

9.2 Επικοινωνία και συνεργασία μέσω ψηφιακών περιβαλλόντων

Μεγάλα Δεδομένα - Big Data

Συλλογή και Επεξεργασία Δεδομένων στο Διαδίκτυο (Μαζικά Δεδομένα - Big Data)

- Τα **Big Data** δεν είναι απλά τα «μεγάλα δεδομένα», δεν είναι μόνο ένα σύνολο δεδομένων που συλλέγονται στο διαδίκτυο, αλλά και το **σύνολο των εργαλείων, τεχνικών και τεχνολογιών**, το οποίο διαχειρίζεται και επεξεργάζεται αυτό τον **τεράστιο όγκο δεδομένων**. Τα Big Data προκύπτουν από πολλαπλές πηγές, όπως άτομα, επιχειρήσεις/οργανισμούς, δικτυακούς τόπους, συσκευές, αισθητήρες, κ.α., και συμβάλλουν στην παραγωγή γνώσης και αξίας.
- Η καινοτομία που προσφέρουν οι τεχνολογίες των **Big Data** δεν αναφέρεται στην ύπαρξη περισσότερων δεδομένων, αλλά στο ότι υπάρχει πλέον η τεχνογνωσία και οι μέθοδοι να τα διαχειριστούμε και να ανακαλύψουμε τη γνώση που κρύβεται σε αυτά. Η ανάλυση των δεδομένων μπορεί να βοηθήσει στην εξαγωγή συμπερασμάτων και στην ανακάλυψη μοτίβων και τάσεων που δεν είναι εμφανείς στο ανθρώπινο μάτι.
- Ο όρος «big data» αναφέρεται σε δεδομένα που χαρακτηρίζονται από μεγάλο όγκο, ποικιλία, ταχύτητα και ακρίβεια. Αυτά τα χαρακτηριστικά αναφέρονται συχνά ως τα «τέσσερα V» των big data.
- **Όγκος δεδομένων (Volume)**: Ο όγκος των big data είναι εξαιρετικά μεγάλος, σε τάξεις μεγέθους τεραμπάιτ (TB - terabytes), πεταμπάιτ (PB - petabytes) ή ακόμη και εξαμπάιτ (EB - hexabytes). Για παράδειγμα, 1 TB είναι ίσο με 1 τρισεκατομμύριο bytes ή ίσο με 1024 GB. Αντίστοιχα 1 PB είναι ίσο με 1024 TB κ.λπ.
- **Ταχύτητα δεδομένων (Velocity)**: Η ταχύτητα με την οποία δημιουργούνται, συλλέγονται και επεξεργάζονται τα δεδομένα. Σε ορισμένες περιπτώσεις, τα δεδομένα πρέπει να αναλύονται σε πραγματικό χρόνο ή κοντά σε πραγματικό χρόνο για να παραχθεί νόημα.
- **Ποικιλία δεδομένων (Variety)**: Τα big data έχουν πολλές και διαφορετικές μορφές. Συνήθως αναφερόμαστε σε τρεις μορφές: τα δομημένα δεδομένα, τα ημιδομημένα δεδομένα και τα μη δομημένα δεδομένα. Τα δομημένα δεδομένα είναι δεδομένα που έχουν σαφώς καθορισμένη δομή, όπως τα δεδομένα που αποθηκεύονται σε παραδοσιακές βάσεις δεδομένων. Τα ημιδομημένα δεδομένα είναι δεδομένα που έχουν κάποια δομή, αλλά όχι τόσο σαφώς καθορισμένη όσο τα δομημένα δεδομένα, όπως για παράδειγμα τα δεδομένα σε ένα φύλλο εργασίας, ή τα δεδομένα που είναι αποθηκευμένα σε μορφότυπους XML και JSON. Τα μη δομημένα δεδομένα είναι δεδομένα που δεν έχουν καμία δομή, όπως το κείμενο σε ένα αρχείο κειμένου, οι εικόνες και τα βίντεο.
- **Ακρίβεια δεδομένων (Veracity)**: Τα big data μπορεί να είναι ανακριβή ή ατελή. Αυτό μπορεί να οφείλεται σε διάφορους παράγοντες, όπως σφάλματα ανθρώπινης καταγραφής, τεχνικές δυσλειτουργίες όπως λανθασμένη κωδικοποίηση των χαρακτήρων, ή από το γεγονός ότι προέρχονται από διαφορετικές πηγές αλλά και από εσκεμμένες παραποιήσεις. Η αξιολόγηση της ακρίβειας των big data είναι μια σημαντική πρόκληση στην ανάλυση αυτών.



Τίτλος: «**Μεγάλα Δεδομένα**»

Έκδοση: **1.5**

Ημερομηνία: **10/09/2025**

Συντονιστής ομάδας σχεδιασμού και ανάπτυξης: **Κέλλυ Σαρρή Πασχαλίδη**

Δημιουργία: **ΕΚΔΟΣΕΙΣ ΕΛΛΗΝΙΚΗ ΓΡΑΦΗ**



Το παρόν αναπτύχθηκε στο πλαίσιο της Πράξης «Συγγραφή, Αξιολόγηση και Ένταξη διδακτικών βιβλίων στο Μητρώο Διδακτικών Βιβλίων και στην Ψηφιακή Βιβλιοθήκη Διδακτικών Βιβλίων» με κωδικό ΟΠΣ (ΜΙΣ) 6010165, του Προγράμματος «Ανθρώπινο Δυναμικό και Κοινωνική Συνοχή 2021-2027» που υλοποιείται από το Ινστιτούτο Εκπαιδευτικής Πολιτικής και συγχρηματοδοτείται από το Ευρωπαϊκό Κοινωνικό Ταμείο.



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Υπουργείο Παιδείας, Θρησκευμάτων
και Αθλητισμού

ΙΕΠ
ΙΝΣΤΙΤΟΥΤΟ
ΕΚΠΑΙΔΕΥΤΙΚΗΣ
ΠΟΛΙΤΙΚΗΣ



Με τη συγχρηματοδότηση
της Ευρωπαϊκής Ένωσης



Πρόγραμμα
Ανθρώπινο Δυναμικό και
Κοινωνική Συνοχή