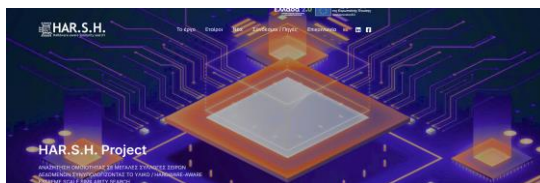


HAR.S.H. – HARdware aware extreme-scale Similarity search Αναζήτηση Ομοιότητας σε Μεγάλες Συλλογές Σειρών Δεδομένων Συνυπολογίζοντας το Υλικό

Δελτίο Τύπου



Οι *σειρές δεδομένων* έχουν αμέτρητες εφαρμογές, σχεδόν σε κάθε τομέα της κοινωνίας, της επιστήμης και της ζωής καθώς εμφανίζονται ως ακολουθίες ήχου, αναπαραστάσεις εικόνων ή βίντεο, οικονομικά και τηλεπικοινωνιακά δεδομένα, δεδομένα παρακολούθησης του περιβάλλοντος και επιστημονικά δεδομένα. Η εκθετική αύξηση των δεδομένων που παράγονται από τέτοιου είδους εφαρμογές σε όλους τους παραπάνω τομείς οδηγεί στην ανάγκη για αποδοτική ανάλυση μεγάλων συλλογών σειρών δεδομένων πολλών διαστάσεων. Ταυτόχρονα, η εξάπλωση της μηχανικής μάθησης έχει επιτρέψει την εκτέλεση απαιτητικών εργασιών πάνω σε δεδομένα συγκεκριμένων μορφών. Ωστόσο, οι μοντέρνες εφαρμογές οδηγούν στην αναγκαιότητα επεξεργασίας δεδομένων που δεν είναι στην ίδια μορφή (π.χ. κείμενα, εικόνες, βίντεο). Η επεξεργασία τέτοιων συλλογών *πολυτροπικών δεδομένων* ενέχει πολλές ερευνητικές προκλήσεις αλλά είναι επιτακτική καθώς μια ενοποιημένη αναπαράσταση των διαφορετικών τύπων δεδομένων θα επιτρέψει την επεξεργασία πληροφορίας σε σημαντικά ευρύτερο φάσμα. Το HAR.S.H. φιλοδοξεί να παρέχει καινοτόμες λύσεις και να έχει σημαντικό αντίκτυπο στις ακόλουθες κατευθύνσεις:

- **Κωδικοποίηση Πολυτροπικών Δεδομένων.** Αντιμετώπιση της ανάγκης κωδικοποίησης αντικειμένων υψηλών διαστάσεων, όπως εικόνες και βίντεο, επιτρέποντας στις σειρές δεδομένων να αποτυπώνουν πολυτροπικά δεδομένα.
- **Εκμετάλλευση της πλήρους υπολογιστικής ισχύος των σύγχρονων υπολογιστικών πλατφορμών.** Αξιοποίηση όλων των υπολογιστικών μονάδων των σύγχρονων υπολογιστικών πλατφορμών, συμπεριλαμβανομένων των πολλαπλών κόμβων, των πυρήνων κάθε κόμβου και των επιταχυντών.
- **Σχεδίαση και υλοποίηση αλγορίθμων και δομών δεδομένων με επίγνωση του υλικού.** Μελέτη της επίδρασης που μπορεί να έχει η αξιοποίηση μοντέρνων και αναδυόμενων τεχνολογιών υλικού στην επεξεργασία μεγάλου όγκου σειρών δεδομένων με δυναμικό ή εύρωστο τρόπο.

Το HARSH θα αποδείξει την αξία της τεχνολογίας που θα παράγει μέσω των εξής πιλοτικών εφαρμογών:

1. **Αναζήτηση παρόμοιων εγγράφων** σε μεγάλες βάσεις δεδομένων.
2. **Ανάλυση φωτογραφιών για εξατομίκευση ταξιδιωτικών προτάσεων.**
3. **Ανάλυση γνώμης πολιτών** μέσα από δεδομένα κοινωνικής δικτύωσης.

Το έργο HAR.S.H. (ΥΠ3ΤΑ-0560901) ξεκίνησε τον Απρίλιο του 2025 και θα είναι ενεργό μέχρι και τον Μάιο του 2026. Εντάσσεται στη δράση «SUB1.1. Συμπράξεις Ερευνητικής Αριστείας – ΣΕΑ».

Το έργο/δράση υλοποιείται στο πλαίσιο του Εθνικού Σχεδίου Ανάκαμψης και Ανθεκτικότητας «Ελλάδα 2.0» με τη χρηματοδότηση της Ευρωπαϊκής Ένωσης – NextGenerationEU.

ΠΕΡΙΣΣΟΤΕΡΕΣ ΠΛΗΡΟΦΟΡΙΕΣ

- harsh-project.eu
- www.linkedin.com/company/harsh-project
- www.facebook.com/harshproject

ΣΥΝΕΡΓΑΖΟΜΕΝΑ ΙΝΣΤΙΤΟΥΤΑ



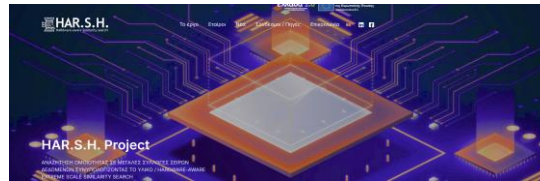
ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ
UNIVERSITY OF CRETE

ΣΤΟΙΧΕΙΑ ΕΠΙΚΟΙΝΩΝΙΑΣ



Καθηγήτρια Παναγιώτα Φατούρου
Τμήμα Επιστήμης Υπολογιστών
Πανεπιστήμιο Κρήτης
Τηλ.: +30 6973991277,
Email: faturu@csd.uoc.gr
URL: www.csd.uoc.gr/~faturu

HAR.S.H.– HARdware aware extreme-scale Similarity search Press Release, May 2025



Ordered sequences of data points, known as *data series*, are one of the most common types of data. They are present in virtually every scientific and social domain. They appear as audio sequences, shape and image data, financial, telecommunications, environmental monitoring and scientific data, and they have many diverse applications, e.g., in health care, earth sciences, astronomy, biology, economics, etc. The exponential growth of data generated by such applications in all of the above areas leads to the need for efficient analysis of large collections of high-dimensional data series. At the same time, the proliferation of machine learning has enabled performing demanding tasks, albeit on data of specific formats. However, modern applications require processing multimodal data, i.e., data that is not in the same format (e.g., text, images, video). The efficient processing of such data collections involves many research challenges but is imperative as a unified representation of the different data types will allow processing of information over a significantly wider range.

HARSH will provide innovative solutions in the following directions:

- **Encoding Multimodal data.** HARSH will address the need to encode high-dimensional objects such as audio, images and videos, enabling data series to capture multi-dimensional data.
- **Utilizing the full computational power of modern computing platforms in an agnostic way.** HARSH will exploit all computing elements of modern computing platforms, including not only multiple nodes, but also the multi-core capabilities of each node, as well as the full capacity of the attached accelerators.
- **Hardware-Aware Algorithms and Data Structures for Data Series Processing.** We will focus on emerging memory technologies and study how the utilization of such technology can influence (or add to) the foundations of data series processing.

HARSH will demonstrate its value proposition in the following ways:

- **Use case 1 – Similar Document/File Finding Application.**
- **Use Case 2 – Photo Analysis for Travel Profile Enhancement Application.**
- **Use case 3 – Public Opinion Analysis Application.**

The HAR.S.H. project (ΥΠ3ΤΑ-0560901) was launched in April 2025 and will be active until May 2026. It is part of the initiative “SUB1.1 – Research Excellence Partnerships (REP)”.

This project is implemented under the National Recovery and Resilience Plan “Greece 2.0”, with funding from the European Union – NextGenerationEU

MORE INFORMATION

- harsh-project.eu
- www.linkedin.com/company/108218606
- www.facebook.com/harshproject

COLLABORATING INSTITUTES



ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ
UNIVERSITY OF CRETE



CONTACT INFORMATION



Professor Panagiota Fatourou
Department of Computer Science
University of Crete
Τηλ.: +30 6973991277,
Email: faturu@csd.uoc.gr
URL: www.csd.uoc.gr/~faturu

