

# ΜΕΛΕΤΗ ΚΑΙ ΥΛΟΠΟΙΗΣΗ ΤΕΧΝΙΚΩΝ ΑΝΑΛΥΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ ΚΕΙΜΕΝΟΥ ΣΕ ΠΟΛΥΜΕΣΙΚΟ ΠΕΡΙΕΧΟΜΕΝΟ

Είναι γεγονός ότι στον σημερινό ψηφιακό κόσμο, ο όγκος των ψηφιακών πληροφοριών τείνει ολοένα και αυξανόμενος. Ο μέσος άνθρωπος/χρήστης έρχεται καθημερινά αντιμέτωπος με ένα «ψηφιακό χάος», είτε αυτό αφορά την εργασία του

All time most popular tags

06 africa amsterdam animals architecture art asia august australia autumn baby barcelona beach berlin birthday black blackandwhite blue boston bw california cameraphone camping canada canon car cat cats chicago china christmas church city clouds color concert dso day dc dog england europe fall family festival film florida flower flowers food france friends fun garden geotagged germany girl graffiti green halloween hawaii hiking holiday home honeymoon hongkong house india ireland island italy japan july kids la lake landscape light live london losangeles macro march me mexico mountain mountains museum music nature new newyork newyorkcity newzealand night nikon nyc ocean paris park party people portrait red river roadtrip rock rome san sanfrancisco scotland sea seattle show sky snow spain spring street summer sun sunset sydney taiwan texas thailand tokyo toronto travel tree trees trip uk urban usa vacation vancouver washington water wedding white winter yellow york zoo

είτε την ψυχαγωγία του. Στο πλαίσιο ενός τέτοιου περιβάλλοντος ο χρήστης «βομβαρδίζεται» καθημερινά από ψηφιακό πολυμεσικό περιεχόμενο: ψηφιακές φωτογραφίες, σελίδες υπερκειμένου (hypertext), blogs, forum, MMS, ψηφιακές ταινίες και σειρές, μέχρι πολύπλοκα αλληλεπιδραστικά περιβάλλοντα δύο ή περισσότερων διαστάσεων (3D).

Εν έτη 2010, το πολυδύναμο αυτό ψηφιακό υλικό αλληλεπιδρά με μια πληθώρα συσκευών χρήστη (όπως λ.χ. Η/Υ, smart phones, set-top-boxes, TVs, κ.α.), οι οποίες αποτελούν πλέον μικρούς υπολογιστές και είναι ικανές για την αναπαραγωγή, αποθήκευση και επεξεργασία του πολυμεσικού αυτού περιεχομένου. Καθώς το τελευταίο αυτό περιεχόμενο συνοδεύεται σχεδόν πάντα από επιπρόσθετες πληροφορίες κειμένου που το εμπλουτίζουν και το κάνουν πιο φιλικό προς τον εκάστοτε χρήστη του (όπως π.χ. γίνεται με την προσθήκη **tags** στις φωτογραφίες), η σχετική ερευνητική κοινότητα επενδύει συνεχώς στις τεχνολογίες που χρησιμοποιούν και αναπτύσσουν τεχνικές ανάλυσης και αξιοποίησης αυτών των συνοδευτικών πληροφοριών σε πραγματικό ή μη χρόνο.

Στο παραπάνω πλαίσιο, λαμβάνοντας υπόψη και την ίδια τη μορφή αυτών των δεδομένων κειμένου, καθίσταται σαφές ότι απαιτείται η εφαρμογή προηγμένων μεθόδων πληροφορικής για την ανάλυση και κυρίως τον αποδοτικό συνδυασμό τους με το υπόλοιπο πολυμεσικό υλικό (φωτογραφίες, βίντεο). Έτσι, στο πλαίσιο της παρούσας διπλωματικής εργασίας, θα πρέπει να μελετηθούν αναλυτικά τα στάδια της ανάλυσης δεδομένων κειμένου και να υλοποιηθούν **μέρος ή όλα** από τα παρακάτω:

- **Μελέτη** της κλασικής βιβλιογραφίας του χώρου και των σύγχρονων τεχνικών ανάλυσης, σύγκρισης και κατηγοριοποίησης δεδομένων κειμένου.
- **Ανάπτυξη και υλοποίηση** (π.χ. Java) τεχνικών προ-επεξεργασίας των δεδομένων.
- **Μελέτη** μεθόδων επιλογής χαρακτηριστικών και μείωσης των διαστάσεων των δεδομένων.
- **Ανάπτυξη και υλοποίηση** (π.χ. Java) σχετικών αλγορίθμων μηχανικής μάθησης για αποδοτική εξαγωγή γνώσης από τα δεδομένα.

## Προαπαιτούμενα

Καλή γνώση προγραμματισμού, Java ή C++/C#

## Επιθυμητή γνώση

Android OS, iPhone OS, Γνώση τεχνολογιών web

## Επικοινωνία

Δρ. Φοίβος Μυλωνάς

τηλ.: 210-7724351

e-mail: [fmylonas@image.ntua.gr](mailto:fmylonas@image.ntua.gr)

