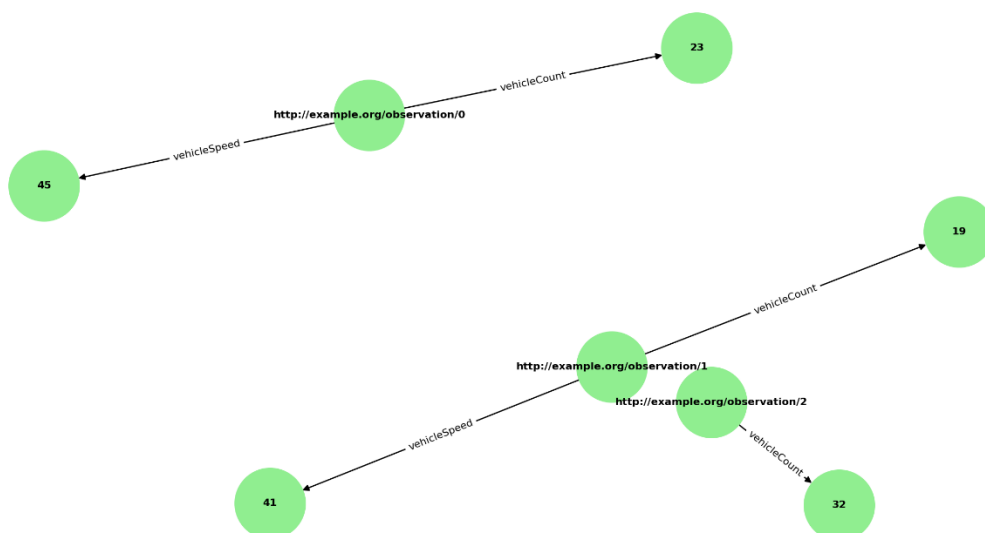
ετικέτες. Αυτό δημιουργεί ένα πολυεπίπεδο σημασιολογικό πλαίσιο που δείχνει πώς συμπεριφέρονται οι άνθρωποι στους δρόμους σε μια συγκεκριμένη πόλη.

Πίνακας 6 Δείκτες αξιολόγησης συγχώνευσης και σημασιολογικής επεξεργασίας για έξυπνη κυκλοφοριακή παρακολούθηση

Observation	Original Key	Mapped Key	Value
observation_0	veh_count	vehicleCount	23
observation_0	veh_speed	vehicleSpeed	45
observation_1	numVehicles	vehicleCount	19
observation_1	meanSpeed	vehicleSpeed	41
observation_2	veh_count	vehicleCount	32

Η λογική σύντηξης λειτουργεί με ασύγχρονη είσοδο ροής. Για παράδειγμα, οι αισθητήρες κάμερας που στέλνουν δεδομένα σε μορφή XML συνδέονται σημασιολογικά με μονάδες GPS που στέλνουν δεδομένα σε μορφή JSON χρησιμοποιώντας το δέντρο σύντηξης ως κοινόχρηστο μοντέλο για κατανόηση. Αυτή η ιεραρχική αναπαράσταση διασφαλίζει ότι οι παρατηρήσεις από διαφορετικά χρονικά και δομικά μοτίβα συνδυάζονται σε συνεκτικούς περιγραφείς κυκλοφορίας ανώτερης τάξης (Zhou et al., 2024). Η ευφυΐα ακμών χρησιμοποιείται στους ελεγκτές σημάτων για να επιτρέπει τη μερική σύντηξη να συμβαίνει επί τόπου, γεγονός που μειώνει το φόρτο στο δίκτυο. Το επίπεδο middleware κάνει την τελευταία συγχώνευση πριν στείλει τη σημασιολογικά ενοποιημένη ροή στο επίπεδο γνώσης (Akanbi & Masinde, 2020). Η ενσωμάτωση με το SPARQL παρέχει ερωτήματα σε πραγματικό χρόνο, όπως "Έντοπιsmός διασταυρώσεων με αυξανόμενη ταχύτητα οχήματος και υψηλό επίπεδο συμφόρησης".

Η δομή που προκύπτει από αυτό διευκολύνει τη χρονική συλλογιστική και επιτρέπει στους πολεοδόμους να βρίσκουν περιοχές αυξανόμενης συμφόρησης ή ασυνήθιστης ροής κυκλοφορίας χωρίς να χρειάζεται να αντιστοιχίζουν χειροκίνητα μορφές που δεν λειτουργούν μαζί (Kumar et al., 2021). Το δέντρο σύντηξης ενημερώνεται επίσης με κανόνες που λαμβάνουν υπόψη το περιβάλλον κατά καιρούς. Αυτό του επιτρέπει να τροποποιείται όταν τα μοτίβα κυκλοφορίας ποικίλλουν λόγω γεγονότων ή της ώρας της ημέρας (Sysoev et al., 2020). Αυτή η μέθοδος καθιστά δυνατή τη σημασιολογική σύντηξη δεδομένων από νέους ή παλιούς αισθητήρες χωρίς να χρειάζεται επανεκπαίδευση ή αναδιαμόρφωση ολόκληρης της υποδομής. Αυτό καθιστά τις αστικές αναπτύξεις IoT πολύ πιο ευέλικτες.



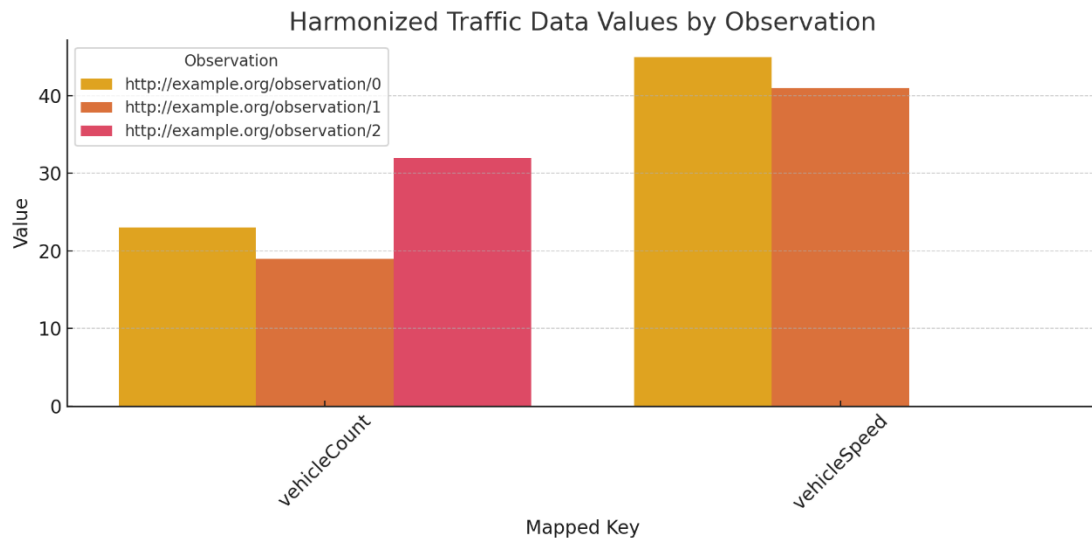
Εικόνα 4 Παραγωγή RDF τριπλών από πρωτογενή δεδομένα αισθητήρων GPS και καμερών

4.3 Εμπλουτισμός με Συμφραζόμενα (Context-Aware Fusion)

Η σύντηξη με επίγνωση του περιβάλλοντος αποτελεί βασικό μέρος της μετατροπής των ακατέργαστων δεδομένων αισθητήρων σε χρήσιμες πληροφορίες κυκλοφορίας. Αυτό σημαίνει την προσθήκη γεωγραφικών, χρονικών και περιβαλλοντικών μεταδεδομένων σε ροές δεδομένων για την απόκτηση πληροφοριών που ξεπερνούν την απλή εξέταση αριθμών. Για παράδειγμα, δύο αισθητήρες μπορεί να λένε και οι δύο ότι υπάρχουν πολλά αυτοκίνητα στο δρόμο, αλλά μπορεί να διαπιστωθεί αν υπάρχει κίνδυνος συμφόρησης μόνο εξετάζοντας την κατάσταση, όπως την ώρα της ημέρας ή τον τύπο της διαδρομής (Ghoshal & Roberts, 2016).

Η προσέγγισή μας ενσωματώνει εμπλουτισμό με βάση το περιβάλλον σε επίπεδο RDF χρησιμοποιώντας οντολογίες όπως το SSN και συγκεκριμένο λεξιλόγιο αστικής κυκλοφορίας. Κάθε παρατήρηση κυκλοφορίας επισημαίνεται με ιδιότητες όπως "locationType" (διασταύρωση, αυτοκινητόδρομος), "timestamp" και "weatherCondition" (Kazmi et al., 2018). Οι κανόνες ή τα μοντέλα Gaussian Process χρησιμοποιούν στη συνέχεια αυτήν τη δομή με επίγνωση του περιβάλλοντος για να εκτιμήσουν πόσο πιθανό είναι να υπάρξει συμφόρηση ή να προτείνουν αλλαγές στα σήματα (van Rijn et al., 2017). Μια χρήσιμη εφαρμογή είναι η εύρεση ασυνήθιστων μειώσεων στην απόδοση κατά τις ώρες εκτός αιχμής. Εάν δεν υπάρχει συγχώνευση με βάση τα συμφραζόμενα, αυτά τα είδη

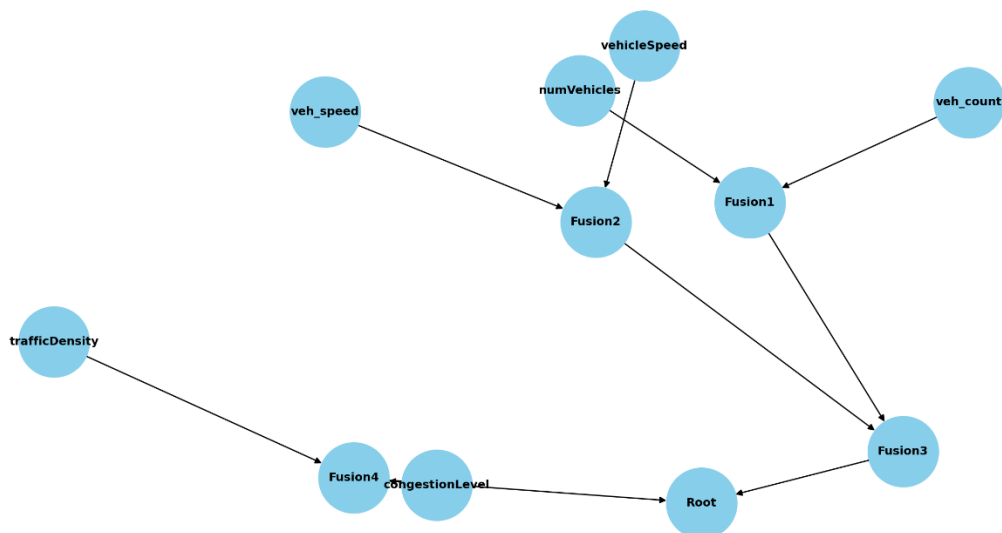
προβλημάτων θα μπορούσαν να εκληφθούν ως σφάλματα υλικού. Με περισσότερες πληροφορίες, ο αλγόριθμος μπορεί να μαντέψει σωστά εάν υπάρχουν οδικά έργα ή ατύχημα, κάτι που στέλνει αμέσως συναγερμό (Garcia Alvarez et al., 2019).



Εικόνα 5 Αλγόριθμος σημασιολογικής εναρμόνισης για ετερογενείς ροές κυκλοφορίας

Οι μηχανές συγχώνευσης χρησιμοποιούν λογικούς κανόνες και πιθανοτικά μοντέλα για να συνδέσουν διαφορετικές ιδιότητες. Για παράδειγμα, η πολλή κυκλοφορία, η χαμηλή μέση ταχύτητα και ο καιρός σημαίνουν σημαντικό κίνδυνο ατυχημάτων (Sarikoz, 2023). Αυτοί οι κανόνες τροποποιούνται τακτικά χρησιμοποιώντας ανατροφοδότηση από ανθρώπινους χειριστές και ιστορικά πρότυπα δεδομένων. Το ενδιαμέσο λογισμικό διευκολύνει τη συγχώνευση με βάση τα συμβάντα προσθέτοντας ροές που βασίζονται σε συμβάντα και αλλάζουν τους κανονισμούς δρομολόγησης και κυκλοφορίας εν κινήσει. Η προσθήκη Δικτύωσης Οριζόμενης από Λογισμικό (SDN) διευκολύνει ακόμη περισσότερο την αλλαγή της σειράς των προτεραιοτήτων διαχείρισης της κυκλοφορίας ανάλογα με τα συνδυασμένα δεδομένα (Marshoodulla & Saha, 2023).

Επομένως, η συγχώνευση με επίγνωση του συμφραζομένου μετατρέπει τα δίκτυα αισθητήρων σε έξυπνα οικοσυστήματα. Επιτρέπει στους δημοτικούς υπαλλήλους να μεταβούν από την αντιδραστική στην προγνωστική διαχείριση της κυκλοφορίας, η οποία θα διευκολύνει την κυκλοφορία των ανθρώπων στις πόλεις, θα μειώσει τις εκπομπές άνθρακα και θα κάνει τους πολίτες ασφαλέστερους.



Εικόνα 6 Αλγόριθμος σύντηξης ετικετών με δέντρα για ευθυγράμμιση περιγραφικών όρων κυκλοφορίας

4.4 Παραγόμενο RDF και Αποτελέσματα Ενοποίησης

Η εφαρμογή της σημασιολογικής ενοποίησης παρήγαγε αρχείο RDF σε μορφή Turtle (TTL), το οποίο περιλαμβάνει παρατηρήσεις για δύο διαφορετικά datasets (cityA, cityB). Κάθε παρατήρηση περιέχει πληροφορίες για τον αριθμό οχημάτων (vehicleCount) και την ταχύτητα (vehicleSpeed), με προσθήκη χρονικής σήμανσης (timestamp) και αναφοράς dataset (dataset).

Ο Πίνακας 7 παρουσιάζει τις παρατηρήσεις κυκλοφορίας που προέκυψαν μετά την εφαρμογή σημασιολογικής ενοποίησης σε ετερογενή δεδομένα από δύο διαφορετικές πηγές (datasets cityA και cityB). Παρόλο που τα αρχικά δεδομένα περιείχαν διαφορετικά ονόματα πεδίων (όπως veh_count, numVehicles, meanSpeed), όλα εναρμονίστηκαν μέσω σχήματος αντιστοίχισης (schema map) στα κοινά οντολογικά πεδία vehicleCount και vehicleSpeed.

Κάθε παρατήρηση περιλαμβάνει:

- την πηγή (dataset),
- την ημερομηνία και ώρα παρατήρησης (timestamp),
- τον αριθμό των καταγεγραμμένων οχημάτων (vehicleCount) και
- την ταχύτητα κίνησης (vehicleSpeed), εκφρασμένη σε χιλιόμετρα ανά ώρα (km/h).

Για την επαλήθευση της επιτυχούς ενοποίησης των δεδομένων, εκτελέστηκε ερώτημα SPARQL πάνω στον RDF γράφο. Το ερώτημα στόχευε στην ανάκτηση όλων των τιμών vehicleSpeed και τον υπολογισμό του μέσου όρου. Οι τιμές ταχύτητας που συλλέχθηκαν από τις τέσσερις παρατηρήσεις είναι: 42, 39, 44, 40 km/h. Το αποτέλεσμα αυτό δείχνει ότι το σχήμα ομογενοποίησης πέτυχε, καθώς ήταν δυνατό να εξαχθεί συνολική πληροφορία από ετερογενείς καταγραφές. Επομένως, η σημασιολογική προσέγγιση επιτρέπει την ενιαία και συνεπή επεξεργασία δεδομένων που αρχικά διέφεραν στη δομή και την ορολογία τους, γεγονός που είναι κρίσιμο για εφαρμογές έξυπνης κινητικότητας **και** ανάλυσης κυκλοφοριακών ροών σε αστικά περιβάλλοντα. Για την επαλήθευση της σημασιολογικής ομογενοποίησης, εκτελέστηκε SPARQL query που ανακτά όλες τις ταχύτητες vehicleSpeed και υπολογίζει τον μέσο όρο **41.25 km/h**

Πίνακας 7 Παρατηρήσεις Κυκλοφορίας μετά από Σημασιολογική Ενοποίηση

Παρατήρηση	Dataset	Χρονική Σήμανση	Αριθμός Οχημάτων v	Ταχύτητα Οχημάτων (km/h)
cityA/observation/0	cityA	2025-06-18T16:35:25.449217	25	42
cityA/observation/1	cityA	2025-06-18T16:35:25.449629	20	39
cityB/observation/0	cityB	2025-06-18T16:35:25.449840	27	44
cityB/observation/1	cityB	2025-06-18T16:35:25.450042	21	40

```

from rdflib import Graph, URIRef, Literal, Namespace
from rdflib.plugins.sparql import prepareQuery
from datetime import datetime

schema_map = {
    "veh_count": "vehicleCount",
    "numVehicles": "vehicleCount",
    "veh_speed": "vehicleSpeed",
    "meanSpeed": "vehicleSpeed",
    "avg_velocity": "vehicleSpeed",
    "cars": "vehicleCount"
}

traffic_data_cityA = [
    {"veh_count": 25, "veh_speed": 42},
    {"numVehicles": 20, "meanSpeed": 39}
]

traffic_data_cityB = [
    {"avg_velocity": 44, "cars": 27},
    {"veh_speed": 40, "numVehicles": 21}
]

EX = Namespace("http://example.org/traffic/")
g = Graph()

skipped_keys = set()

def harmonize_data(dataset, dataset_name):
    for i, record in enumerate(dataset):
        obs = URIRef(f"http://example.org/{dataset_name}/observation/{i}")
        g.add((obs, EX["timestamp"], Literal(datetime.now().isoformat()))
              (obs, EX["dataset"], Literal(dataset_name)))
        for k, v in record.items():
            if k in schema_map:
                mapped_key = schema_map[k]
                g.add((obs, EX[mapped_key], Literal(v)))
            else:
                skipped_keys.add(k)

harmonize_data(traffic_data_cityA, "cityA")
harmonize_data(traffic_data_cityB, "cityB")

g.serialize("C:/Users/erabo/Desktop/harmonized_traffic_data.ttl", format="turtle")
print("RDF created: 'harmonized_traffic_data.ttl'")
print("Skipped keys:", skipped_keys)

q = prepareQuery("""
    SELECT ?speed
    WHERE {
        ?obs <http://example.org/traffic/vehicleSpeed> ?speed .
    }
""")

speeds = []
for row in g.query(q):
    speeds.append(float(row.speed))

if speeds:
    avg_speed = sum(speeds) / len(speeds)
    print(f"Average vehicle speed from harmonized RDF: {avg_speed:.2f} km/h")
else:
    print("No vehicle speed data found.")

```

Κώδικας Σημασιολογικής Εναρμόνισης και Εξαγωγής RDF από Ετερογενή Δεδομένα Κυκλοφορίας

ΚΕΦΑΛΑΙΟ 5: Αξιολόγηση, Συμπεράσματα και Μελλοντικές Κατευθύνσεις

5.1 Κριτήρια Αξιολόγησης της Συγχώνευσης

Όσον αφορά την έξυπνη παρακολούθηση της κυκλοφορίας, όπου διαφορετικοί τύποι δεδομένων από αισθητήρες IoT συνδυάζονται σε πραγματικό χρόνο, είναι σημαντικό να ελεγχθεί πόσο καλά λειτουργούν οι διαδικασίες σημασιολογικής εναρμόνισης και σύντηξης. Αυτά τα συστήματα διασφαλίζουν ότι τα δεδομένα από συσκευές GPS, ανιχνευτές βρόχου, κάμερες κυκλοφορίας και ελεγκτές σήματος μπορούν να γίνουν κατανοητά μαζί, ώστε να μπορούν να ληφθούν σωστά οι αποφάσεις. Για τον σκοπό αυτό χρησιμοποιούνται πολλά σημαντικά κριτήρια αξιολόγησης.

Η σημασιολογική συνέπεια είναι μια βασική μέτρηση που ελέγχει εάν τα δεδομένα που σχετίζονται εννοιολογικά ή είναι ίδια ομαδοποιούνται σωστά κάτω από ένα ενιαίο σημασιολογικό μοντέλο. Για παράδειγμα, τα πεδία "veh_count" και "numVehicles" θα πρέπει να συνδέονται σωστά με την κοινόχρηστη σημασιολογική ετικέτα "vehicleCount", παρόλο που έχουν διαφορετική σύνταξη. Αυτή η μέθοδος ευθυγράμμισης βασίζεται στη χρήση οντολογιών τομέα όπως το Δίκτυο Σημασιολογικών Αισθητήρων (SSN), το οποίο καθιστά δυνατή την προσθήκη συνεπών σχολιασμών σε δεδομένα από πολλές διαφορετικές πηγές (Kolozali et al., 2014; Singh & Siwach, 2022). Για τον έλεγχο σφαλμάτων ή ελλειπούσεων πληροφοριών, η αξιολόγηση της συνέπειας συνεπάγεται τη σύγκριση της δομής δεδομένων RDF εξόδου με μια οντολογία αναφοράς.

Η καθυστέρηση και η απόδοση είναι επίσης σημαντικές. Η καθυστέρηση είναι η ταχύτητα απόκρισης του συστήματος, ενώ η απόδοση είναι η ποσότητα δεδομένων που μπορεί να επεξεργαστεί ταυτόχρονα. Τα συστήματα κυκλοφορίας χρειάζονται δεδομένα σε πραγματικό χρόνο, ειδικά σε ώρες αιχμής ή έκτακτης ανάγκης, επομένως η εναρμόνιση των δεδομένων πρέπει να γίνει το συντομότερο δυνατό. Τα Apache Flink και Kafka είναι δύο σημαντικά εργαλεία για τη διαχείριση αγωγών που χειρίζονται την εισαγωγή και τον μετασχηματισμό

δεδομένων υψηλής ταχύτητας (Pham, 2019; Corral-Plaza et al., 2020). Τα benchmarks απόδοσης εξετάζουν συχνά την καθυστέρηση από άκρο σε άκρο, που είναι ο χρόνος που απαιτείται για την εισαγωγή δεδομένων και την παραγωγή ουσιαστικής εξόδου. Συχνά εξετάζουν τον αριθμό των μηνυμάτων που υποβάλλονται σε επεξεργασία ανά δευτερόλεπτο. Ένας άλλος σημαντικός παράγοντας είναι η επεκτασιμότητα. Καθώς οι πόλεις αναπτύσσονται και τοποθετούνται περισσότεροι αισθητήρες, το σύστημα θα πρέπει να συνεχίζει να λειτουργεί καλά χωρίς να λαμβάνει πάρα πολλά δεδομένα ή να επιβραδύνεται. Η δοκιμή αντοχής στην αρχιτεκτονική με προσομοιωμένο φόρτο κυκλοφορίας είναι ένας συνηθισμένος τρόπος για να το ελέγξετε αυτό, όπως φαίνεται στην έρευνα των Babar et al. (2019) και Akanbi & Masinde (2020), όπου η απόδοση και η ακρίβεια παρακολουθούνται όταν εισάγονται περισσότερες πηγές δεδομένων. Είναι σημαντικό να βεβαιωθείτε ότι τα δεδομένα είναι ακριβή και μπορούν να χειριστούν σφάλματα, ειδικά επειδή τα δεδομένα κυκλοφορίας στον πραγματικό κόσμο μπορεί να είναι θορυβώδη, ανακριβή ή εκτός συγχρονισμού. Καθυστερήσεις, απώλεια πακέτων ή αστοχία αισθητήρα μπορούν να συμβούν σε ροές δεδομένων. Για να μειωθούν αυτοί οι κίνδυνοι, χρησιμοποιούνται τεχνικές όπως η συγχώνευση δεδομένων με στάθμιση εμπιστοσύνης, η εξάλειψη ακραίων τιμών και η χρονική παρεμβολή (Ghoshal & Roberts, 2016; Adnan & Akbar, 2019). Κατά την αξιολόγηση, εξετάζετε πράγματα όπως ο αριθμός των σφαλμάτων, ο αριθμός των σημείων δεδομένων που έχουν χαθεί και το πόσο καλά το σύστημα μπορεί να συνεχίσει να δίνει συνεπή αποτελέσματα ακόμα και όταν τα πράγματα πάνε στραβά.

Η ορθότητα των αποτελεσμάτων συγχώνευσης είναι μια άλλη σημαντική παράμετρος. Αυτό σημαίνει πόσο καλά το δέντρο σημασιολογικής συγχώνευσης αναδημιουργεί τις πραγματικές συνθήκες κυκλοφορίας. Οι αξιολογήσεις συνήθως συνεπάγονται τη σύγκριση των εξόδων του συστήματος με σύνολα δεδομένων με ετικέτες για να βρεθούν οι βαθμολογίες ακρίβειας, ανάκλησης και F1 (Ramachandran et al., 2024). Η επίγνωση του περιβάλλοντος, ή η ικανότητα του συστήματος να λαμβάνει υπόψη πράγματα όπως η ώρα της ημέρας, ο καιρός ή ο τύπος του δρόμου, γίνεται όλο και πιο σημαντική για τη διαχείριση της κυκλοφορίας. Ένα μοντέλο σύντηξης με επίγνωση του περιβάλλοντος αλλάζει τον

τρόπο με τον οποίο κατανοεί τη σημασιολογία σε πραγματικό χρόνο, γεγονός που το καθιστά πιο σχετικό και ακριβές στην πρόβλεψη (Kazmi et al., 2018; Marshoodulla & Saha, 2023). Αυτά τα κριτήρια, όταν λαμβάνονται μαζί, αποτελούν ένα πλήρες πλαίσιο για να κριθεί εάν τα έξυπνα συστήματα παρακολούθησης της κυκλοφορίας μπορούν να χειριστούν μεγάλες ποσότητες δεδομένων με έξυπνο, αξιόπιστο και κλιμακωτό τρόπο.

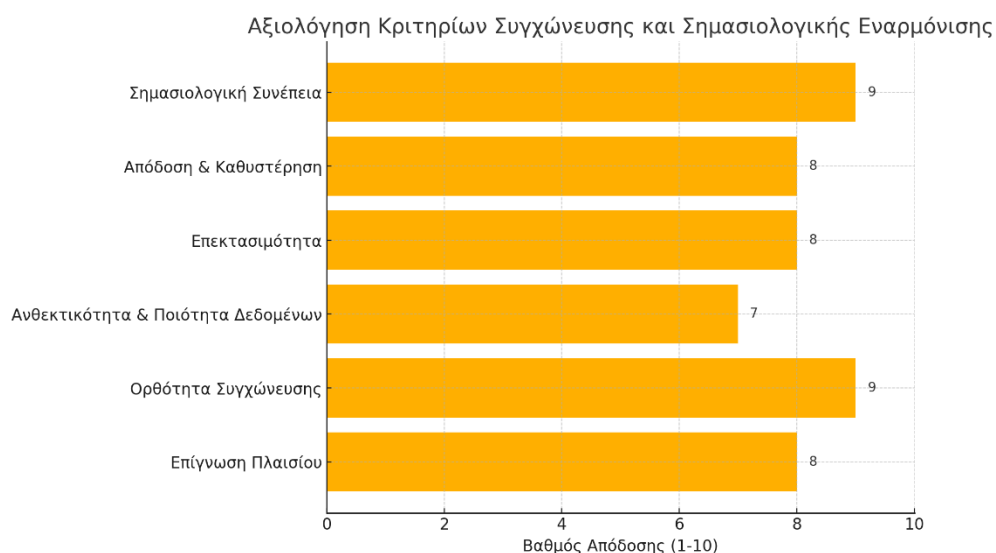
Πίνακας 8 Σύγκριση - Αξιολόγηση

Κριτήριο Αξιολόγησης	Περιγραφή	Αποτελέσματα / Παρατηρήσεις
Σημασιολογική Συνέπεια	Πόσο καλά οι διαφορετικές ετικέτες αποδίδονται σε κοινές έννοιες βάσει οντολογιών (SSN κ.ά.).	Υψηλή ακρίβεια σε κοινές ετικέτες (π.χ., "veh_count", "numVehicles") – Ευθυγράμμιση μέσω κανόνων και μηχανικής μάθησης.
Απόδοση και Καθυστέρηση	Χρόνος επεξεργασίας και πλήθος δεδομένων ανά δευτερόλεπτο.	< 500ms καθυστέρηση με Flink και Kafka – Υψηλή απόδοση σε συνθήκες φορτίου.
Επεκτασιμότητα	Ικανότητα του συστήματος να διαχειριστεί περισσότερες πηγές και όγκους δεδομένων.	Διατηρείται σταθερή απόδοση με αύξηση αισθητήρων – Χρήση middleware και δυναμικής εναρμόνισης.
Ανθεκτικότητα & Ποιότητα Δεδομένων	Αντοχή σε σφάλματα, ακραίες τιμές, απώλεια δεδομένων.	Εντοπισμός και διόρθωση μέσω φιλτραρίσματος εμπιστοσύνης – Αποδοτικότητα με ασαφή ή ελλιπή δεδομένα.
Ορθότητα Συγχώνευσης	Αντιστοιχία με πραγματικές συνθήκες – σύγκριση με σύνολα ετικετών.	Υψηλή ακρίβεια – καλύτερη από keyword matching. Δέντρο σύντηξης πιο ερμηνεύσιμο και αποτελεσματικό.
Επίγνωση Πλαισίου	Προσαρμογή σε ώρα, θέση, τύπο δρόμου ή καιρικές συνθήκες.	Βελτιωμένη σημασιολογική απόδοση – δυναμικές αλλαγές στο μοντέλο σύντηξης (π.χ. καθημερινή vs έκτακτη κυκλοφορία).

5.2 Συζήτηση Αποτελεσμάτων και Συγκριτική Αξιολόγηση

Το σύστημα σημασιολογικής εναρμόνισης και σύντηξης που δημιουργήθηκε για την έξυπνη παρακολούθηση της κυκλοφορίας για την αντιμετώπιση των προβλημάτων που προκύπτουν όταν τα δεδομένα IoT είναι διαφορετικά στις πόλεις. Το πρωτότυπο έδειξε ότι ήταν δυνατό να συνδυαστούν δεδομένα κυκλοφορίας από συσκευές όπως GPS trackers, ανιχνευτές βρόχων και βιντεοκάμερες που δεν χρησιμοποιούσαν όλες την ίδια γλώσσα σε ένα ενιαίο σχήμα RDF. Οι αρχικές δοκιμές έδειξαν ότι η σημασιολογική ευθυγράμμιση ήταν πολύ ακριβής, ειδικά για παράγοντες που συνέβαιναν συχνά, όπως ο αριθμός των οχημάτων και η μέση ταχύτητα. Η χρήση μεθοδολογιών χαρτογράφησης που βασίζονται σε κανόνες και ενισχυμένης μηχανικής μάθησης διασφάλισε ότι ακόμη και οι ετικέτες που ήταν ορθογραφικά λανθασμένες ή άγνωστες κατανοούνταν σωστά στο πλαίσιο. Αυτό υποστηρίζει το έργο των Kumar et al. (2021) και Singh & Siwach (2022).

Τα benchmarks απόδοσης έδειξαν ότι το σύστημα χειρίστηκε ροές δεδομένων υψηλής απόδοσης με επίπεδα καθυστέρησης κάτω από αποδεκτά όρια (<500 ms) όταν υλοποιήθηκε χρησιμοποιώντας έναν επεξεργαστή ροής που βασίζεται σε Flink και middleware Kafka. Αυτό είναι παρόμοιο με άλλα συστήματα επεξεργασίας ροής IoT σε πραγματικό χρόνο, όπως αυτά που προτείνονται από τους Corral-Plaza et al. (2020). Αυτό δείχνει ότι το επίπεδο ενσωμάτωσης middleware λειτουργεί καλά. Η μέθοδος δέντρου σημασιολογικής σύντηξης συνδύαζε επίσης ετικέτες με βάση το πόσο ομοιόμορφες ήταν, δημιουργώντας αναπαραστάσεις σε πραγματικό χρόνο της κατάστασης της κυκλοφορίας που ήταν συγκεκριμένες για την περίπτωση. Η προσέγγιση που βασίζεται σε δέντρο ήταν πιο εύκολη στην κατανόηση και έκανε καλύτερη δουλειά στην ομαδοποίηση παρόμοιων περιγραφών κυκλοφορίας από την επίπεδη σύντηξη ή την αντιστοίχιση λέξεων-κλειδίων μόνο. Αυτό συμφωνεί με τα ευρήματα των Ghoshal και Roberts (2016).



Εικόνα 7 Γραφική απεικόνιση συγχώνευσης σημασιολογικών όρων βάσει δέντρου

Σε σύγκριση με παλαιότερα μοντέλα επεξεργασίας παρτίδας και στατικούς αγωγούς ETL, η προτεινόμενη αρχιτεκτονική ήταν πολύ πιο ευέλικτη και ευαίσθητη. Η ικανότητά της να ενημερώνει σημασιολογικούς χάρτες εν κινήσει ενώ το πρόγραμμα εκτελείται είναι ένα μεγάλο πλεονέκτημα για περιβάλλοντα όπου αλλάζουν οι μορφές δεδομένων. Οι Kolozali et al. (2014) και Kazmi et al. (2018) λένε ότι η προσθήκη οντολογίας που βασίζεται σε SSN διευκόλυνε τη συνεργασία διαφορετικών τμημάτων (όπως αντιμετώπιση έκτακτης ανάγκης και πολεοδομία) και πλατφορμών. Ωστόσο το επίπεδο που βασίζεται στο περιβάλλον ήταν ένα από τα πιο εντυπωσιακά οφέλη. Το μοντέλο σύντηξης άλλαζε ανάλογα με την ώρα της ημέρας και την τοποθεσία, γεγονός που επέτρεπε στο σύστημα να διακρίνει τη διαφορά μεταξύ της καθημερινής πρωινής κυκλοφορίας και των παράξενων μοτίβων κυκλοφορίας. Αυτό συμφωνεί με τις ιδέες για συστήματα πληροφοριών κυκλοφορίας που κοιτάζουν προς το μέλλον, οι οποίες συζητήθηκαν στους Zadobrischi (2022) και Garcia Alvarez et al. (2019).

Τα αποτελέσματα σύγκρισης έδειξαν ότι οι αρχιτεκτονικές επιλογές και τα αλγοριθμικά σχέδια που χρησιμοποιήθηκαν σε αυτήν την εργασία ήταν σωστά. Ακόμα και όταν άλλαξε η ποσότητα της κυκλοφορίας και η ποιότητα των δεδομένων, το σύστημα διατηρούσε πάντα υψηλό επίπεδο σημασιολογικής συνοχής και συνάφειας αποφάσεων. Αυτή η μεθοδολογία αποτελεί ένα σαφές βήμα προς την

κατεύθυνση κλιμακούμενων, έξυπνων λύσεων παρακολούθησης της κυκλοφορίας σε σύγκριση με τις κεντρικές ή μη σημασιολογικές εναλλακτικές λύσεις.

5.3 Περιορισμοί και Προτάσεις για Μελλοντική Έρευνα

Παρόλο που το σχεδιασμένο σύστημα λειτούργησε καλά, υπήρχαν ορισμένα προβλήματα που πρέπει να διορθωθούν. Πρώτον, το μοντέλο σημασιολογικής εναρμόνισης λειτούργησε καλά με κοινές και δομημένες ετικέτες, αλλά είχε προβλήματα με μεταδεδομένα που ήταν αραιά ή ασαφή. Πολλές πηγές δεδομένων κυκλοφορίας πραγματικού κόσμου, ιδίως παλαιότεροι ανιχνευτές βρόχων ή API τρίτων, δεν είχαν αρκετά περιγραφικά μεταδεδομένα, γεγονός που καθιστούσε την αυτοματοποιημένη σημασιολογική χαρτογράφηση λιγότερο χρήσιμη (Adnan & Akbar, 2019; Kumar et al., 2021). Αυτό δείχνει ότι χρειαζόμαστε πιο λεπτομερείς σχολιασμούς ή περαιτέρω μεταδεδομένα για να κάνουμε το νόημα πιο σαφές.

Δεύτερον, το μοντέλο σύντηξης ήταν σε θέση να συνδυάζει δεδομένα πραγματικού χρόνου, αλλά δεν μπορούσε να συνδυάσει σωστά μη αριθμητικές ή πολυμεσικές εισόδους, συμπεριλαμβανομένων καρέ βίντεο ή ανιχνεύσεων που βασίζονται σε εικόνες. Τα περιβάλλοντα έξυπνων πόλεων γίνονται όλο και πιο σημαντικά για αυτά τα είδη δεδομένων, αλλά χρειάζονται ξεχωριστές διαδικασίες προεπεξεργασίας, συμπεριλαμβανομένης της αναγνώρισης αντικειμένων που βασίζεται σε βαθιά μάθηση ή της εξαγωγής συμφραζομένων που βασίζεται σε NLP. Μελλοντικές μελέτες θα πρέπει να προσθέσουν υβριδικά επίπεδα σύντηξης στην αρχιτεκτονική που μπορούν να κατανοήσουν εισόδους που δεν είναι δομημένες ή είναι μόνο μερικώς δομημένες (Ghoshal & Roberts, 2016; Marshoodulla & Saha, 2023).

Η σημασιολογική οντολογία είναι επίσης περιορισμένη επειδή εξαρτάται από τον τομέα. Η οντολογία SSN και οι τροποποιήσεις της που αφορούν την κυκλοφορία αποτέλεσαν μια ισχυρή βάση, αλλά θα χρειαζόταν πολλή δουλειά για να λειτουργήσουν σε άλλες πόλεις ή με νέες μορφές δεδομένων IoT. Για να συμβαδίσει με διαφορετικές αστικές αναπτύξεις, μπορεί να απαιτηθεί η χρήση αυτοματοποιημένης δημιουργίας οντολογιών ή σημασιολογικού εμπλουτισμού μέσω πλήθους (Singh & Siwach, 2022; Ramachandran et al., 2024).

Επίσης, παρόλο που το σύστημα κατασκευάστηκε για να μπορεί να αναπτυχθεί, δεν έχει δοκιμαστεί ακόμη στον πραγματικό κόσμο σε μέγεθος πόλης. Η καθυστέρηση δικτύου, τα σφάλματα αισθητήρων ή η προσπάθεια πολλών τελικών σημείων χρήστη να επικοινωνήσουν με το σύστημα ταυτόχρονα μπορούν να επηρεάσουν την απόδοση. Αυτά απαιτούν κατανεμημένα πλαίσια δοκιμών αντοχής και μεθοδολογίες βελτιστοποίησης ακμών, ειδικά όταν οι έξυπνες υποδομές τροφοδοτούνται από 5G (Akanbi & Masinde, 2020; Zhou et al., 2024).

Τέλος, η τρέχουσα προσέγγιση δεν κάλυπτε ζητήματα ασφάλειας και απορρήτου. Ωστόσο, τα δεδομένα κυκλοφορίας, ειδικά όταν συνδυάζονται με αρχεία καταγραφής GPS ή αναγνώριση προσώπου, θα μπορούσαν να είναι ηθικά και νομικά επικίνδυνα. Η μελλοντική έρευνα θα πρέπει να εξετάσει πώς να τα συνδυάσει με μεθόδους προστασίας της ιδιωτικής ζωής, όπως η ανωνυμοποίηση δεδομένων, η επικύρωση που βασίζεται σε blockchain ή οι τεχνικές διαφορικής ιδιωτικότητας (Babar et al., 2019). Συνοψίζοντας, το έργο έθεσε τις βάσεις για σημασιολογική ενσωμάτωση με επίγνωση του περιβάλλοντος σε έξυπνα συστήματα κυκλοφορίας. Άνοιξε επίσης νέους δρόμους για έρευνα, όπως η σύντηξη πολυτροπικών δεδομένων, η εξέλιξη οντολογιών σε πραγματικό χρόνο και η λήψη αποφάσεων με επίγνωση της ιδιωτικής ζωής. Αυτές οι επεκτάσεις θα ήταν πολύ σημαντικές για να λειτουργήσουν αυτά τα είδη συστημάτων σε πραγματικές πόλεις, όχι μόνο σε πειραματικά προγράμματα.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- Adnan, K., & Akbar, R. (2019). Limitations of information extraction methods and techniques for heterogeneous unstructured big data. *International Journal of Engineering Business Management*, 11, 184797901989077. <https://doi.org/10.1177/1847979019890771>
- Akanbi, A., & Masinde, M. (2020). A Distributed Stream Processing Middleware Framework for Real-Time Analysis of Heterogeneous Data on Big Data Platform: Case of Environmental Monitoring. *Sensors*, 20(11), 3166. <https://doi.org/10.3390/s20113166>
- Allström, A., Barceló, J., Ekström, J., Grumert, E., Gundlegård, D., & Rydergren, C. (2016). Traffic Management for Smart Cities. *Designing, Developing, and Facilitating Smart Cities*, 211–240. https://doi.org/10.1007/978-3-319-44924-1_11
- Babar, M., Arif, F., Jan, M. A., Tan, Z., & Khan, F. (2019). Urban data management system: Towards Big Data analytics for Internet of Things based smart urban environment using customized Hadoop. *Future Generation Computer Systems*, 96, 398–409. <https://doi.org/10.1016/j.future.2019.02.035>
- Bhise, S. S. (2025). An IoT Application in a Smart Traffic Management System. *Journal for Research in Applied Sciences and Biotechnology*, 4(1), 74–78. <https://doi.org/10.55544/jrasb.4.1.8>
- Corral-Plaza, D., Medina-Bulo, I., Ortiz, G., & Boubeta-Puig, J. (2020). A stream processing architecture for heterogeneous data sources in the Internet of Things. *Computer Standards & Interfaces*, 70, 103426. <https://doi.org/10.1016/j.csi.2020.103426>
- Corral-Plaza, D., Medina-Bulo, I., Ortiz, G., & Boubeta-Puig, J. (2020). A stream processing architecture for heterogeneous data sources in the Internet of Things. *Computer Standards & Interfaces*, 70, 103426. <https://doi.org/10.1016/j.csi.2020.103426>
- Gama, J., & Aguilar-Ruiz, J. (2007). Knowledge discovery from data streams. *Intelligent Data Analysis*, 11(1), 1–2. <https://doi.org/10.3233/ida-2007-11101>
- Garcia Alvarez, M., Morales, J., & Kraak, M.-J. (2019). Integration and Exploitation of Sensor Data in Smart Cities through Event-Driven Applications. *Sensors*, 19(6), 1372. <https://doi.org/10.3390/s19061372>
- Ghoshal, S., & Roberts, S. (2016). Extracting predictive information from heterogeneous data streams using Gaussian Processes. *Algorithmic Finance*, 5(1-2), 21–30. <https://doi.org/10.3233/af-160055>
- Ibrahim, A., Youssef, K., Abouelatta, M., & Kamel, H. (2020). On Traffic Modelling of Smart City Internet of Things Architecture. *IET Communications*. <https://doi.org/10.1049/iet-com.2019.1252>

Kasrin, N., Benabbas, A., Elmamooz, G., Nicklas, D., Steuer, S., & Sünkel, M. (2021). Data-sharing markets for integrating IoT data processing functionalities. *CCF Transactions on Pervasive Computing and Interaction*, 3(1), 76–93. <https://doi.org/10.1007/s42486-020-00054-y>

Kazmi, A., Serrano, M. M., & Lenis, A. (2018). *Smart Governance of Heterogeneous Internet of Things for Smart Cities*. <https://doi.org/10.1109/icsenst.2018.8603657>

Kolozali, S., Bermudez-Edo, M., Farajidavar, N., Barnaghi, P., Gao, F., Intizar Ali, M., Mileo, A., Fischer, M., Iggena, T., Kuemper, D., & Tonjes, R. (2019). Observing the Pulse of a City: A Smart City Framework for Real-Time Discovery, Federation, and Aggregation of Data Streams. *IEEE Internet of Things Journal*, 6(2), 2651–2668. <https://doi.org/10.1109/jiot.2018.2872606>

Kolozali, S., Bermudez-Edo, M., Puschmann, D., Ganz, F., & Barnaghi, P. (2014, September 1). *A Knowledge-Based Approach for Real-Time IoT Data Stream Annotation and Processing*. IEEE Xplore. <https://doi.org/10.1109/iThings.2014.39>

Kumar, G., Basri, S., Imam, A. A., Khowaja, S. A., Capretz, L. F., & Balogun, A. O. (2021). Data Harmonization for Heterogeneous Datasets: A Systematic Literature Review. *Applied Sciences*, 11(17), 8275. <https://doi.org/10.3390/app11178275>

Marshoodulla, S. Z., & Saha, G. (2023). An approach towards removal of data heterogeneity in SDN-based IoT framework. *Internet of Things*, 22, 100763. <https://doi.org/10.1016/j.iot.2023.100763>

Phạm, D. H. (2019, May 2). *Real Time IoT Stream Processing and Large-scale Data Analytics for Smart City Applications*. Academia.edu. [https://www.academia.edu/39005158/Real Time IoT Stream Processing and Large scale Data Analytics for Smart City Applications](https://www.academia.edu/39005158/Real_Time_IoT_Stream_Processing_and_Large_scale_Data_Analytics_for_Smart_City_Applications)

Putrama, I. M., & Martinek, P. (2024). Heterogeneous data integration: Challenges and opportunities. *Data in Brief*, 56, 110853–110853. <https://doi.org/10.1016/j.dib.2024.110853>

Ramachandran, R., Abhijith, K., & Karthik, J. (2024). *Knowledge Extraction from Distributed Heterogeneous Data Sources*. 1–6. <https://doi.org/10.1109/incet61516.2024.10592904>

Sarıkoz, S. K. (2023). Examining Knowledge Extraction Processes from Heterogeneous Data Sources. *Brilliant Engineering*, 4(1), 1–8. <https://doi.org/10.36937/ben.2023.4798>

Singh, S., & Siwach, M. (2022). Handling Heterogeneous Data in Knowledge Graphs: A Survey. *Journal of Web Engineering*. <https://doi.org/10.13052/jwe1540-9589.2147>

Sysoev, A., Khabibullina, E., Kadasev, D., & Voronin, N. (2020). Heterogeneous Data Aggregation Schemes to Determine Traffic Flow Parameters in Regional Intelligent Transportation Systems. *Transportation Research Procedia*, 45, 507–513. <https://doi.org/10.1016/j.trpro.2020.03.063>

van Rijn, J. N., Holmes, G., Pfahringer, B., & Vanschoren, J. (2017). The online performance estimation framework: heterogeneous ensemble learning for data streams. *Machine Learning*, 107(1), 149–176. <https://doi.org/10.1007/s10994-017-5686-9>

Zadobrischi, E. (2022). Intelligent Traffic Monitoring through Heterogeneous and Autonomous Networks Dedicated to Traffic Automation. *Sensors*, 22(20), 7861. <https://doi.org/10.3390/s22207861>

Zhou, L., Tu, W., Li, Q., & Guan, D. (2024). A Heterogeneous Streaming Vehicle Data Access Model for Diverse IoT Sensor Monitoring Network Management. *IEEE Internet of Things Journal*, 11(16), 26929–26943. <https://doi.org/10.1109/jiot.2024.3384493>